Una base para el Álgebra Lineal

Ángel Miguel Amores Lázaro

Índice general

1. Matrices y sistemas de lineales **1**

1.1. Cuerpos . . . 1

1.1.1. Números complejos . . . 1

1.1.2. Clases de restos módulo n . . . 6

1.1.3. La ecuación de segundo grado . . . 10

1.2. Una notación esencial: los sumatorios . . . 11

1.3. Matrices . . . 13

1.3.1. Conceptos generales . . . 13

1.3.2. Tipos especiales de matrices . . . 20

1.3.3. Un importante comentario sobre la notación . . . 22

1.4. Sistemas de ecuaciones lineales . . . 22

1.4.1. Cuestiones generales . . . 23

1.4.2. Matrices y sistemas en forma escalonada y reducida . . . 24

1.4.3. Operaciones fila en matrices y sistemas lineales . . . 26

1.4.4. Transformación de matrices . . . 28

1.4.5. Solución de sistemas lineales . . . 31

1.4.6. La información que dan los sistemas lineales . . . 33

1.5. Matrices elementales e inversión de matrices . . . 34

1.6. Operaciones con columnas . . . 39

1.7. El principio de inducción . . . 42

1.8. Más demostraciones por inducción . . . 45

2. Espacios vectoriales **49**

2.1. El espacio vectorial estándar . . . 49

2.1.1. Los teoremas principales . . . 53

2.1.2. Espacios dados en implícitas . . . 55

2.1.3. Espacios dados en paramétricas . . . 55

2.2. Espacios vectoriales . . . 58

2.3. Subespacios vectoriales . . . 61

2.4. Bases . . . 63

2.5. El rango de una matriz . . . 67

2.6. Intersección y suma de subespacios vectoriales . . . 71

2.6.1. Intersección de subespacios vectoriales . . . 71

2.6.2. Suma de subespacios vectoriales. Sumas directas . . . 72

2.7. Sobre subespacios afines . . . 77

3. Funciones lineales **83**

3.1. Cuestiones generales . . . 83

3.2. Funciones lineales y matrices . . . 87

3.3. Rangos y dimensiones . . . 90
3.4. Espacios de funciones lineales . . . 94
3.5. Efectos de cambios de base . . . 101
3.6. La traza de un endomorfismo . . . 106
3.7. Referencias y funciones afines . . . 106

4. Determinantes **115**
4.1. Permutaciones . . . 115
4.1.1. Cuestiones básicas sobre permutaciones . . . 115
4.1.2. Exposición más detallada de la signatura . . . 119
4.2. Determinantes . . . 123
4.3. El determinante como función de los vectores columna . . . 125
4.4. Sobre cálculo de determinantes y sus aplicaciones . . . 131
4.4.1. Cálculo de determinantes por cajas . . . 131
4.4.2. Cálculo de la inversa y desarrollo por filas o columnas . . . 132
4.4.3. Otras cuestiones sobre determinantes . . . 134
4.5. El determinante de un endomorfismo . . . 139

5. Autovalores y autovectores **141**
5.1. Una breve visita a los polinomios . . . 141
5.2. Autovalores y autovectores . . . 146
5.3. El polinomio característico . . . 147
5.4. Valores propios de matrices simétricas y hermitianas . . . 150
5.5. Diagonalización . . . 151
5.6. El teorema de Cayley-Hamilton . . . 157
5.7. Diagonalización y fórmulas recursivas . . . 159
5.8. Ecuaciones diferenciales matriciales. El caso fácil . . . 160
5.8.1. El caso real . . . 160
5.8.2. El caso complejo . . . 162
5.9. El polinomio minimal . . . 164
5.10. Una segunda visita a la diagonalización . . . 165
5.11. Triangulación . . . 168
5.12. Un primer contacto con la forma de Jordan . . . 171
5.13. Triangulación en el caso real . . . 173
5.14. Ecuaciones lineales homogéneas con coeficientes constantes . . . 174
5.15. Ecuaciones lineales no homogéneas . . . 178

6. Potencias y exponencial de endomorfismos **183**
6.1. Sobre polinomios . . . 183
6.2. Expresión de un endomorfismo con proyecciones . . . 185
6.3. La descomposición $L = D + N$ . . . 186
6.4. Un breve contacto con las ecuaciones diferenciales lineales . . . 193
6.4.1. Qué son estas ecuaciones y cómo son sus soluciones . . . 193
6.4.2. La exponencial de una matriz y su cálculo . . . 193

7. La forma de Jordan **199**
7.1. Planteamiento y pasos iniciales . . . 199
7.2. La descomposición primaria . . . 201
7.3. El caso nilpotente . . . 202
7.4. La tabla $*$ de una base de Jordan (caso nilpotente) . . . 204

7.5. Matrices y bases de Jordan en general . . . 208
7.6. Otras cuestiones sobre la forma de Jordan . . . 213

8. Formas bilineales simétricas 219
8.1. Funciones bilineales . . . 219
8.1.1. Matrices de una función bilineal . . . 220
8.1.2. Efectos de cambios de base . . . 222
8.1.3. Distinciones entre endomorfismos y funciones bilineales . . . 225
8.2. Formas bilineales simétricas . . . 226
8.3. Formas cuadráticas . . . 234
8.3.1. La reducción de Lagrange . . . 235
8.4. Una aplicación al Análisis . . . 239
8.5. Cuádricas afines . . . 241
8.5.1. Centros de una cuádrica . . . 242
8.5.2. Ecuación normalizada si hay centros . . . 242
8.5.3. Ecuación normalizada si no hay centros . . . 244
8.5.4. Tipos de cónicas y cuádricas en dimensión 2 y 3 . . . 245
8.5.5. Cálculos alternativos para clasificar cuádricas . . . 248

9. Productos euclidianos 251
9.1. Fundamentos . . . 251
9.1.1. Principales definiciones y ejemplos . . . 251
9.1.2. Bases ortogonales y ortonormales . . . 253
9.1.3. Matrices y determinantes de Gram . . . 256
9.1.4. Matrices ortogonales . . . 257
9.2. Un primer contacto con el producto vectorial . . . 258
9.3. El producto vectorial. Orientaciones . . . 260
9.4. Sumas, proyecciones y simetrías ortogonales . . . 263
9.5. La ortogonalización de Gram-Schmidt . . . 267
9.6. Factorizaciones . . . 269
9.6.1. La factorización QR . . . 269
9.6.2. La factorización de Cholesky . . . 270
9.7. Distancias y desigualdades . . . 271
9.8. Cuestiones diversas . . . 276
9.8.1. Ángulos no orientados . . . 276
9.8.2. Semiespacios y semiplanos . . . 278
9.8.3. Aproximaciones óptimas . . . 278
9.8.4. El criterio de Sylvester . . . 279
9.9. Conceptos afines euclidianos . . . 280

10. Funciones en espacios euclidianos 285
10.1. Adjunto de un endomorfismo . . . 285
10.2. Endomorfismos normales . . . 287
10.3. Los teoremas espectrales . . . 291
10.3.1. Teoremas espectrales para endomorfismos . . . 291
10.3.2. Teoremas espectrales para formas simétricas . . . 294
10.3.3. Un resumen de los teoremas espectrales . . . 298
10.4. Cuádricas en el espacio euclidiano . . . 298
10.4.1. Ecuación normalizada si hay centros . . . 299
10.4.2. Ecuación normalizada si no hay centros . . . 300

10.5. Isometrías o transformaciones ortogonales . . . 303
10.6. Isometrías del plano euclidiano . . . 304
10.7. Ángulos y orientaciones en el plano . . . 305
10.7.1. Orientación de bases del plano . . . 307
10.7.2. Ángulo de una rotación del plano orientado . . . 309
10.8. Teorema de estructura de isometrías . . . 311
10.8.1. Clasificación de las isometrías en dimensión dos . . . 313
10.8.2. Clasificación de las isometrías en dimensión tres . . . 313
10.9. Factorización con simetrías respecto a hiperplanos . . . 317
10.10Rotaciones, rotosimetrías, y producto vectorial . . . 321
10.11Endomorfismos antisimétricos . . . 324
10.12Teorema estructural de isometrías afines . . . 326
10.13Descripción de las isometrías . . . 330
10.13.1El caso bidimensional . . . 331
10.13.2El caso tridimensional . . . 331
10.14Problemas sobre isometrías afines . . . 332
10.15Factorización de isometrías afines . . . 334
10.16Más problemas sobre isometrías afines . . . 335

Prefacio

El autor ha dado en la Universidad Complutense durante tres años la asignatura de Álgebra Lineal del curso del Doble Grado en Economía, Matemáticas y Estadística y otro año más el mismo curso pero para los Dobles Grados de Matemáticas y Física y de Matemáticas e Informática. Este curso 2018-19 será para los mismos Dobles Grados pero con otros estudiantes de los demás Grados que tiene la Facultad de Matemáticas. Desde el principio ha habido un texto que ha ido evolucionando, en parte por la mayor experiencia del profesor, y en parte por la reacción de los estudiantes. El que ahora se presenta es prácticamente el del curso pasado 2017-18 para los Dobles Grados mencionados y, salvo las inevitables correcciones, será la versión definitiva. Las modificaciones al ir pasando de curso han sido más de cómo se cuenta que de qué se cuenta, pero cree el autor que esto tiene mucha trascendencia, sobre todo al iniciar los estudios. Vaya por delante que se entrega a los alumnos de la asignatura a través del Campus Virtual sin coste alguno para ellos ni beneficio económico para quien lo escribe.

Hay muchos libros de Álgebra Lineal que superan a este en muchos aspectos, pero siempre parece que el "libro del profesor" es inevitable. Efectivamente, creemos que es mucho pedir que un estudiante de Primer Año ande por ahí buscando la materia en uno u otro texto por muchas facilidades que den las herramientas informáticas. Mientras sea buscar datos, la cosa va bien, pero compaginar diversas notaciones, enfoques y enunciados de teoremas, aunque al final cubran la misma materia, es muy difícil. Hay otra alternativa, que es adoptar un solo texto, decirle al alumno que lo adquiera y seguirlo con todas las consecuencias. No lo hacemos porque no hay ninguno que nos satisfaga plenamente, y no hablamos de su valor científico (los hay excelentes) sino de cómo y cuánto se adapta al curso. Básicamente se pretende que el estudiante se vea libre en buena parte de copiar apuntes y tenga al final del curso una serie de temas básicos entre los que se mueva con familiaridad y sean un bastidor o cimiento sobre el que pueda añadir mucho más según le convenga. No hay decir que cualquiera con ciertas necesidades o curiosidad puede encontrar enseguida que este texto se le queda pequeño en el futuro. Tras todo esto hay que decir que el texto tiene demasiada materia para poderla dar en un solo curso, e iremos advirtiendo al lector de la materia del texto que no es materia de examen.

Hemos procurado que el lector reciba una buena cantidad de información sobre el porqué de lo que se hace, evitando que el texto sea solo una colección de herramientas. Los problemas están intercalados con la teoría, lo que les quita cierta dificultad porque cualquiera piensa que lo que le pregunten se resolverá con lo que le acaban de contar. No siempre es así. El Álgebra Lineal es muy algorítmica (entiéndase, calculista) y es bastante fácil plantear problemas basados en un puro cálculo. Con frecuencia le explicamos al lector como ponerse problemas a sí mismo. No es cuestión de desvelar trucos del oficio, sino de mostrarle de un modo más eficaz que la pura repetición de cálculos con variantes, las ideas subyacentes. Sin duda, cuando uno sabe poner y ponerse problemas de este tipo es cuando de verdad entiende su naturaleza y solución. Decimos esto también porque puede parecer que hay pocos problemas en relación con la cantidad de teoría. Hay muchos más de los que parece si se añaden estos problemas que no se numeran como tales y que el lector puede hacer a su gusto, y también hay muchas comprobaciones que debe hacer de propiedades sencillas de ciertas estructuras.

Capítulo 1

Matrices y sistemas de lineales

1.1. Cuerpos

En Álgebra Lineal se manejan varios tipos de números, siendo los más interesantes aquellos que forman un **cuerpo**. No damos aún la definición técnica de tales números pero, de una manera aproximada, son aquellos que se pueden sumar y multiplicar con las reglas usuales de la Aritmética y en donde todos ellos, excepto 0, tienen inverso multiplicativo. Para no entrar a un nivel demasiado abstracto, vamos a suponer que el lector sabe manejar los "números", que son en realidad los **números reales** y se representarán por $\mathbb{R}$. Ejemplos de estos números son

$$0, +1, -1, +2, -2, \ldots \qquad \frac{1}{3}, -\frac{3}{5}, \ldots, \pi, \sqrt{\pi}, \sqrt{3}, 2^{-1/3}, \ldots$$

aunque en los cálculos aparecerán casi siempre $0, +1, -1, +2, -2, \ldots$ para simplificar. Dentro de los números **reales** $\mathbb{R}$ están los **naturales** $\mathbb{N} = 1, 2, 3, \ldots$, los **enteros** $\mathbb{Z} = 0, \pm 1, \pm 2, \pm 3, \ldots$, y los **racionales** $\mathbb{Q}$, que son las fracciones m/n con $m, n \in \mathbb{Z}$ y $n \neq 0$. Las letras $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ y $\mathbb{R}$ denotan los conjuntos de estos números y $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. Muchos autores consideran que 0 es un número natural y para ellos $\mathbb{N}$ es $0, 1, 2, \ldots$ pero nosotros denotaremos este conjunto por $\mathbb{N}_0$, luego $\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z}$. Una cosa es saber manejar estos números y otra mucho más difícil es definir de manera precisa qué son, de modo que sepamos exactamente qué se puede hacer con ellos. Los problemas que han justificado históricamente esta necesidad de rigor quedan muy lejos y es por lo que adoptamos esta actitud más laxa.

Vamos a construir de modo no totalmente formal pero sí con cierto detalle los **números complejos** y **las clases de restos módulo** p (con p primo) a lo que dedicaremos las dos secciones siguientes. En la sección sobre los números complejos se intercalará la noción de cuerpo. La última sección trata la ecuación de segundo grado, que nos será necesaria y ayudará a entender estos conceptos.

1.1.1. Números complejos

Ahora explicaremos (o recordaremos) los **números complejos**, cuyo conjunto se denota por $\mathbb{C}$. Damos primero una definición "operativa" (para echar cuentas) de lo que son y un poco más adelante vendrá la definición rigurosa.

Suponemos al lector cierta habilidad de cálculo con polinomios en una variable, digamos X, con coeficientes en $\mathbb{R}$. Sabemos sumar polinomios, multiplicarlos por números, y multiplicarlos entre sí. La definición "para empezar" de los números complejos es la que sigue. Consideremos polinomios *de primer grado* donde en lugar de X aparece la letra i. Estos van a ser los números complejos. Los sumamos entre sí y multiplicamos por números reales como se hace con polinomios. El problema es cómo multiplicar complejos pues si los multiplicamos como polinomios encontraríamos enseguida polinomios de grado ≥ 2 que no son números complejos. Hacemos el convenio de que cada vez que al multiplicar aparezca i^2 lo podemos sustituir por -1. Por ejemplo,

$$(1+2i)(1-3i) = 1 - i - 6i^2 = 7 - i, \qquad (1+2i)^2(1-3i) = \begin{cases} (1+2i)(7-i) = 9 + 13i \\ (-3+4i)(1-3i) = 9 + 13i \end{cases}$$

Como se ve, las dos maneras de calcular $(1+2i)^2(1-3i)$ llevan al mismo resultado $9+13i$ aunque no sabemos si es casual o no. Admitamos de momento que no habrá problemas al elegir el tipo de vías alternativas que tomamos con números reales. El enunciado más formal de la regla del producto es

$$(a+bi)(u+vi) = (au-bv)+(av+bu)i, \tag{1.1}$$

y el signo $-$ en $-bv$ aparece porque $(bi)(vi) = bv\left(i^2\right) = -bv$.

Podemos considerar que $\mathbb{R}$ es un subconjunto de $\mathbb{C}$, identificando $r \in \mathbb{R}$ con $r+0i$, que se escribe como r igual que el polinomio $a+0X$ es a. Evidentemente $0 = 0+0i$ y $1 = 1+0i$ tienen las propiedades análogas de 0 y 1 en $\mathbb{R}$. Por ejemplo $0+\sigma = \sigma+0 = \sigma$ para todo $\sigma \in \mathbb{C}$, y $\sigma 1 = 1\sigma = \sigma$. Sabemos que si r es real o racional no nulo, existe el inverso multiplicativo; o sea, otro número s tal que $rs = sr = 1$, y se usa la notación $s = r^{-1}$ o $s = 1/r$. Nos preguntamos si lo mismo es cierto en $\mathbb{C}$. Sea $\sigma = a+bi \neq 0$ y buscamos $\tau = u+vi$ tal que $\sigma\tau = 1 = 1+0i$ (al ser el producto conmutativo, $\tau\sigma = 1$ también). Usaremos que $\sigma \neq 0$ equivale a que tanto a como b son $\neq 0$, luego $a^2+b^2 \neq 0$. Se constata que

$$(a+bi)(a-bi) = a^2-b^2i^2-abi+abi = a^2+b^2, \qquad (a+bi)\frac{a-bi}{a^2+b^2} = 1+0i = 1.$$

Hemos hecho algunas simplificaciones como escribir

$$\frac{a-bi}{a^2+b^2} \text{ en vez de } \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$$

que sería lo estrictamente correcto, pero de un modo u otro, este complejo es σ^{-1} para $\sigma \neq 0 = 0+0i$. Por ejemplo, suponiendo en el segundo caso que $h \in \mathbb{R}$ y $h \neq 0$,

$$\frac{1}{1+2i} = \frac{1-2i}{1^2+2^2} = \frac{1-2i}{5} = \frac{1}{5} - \frac{2}{5}i, \qquad \frac{1}{h+hi} = \frac{h-hi}{h^2+h^2} = \frac{h-hi}{2h^2} = \frac{1-i}{2h}.$$

Es muy conveniente el identificar el complejo $\sigma = a+bi$ con el punto (a,b) del plano $\mathbb{R}^2$. Entonces $\sqrt{a^2+b^2}$ representa la distancia al origen y se llama el **módulo**, **norma** o **valor absoluto** de σ. Se representa por $|\sigma|$ (como el valor absoluto de un número real) y es obvio que $\sigma \neq 0$ equivale a $|\sigma| \neq 0$. Dado el número complejo $\sigma = a+bi$ definimos su **conjugado** como $\bar{\sigma} = a-bi$ (con barra sobre σ).[1] Si vemos σ como el punto (a,b), $\bar{\sigma}$ es el simétrico de σ respecto al eje OX. Es inmediato que

$$\sigma\bar{\sigma} = a^2+abi-abi+b^2i^2 = a^2+b^2, \qquad \sigma\bar{\sigma} = |\sigma|^2, \qquad \sigma^{-1} = \frac{\bar{\sigma}}{|\sigma|^2} \quad (\sigma \neq 0).$$

El módulo y el conjugado tienen una serie de propiedades muy sencillas y que se usan de modo continuo,

$$\overline{\sigma+\tau} = \bar{\sigma}+\bar{\tau}, \quad \overline{\sigma\tau} = \bar{\sigma}\bar{\tau}, \quad \overline{(\sigma^{-1})} = \frac{1}{\bar{\sigma}} = (\bar{\sigma})^{-1}, \quad \overline{(\bar{\sigma})} = \sigma, \quad |\sigma\tau| = |\sigma|\,|\tau|, \quad \left|\sigma^{-1}\right| = |\sigma|^{-1},$$

aunque es *falso* que sea $|\sigma+\tau| = |\sigma|+|\tau|$. Se recuerdan como "el conjugado de la suma (producto, inverso) es la suma (producto, inverso) de conjugados", que "conjugar dos veces deja el numero invariante", que "el módulo del producto es el producto de los módulos" y que "el módulo del inverso es el inverso del módulo". Finalmente, se definen la **parte real** y **parte imaginaria** de σ como

$$\mathrm{Re}(\sigma) = \frac{\sigma+\bar{\sigma}}{2}, \qquad \mathrm{Im}(\sigma) = \frac{\sigma-\bar{\sigma}}{2i}.$$

Es fácil comprobar que si nos dan $\sigma = a+bi$, sus partes real e imaginaria son a y b, que son números reales gracias al $2i$ en el denominador. Se dice que $\sigma \in \mathbb{C}$ es **real** si $\mathrm{Im}(\sigma) = 0$, que equivale a $\mathrm{Re}(\sigma) = \sigma$; y que σ es **imaginario** si $\mathrm{Re}(\sigma) = 0$. Al considerar $\mathbb{R} \subset \mathbb{C}$ como explicamos más arriba, los elementos de $\mathbb{R}$ son justamente los números reales con parte imaginaria nula. Si $\mathrm{Re}(\sigma) = 0$, σ es **imaginario puro.** Un complejo γ se llama **unitario** si $|\gamma| = 1$, que equivale a $\mathrm{Re}(\gamma)^2 + \mathrm{Im}(\gamma)^2 = 1$ o $a^2+b^2 = 1$ si $\gamma = a+bi$ con $a,b \in \mathbb{R}$. Evidentemente γ es unitario si y solo si su inverso es su conjugado.

Problema 1 *Probar todas las fórmulas vistas hasta ahora y no demostradas.*

[1] A veces se pone σ^* en lugar de $\bar{\sigma}$ por razones tipográficas.

Abordamos ahora los números complejos desde una perspectiva más rigurosa. Hay dos estructuras algebraicas abstractas, llamadas **anillo** y **cuerpo**, que ahora definiremos.[2] El anillo tiene como ejemplo principal a $\mathbb{Z}$, los números enteros, y el cuerpo tiene como ejemplos a $\mathbb{Q}, \mathbb{R}$ y $\mathbb{C}$, quedando los números naturales fuera de estas definiciones. La de cuerpo es más restrictiva, pues todo cuerpo es un anillo.

Definimos un **cuerpo** como un conjunto $\Bbbk$ donde para cada par de elementos a, b se definen otros dos elementos denotados por $a+b$ y $a \cdot b$, llamados la **suma** y **producto** de a y b. Estas dos asignaciones u operaciones en $\Bbbk$ cumplen las siguientes propiedades

1. *Asociatividad de la suma.* Para todo $a, b, c \in \Bbbk$ se tiene $a + (b + c) = (a + b) + c$.
2. *Conmutatividad de la suma.* Para todo $a, b \in \Bbbk$ se tiene $a + b = b + a$.
3. *Existencia de cero.* Existe un elemento, que denotaremos por 0, tal que para todo $a \in \Bbbk$ se tiene $a + 0 = 0 + a = a$.
4. *Existencia de opuesto o inverso aditivo.* Para todo $a \in \Bbbk$ existe un elemento que denotaremos por $-a$ tal que $a + (-a) = (-a) + a = 0$.
5. *Asociatividad del producto.* Para todo $a, b, c \in \Bbbk$ se tiene $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
6. *Conmutatividad del producto.* Para todo $a, b \in \Bbbk$ se tiene $a \cdot b = b \cdot a$.
7. *Existencia de unidad.* Existe un elemento, que denotaremos por 1 y *es distinto de* 0, tal que para todo $a \in \Bbbk$ se tiene $a \cdot 1 = 1 \cdot a = a$.
8. *Existencia de inverso multiplicativo.* Para todo $a \in \Bbbk$ con $a \neq 0$, existe un elemento que denotaremos por a^{-1} tal que $a \cdot a^{-1} = a^{-1} \cdot a = 1$.
9. *Distributividad.* Para todo $a, b, c \in \Bbbk$ se tiene $a \cdot (b + c) = a \cdot b + a \cdot c$ y $(a + b) \cdot c = a \cdot c + b \cdot c$.

Casi siempre simplificaremos la notación del producto quitando el punto, aunque lo mantendremos a veces por claridad o énfasis. Se simplifica también $a + (-b)$ a $a - b$. Las notaciones $a^n = a \cdot a \cdot \cdots \cdot a$ (n veces) o $a^{-n} = \left(a^{-1}\right) \cdot \left(a^{-1}\right) \cdot \cdots \cdot \left(a^{-1}\right)$ (n veces) para $n = 1, 2, 3, \ldots$ se mantienen también y $\frac{1}{a}$ como a^{-1}, el inverso de $a \neq 0$.

Si pensamos en los números reales $\mathbb{R}$, vemos que las 9 propiedades han sido desde siempre cosa obvia y lo mismo se puede decir de los números racionales $\mathbb{Q}$. Sin embargo $\mathbb{Z}$ cumple todo excepto la condición **8**, luego no es un cuerpo (es un anillo como luego definiremos). La definición de cuerpo interesa, no como recordatorio de propiedades evidentes que nada nuevo nos dicen, sino como condiciones mínimas que ciertos entes matemáticos deben verificar para poder "ser intuidos como números". La ventaja de fijar los axiomas es que cualquier teorema que hayamos podido probar, digamos que para $\mathbb{R}$, y que solo haya usado los axiomas **1-9**, sirve automáticamente para cualquier otro cuerpo. Por ejemplo, *los teoremas para sistemas lineales serán válidos en un cuerpo arbitrario.*

Ya imaginará el lector que $\mathbb{C}$ es un cuerpo. El lector puede olvidar momentáneamente lo que hemos dicho sobre los números complejos hasta ahora y pensar cómo sería $\mathbb{C}$ de ser cierto (¡que lo es!) el siguiente teorema.

Teorema 1 *Hay un cuerpo $\mathbb{C}$ y solo uno que verifica las siguientes propiedades*

1. *$\mathbb{C}$ contiene a $\mathbb{R}$, el cuerpo de los números reales.*
2. *Hay en $\mathbb{C}$ un elemento $i \notin \mathbb{R}$, que llamaremos la **unidad imaginaria**, tal que todo $\sigma \in \mathbb{C}$ se escribe de la forma $\sigma = a + bi$ con $a, b \in \mathbb{R}$ e $i^2 = -1$*

Demostración. (esquema) Supongamos primero que al menos uno de estos cuertos existe y veamos que no hay más que uno. Si se escribe $\sigma = a_1 + b_1 i = a_2 + b_2 i$, tiene que ser $a_1 = a_2$ y $b_1 = b_2$. En efecto, como $a_1 - a_2 = -i\,(b_1 - b_2)$, elevando al cuadrado,

$$(a_1 - a_2)^2 = (-i)^2 (b_1 - b_2)^2 = -(b_1 - b_2)^2$$

[2] En inglés se usa *field* (campo) y en muchas traducciones al español aparece "campo". Es de los pocos casos en los que el francés *corps* ha prevalecido sobre el inglés.

Al ser $(a_1-a_2)^2 \geq 0$ y $-(b_1-b_2)^2 \leq 0$ se debe cumplir que $(a_1-a_2)^2 = (b_1-b_2)^2 = 0$ y $a_1 = a_2$ y $b_1 = b_2$. Escribiendo los $\sigma, \tau \in \mathbb{C}$ en la forma *única* $\sigma = a+bi$, $\tau = u+vi$ con $a, b, u, v \in \mathbb{R}$, se verifica que las fórmulas de $\sigma + \tau$ y $\sigma\tau$, *que no nos da directamente el teorema*, tienen que ser las que ya conocemos de la introducción informal. Por ejemplo,

$$\begin{aligned}\sigma\tau &= (a+bi)(u+vi) \overset{9}{=} a(u+vi) + bi(u+vi) \overset{9}{=} au + avi + biu + (bi)(vi) \\ &\overset{6}{=} au + avi + bui + bvii \overset{2}{=} au + bvii + avi + bui \overset{i^2=-1}{=} au + bv(-1) + avi + bui \\ &\overset{9}{=} (au - bv) + (av + bu)i\end{aligned}$$

indicando con $\overset{n}{=}$ que en ese paso se usa el axioma n de cuerpo. La comprobación de que con los axiomas de cuerpo, la fórmula $\sigma + \tau = (a+u) + (b+v)i$ es más rápida. No vamos a hacer la demostración formal del teorema, que se cita simplemente para que el lector sepa lo que puede hacer *operando* con los elementos de $\mathbb{C}$, que es prácticamente lo mismo que puede hacer con los números reales. Sin embargo, al *sacar conclusiones* de las ecuaciones e igualdades, la situación cambia y, a veces, mucho. ♣

Hay una ambigüedad que induce en muchos casos a confusión. Hemos sido cuidadosos reservando las letras griegas para elementos de $\mathbb{C}$ y las latinas para elementos de $\mathbb{R}$, de modo que si vemos $\sigma + \tau i = \theta$ pensemos que tenemos dos complejos σ y τ, que multiplicamos τ por i y le sumamos σ, obteniendo θ. Digamos ahora que el convenio queda anulado y que en $a+bi$ cabe la posibilidad de que a y b no sean forzosamente números reales. Esto crea una diferencia. Si tenemos $\sigma \in \mathbb{C}$ escrito como $\sigma = a+bi$ y nos dicen que $a, b \in \mathbb{R}$, los números $a, b \in \mathbb{R}$ son únicos (visto al probar el teorema) de modo que si, por ejemplo, $\sigma = 0$, debe ser $a = b = 0$. Sin embargo, si nos dan $\sigma = a+bi$ pero pudiendo ser a, b complejos no reales, se tiene que a y b no vienen unívocamente determinados por σ. Puede suceder que $\sigma = a_1 + b_1 i = a_2 + b_2 i$ siendo $a_1 \neq a_2$ o $b_1 \neq b_2$. Por ejemplo,

$$0 = 0 + 0i = 1 + i \cdot i, \qquad a_1 = b_1 = 0, \quad a_2 = 1,\ b_2 = i.$$

Hay que tratar también con cuidado las sumas de cuadrados. Hay una relación de orden en los elementos de $\mathbb{R}$, mientras que $\mathbb{C}$ no tiene relación de orden entre sus elementos. En $\mathbb{R}$, si $a \neq 0$, se tiene que $a^2 > 0$ y si $a^2 + b^2 = 0$ se tiene $a = b = 0$. En $\mathbb{C}$ se pueden considerar σ^2 y τ^2, pero $\sigma^2 + \tau^2 = 0$ *no implica* que sea $\sigma = \tau = 0$, como prueban $\sigma = 1$ y $\tau = i$.

Problema 2 *Realizar las siguientes operaciones*

$$(2-i)\,i, \quad \left(1+\sqrt{2}i\right)\left(1-\sqrt{2}i\right), \quad \left(1+\sqrt{2}i\right)^3, \quad \left(1+\sqrt{2}i\right)^{-1}, \quad \left(1+\sqrt{2}i\right)^{-2}.$$

Además hacemos algunas preguntas:

1. *Si tenemos $\gamma = a+bi$ con a, b números reales ambos no nulos. ¿Puede suceder que $\gamma^2 \in \mathbb{R}$?*
2. *Calcular la raíz cuadrada (si existe) de $\gamma = 1+i$; es decir, σ tal que $\sigma^2 = \gamma$.*

En **2** del problema precedente hemos hablado de raíces cuadradas. Si $\Bbbk$ es un cuerpo *arbitrario* diremos que $a \in k$ tiene una **raíz cuadrada** r si r *es un elemento de* $\Bbbk$ y $r^2 = a$. Parece excesivo insistir con la cursiva en que $r \in \Bbbk$, pero es conveniente con los cuerpos $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. En efecto, preguntarnos si 2 tiene raíz cuadrada es bastante ambiguo puesto que $2 \in \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. ¿Cuál es el cuerpo $\Bbbk$? Si $\Bbbk = \mathbb{Q}$ es probable que el lector sepa de los cursos de Análisis que no existe ningún $r \in \mathbb{Q}$; o sea, una *fracción*, tal que $r^2 = 2$. Sin embargo si vemos 2 como *número real* sí existe un *número real* r (aproximadamente 1,41, pero tiene infinitas cifras decimales) tal que $r^2 = 2$. De hecho, todo $a \in \mathbb{R}$ con $a \geq 0$ tiene una raíz cuadrada. Si $a < 0$, *a entendido como número real*, no tiene raíz cuadrada, pero si se ve a como caso particular de número complejo, sí que la tiene. Puede haber o no haber raíces cuadradas, pero si las hay, hay exactamente dos.

Problema 3 *Probar el aserto anterior en $\Bbbk$ arbitrario. Indicación: suma por diferencia, diferencia de cuadrados.*

Advertimos que la notación $\sqrt{a}$ para la raíz cuadrada de a debe usarse con cuidado. Si $a \in \mathbb{R}$ y $a > 0$, hay dos raíces, una positiva y otra negativa y $\sqrt{a}$ representa la positiva, siendo esto una definición. Sin

embargo, para hablar de positivo y negativo hay de disponer de una relación de orden $>$ en $\Bbbk$. Dado que $\mathbb{C}$, y en general $\Bbbk$ arbitrario no la tienen, $\sqrt{a}$ es una de las dos raíces cuadradas (suponiendo que existan).

No es obvio, pero sí es cierto, que todo $a \in \mathbb{C}$ tenga raíz cuadrada. Se podría ver con un cálculo similar al hecho para $1+i$ en el problema precedente, pero lo vamos a hacer en el teorema que sigue por un procedimiento más directo, que sirve para usar los conceptos de módulo, conjugado, etc. Este teorema es un buen ejemplo de que *muchas veces no es conveniente hacer cálculos desglosando un complejo en su parte real e imaginaria.*

Teorema 2 *Todo complejo unitario $u \neq -1$ cumple*

$$\frac{(1+u)^2}{|1+u|^2} = u, \quad \text{y por tanto } \frac{1+u}{|1+u|} \text{ es una raíz cuadrada de } u.$$

Con más generalidad, si z no es *un número real negativo, para $u = z/|z|$,*

$$\sqrt{|z|}\frac{1+u}{|1+u|} = \sqrt{|z|}\frac{1+\frac{z}{|z|}}{\left|1+\frac{z}{|z|}\right|} = \sqrt{|z|}\frac{|z|+z}{||z|+z|}$$

es una raíz cuadrada de z.

Demostración. Son simples comprobaciones de una idea feliz,

$$\begin{aligned}|1+u|^2 u &= (1+u)\,\overline{(1+u)}u = (1+u)(1+\bar{u})\,u = (1+u+\bar{u}+\bar{u}u)\,u = (2+u+\bar{u})\,u \\ &= 2u+u^2+\bar{u}u = 2u+u^2+1 = (1+u)^2.\end{aligned}$$

Esto vale para todo u aunque para poder dividir y llegar a la primera fórmula hace falta que $1+u \neq 0$. La segunda fórmula resulta de que para $z \neq 0$ podemos poner $z = |z|\,u$ con $u = z/|z|$. ♣

Advertimos que las fórmulas explícitas de las raíces cuadradas son "feas" pues aparecen fácilmente raíces cuartas (o sea $a^{1/4} = \sqrt[4]{a}$ con $a \geq 0$). Calculamos para $z = 1+i$, que no es unitario, su raíz cuadrada. Los pasos son

$$|z| = \sqrt{1^2+1^2} = \sqrt{2}, \qquad \sqrt{|z|} = 2^{1/4},$$

$$|z|+z = \left(\sqrt{2}+1\right)+i, \qquad \left|\left(\sqrt{2}+1\right)+i\right| = \sqrt{\left(\sqrt{2}+1\right)^2+1^2} = \sqrt{2\sqrt{2}+4} = \sqrt{2}\sqrt{\sqrt{2}+2}$$

$$\sqrt{|z|}\frac{|z|+z}{||z|+z|} = 2^{1/4}\frac{\left(\sqrt{2}+1\right)+i}{\sqrt{2}\sqrt{\sqrt{2}+2}} = \frac{\left(\sqrt{2}+1\right)+i}{2^{1/4}\sqrt{\sqrt{2}+2}}.$$

Como se ve, salen fórmulas muy complicadas y el escribir los complejos con senos y cosenos, como pronto veremos, es una ayuda.

Problema 4 *Sea $a \in \mathbb{R}$ no nulo. Calcular una raíz cuadrada de $z = ai$. Por otra parte, supongamos que $z \in \mathbb{C}$ tiene como raíz cuadrada un número imaginario. ¿Tiene que ser z real?*

Hay una alternativa que es utilizar que si $u = x+yi$, con $x, y \in \mathbb{R}$, es unitario, se tiene $x^2+y^2 = 1$; o sea, el complejo u, visto como punto del plano $\mathbb{R}^2$, está en el círculo de centro el origen y radio 1, que denotaremos por $\mathbb{U}$. Estos puntos de $\mathbb{U}$ se pueden escribir como $(x,y) = (\cos\alpha, \sin\alpha)$ para $\alpha \in \mathbb{R}$ determinado salvo múltiplo entero de 2π. Cualquier $z \neq 0$ en $\mathbb{C}$ se escribe como

$$z = |z|\,\frac{z}{|z|} = |z|\,(\cos\alpha, \sin\alpha).$$

Esta expresión trigonométrica de $\mathbb{C}$ tiene muchas ventajas cuando hay que multiplicar, debido al gran número de propiedades del coseno y seno. Volvamos al ejemplo de $z = 1+i$ donde nos piden calcular una de sus raíces cuadradas. Calculamos $|z| = \sqrt{2}$ y a ojo $z/|z| = \cos(\pi/4) + i\sin(\pi/4)$ (¡no siempre será tan fácil!). En general, si $u = \cos\theta + i\sin\theta$ y $v = \cos(\theta/2) + i\sin(\theta/2)$ tenemos que

$$v^2 = \left(\cos^2\frac{1}{2}\theta - \sin^2\frac{1}{2}\theta\right) + i\left(2\cos\frac{1}{2}\theta\sin\frac{1}{2}\theta\right) = \cos\left(2\frac{1}{2}\theta\right) + i\sin\left(2\frac{1}{2}\theta\right) = u,$$

luego $v^2 = u$. En todo caso conviene observar que si se usa para $\sqrt{z}$ la fórmula del teorema 2 nos basta una sencilla calculadora que dé raíces cuadradas mientras que lo recién expuesto nos obliga a calcular el arco coseno (excepto si es fácil intuir θ) de mayor dificultad.

Problema 5 *Si $u = \cos\theta + i\sin\theta$ probar que la potencia $n-$ésima es $u^n = \cos(n\theta) + i\sin(n\theta)$. Nota: se hace por inducción. Si el lector no sabe qué es esto puede dejar el problema para cuando se explique más adelante en el capítulo. Se suponen conocidas las fórmulas del coseno y seno suma.*

Teorema 3 *En un cuerpo arbitrario $\Bbbk$ se verifica para todo $a, b \in \Bbbk$,*

$$a \cdot 0 = 0, \quad a \cdot (-b) = -(ab), \quad \text{si } a \cdot b = 0, \text{ entonces } a = 0 \text{ o } b = 0.$$

Demostración. Se tiene $0 = 0 + 0$, luego $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$, restando $a \cdot 0$ a ambos lados queda $a \cdot 0 = 0$. Para lo segundo, $a \cdot (b + (-b)) = a \cdot 0 = 0$ como acabamos de ver. Entonces

$$0 = a \cdot (b + (-b)) = a \cdot b + a \cdot (-b)$$

y esto indica que $a \cdot (-b)$ es el inverso aditivo de $a \cdot b$. Finalmente, si $a \cdot b = 0$ y $a \neq 0$, existe a^{-1}. Ya sabemos que $a^{-1} \cdot 0 = 0$ luego $0 = a^{-1} \cdot (a \cdot b) = \left(a \cdot a^{-1}\right) \cdot b = 1 \cdot b = b$. ♣

Recordemos que se dijo que $a+(-b)$ se escribiría como $a-b$. De este teorema salen muchas fórmulas de uso continuo. Por conmutatividad, $0 \cdot a = 0$ y $(-b) \cdot a = -ba$. El siguiente problema le resultará incómodo al lector pues todo parece tan obvio que uno se sorprende de que le pidan demostrar esas cosas. Lo que se trata es de cerciorarse de que propiedades que se han usado sin vacilación en $\mathbb{R}$ son válidas en $\Bbbk$ arbitrario y "válida" quiere decir que se deriva de los axiomas de cuerpo y de lo que hasta ahora se ha probado. Por supuesto, una vez probadas, se olvida el camino seguido y se aplican tranquilamente.

Problema 6 *En un cuerpo arbitrario $\Bbbk$ se cumple que*

1. *El cero y la unidad son únicos (solo un elemento cumple las propiedades que los definen).*
2. *Los inversos aditivos y multiplicativos son únicos.*
3. *Se tiene la **ley de cancelación**: si $ab = ac$ o $ba = ca$ con $a \neq 0$ debe ser $b = c$.*
4. *Para todo a, $(-(-a)) = a$ y $\left(a^{-1}\right)^{-1} = a$ (en este segundo caso suponemos $a \neq 0$).*
5. *$(-a)(-b) = ab$ y $a(b - c) = ab - ac$.* ♦

Solución. Si hubiese dos unidades 1 y $1'$ se tendría $1 \overset{1}{=} 1 \cdot 1' \overset{2}{=} 1'$. Se usa en $\overset{1}{=}$ que como 1 es unidad, para todo a es $1 \cdot a = a$ y se toma en particular $a = 1'$. Para $\overset{2}{=}$ se usa que como $1'$ es unidad, se cumple para todo a que $a \cdot 1' = a$ y se toma $a = 1$.

En **4**, recordamos que v es el inverso multiplicativo de u si $uv = vu = 1$. Si $u = a^{-1}$ y nos piden probar $u^{-1} = a$ deberemos comprobar si $\left(a^{-1}\right) \cdot a = 1$ y $a \cdot \left(a^{-1}\right) = 1$, afirmaciones ciertas que prueban $\left(a^{-1}\right)^{-1} = a$.

Las restantes afirmaciones quedan para el lector. ♦

1.1.2. Clases de restos módulo n

¿Qué significa clasificar los elementos de un conjunto X, que supondremos no vacío? Debemos encontrar subconjuntos no vacíos $A_i \subset X$, con i recorriendo un conjunto de indices I, de modo que la unión de los A_i sea todo X y $A_i \cap A_j = \emptyset$ para $i \neq j$; o sea, los A_i diferentes son disjuntos. Los A_i se llamarán **clases** y cada $x \in X$ es miembro de una clase y solo una. Denotamos por $[x]$ a esa clase donde está x y el conjunto cuyos elementos son las clases es el **conjunto cociente.** Si X es un conjunto de lapiceros cuyos colores son los del parchís (amarillo, azul, rojo y verde) tenemos como I el conjunto de estos colores y A_i es el conjunto de los lapiceros de X de color I. Las Matemáticas requieren ejemplos más sofisticados. En general, se define para todos los pares de elementos x, y de X, una propiedad $P(x, y)$ y se dice que x está relacionado con y si $P(x, y)$ es cierta y se escribe como $x \sim y$. Si no ponemos ninguna restricción a P, al definir la clase de x como formada por los y relacionados con x, es posible que la unión de las clases no dé todo X o que las clases diferentes no sean disjuntas. Para tener una verdadera clasificación se necesita que P, o si se prefiere, que $\sim$ que indica cuando x e y están relacionados, verifique tres propiedades:

1. **Reflexividad**. Para todo $x \in X$ se tiene $x \sim x$.

2. **Simetría**. Si $x \sim y$ entonces $y \sim x$.

3. **Transitividad**. Si $x \sim y$ e $y \sim z$, entonces $x \sim z$.

Si la propiedad P o la relación $\sim$ cumple estas tres propiedades, se dice que P o $\sim$ es, una **relación de equivalencia**. En este caso, se define para cada $x \in X$ su clase $[x] = \{y \in X \mid x \sim y\}$, que es un subconjunto de X en donde está x (reflexividad). Si $x \in A_i$ se dice que x es un **representante** de la clase A_i, pudiendo suceder que siendo $x \neq x'$ se tenga $[x] = [x']$.

Una relación de equivalencia da una clasificación y recíprocamente. En efecto, si las A_i son las diferentes clases de $\sim$, tomamos cualquier $x \in X$ y al ser $x \sim x$ se sigue $x \in [x]$ que será una de las A_i. Vemos pues que la unión de las clases es todo X. Si con $[x] \neq [z]$ tuviésemos un $y \in [x] \cap [z]$ deduciríamos de $x \sim y$ y $z \sim y$ que $x \sim z$ (por transitividad) y $x \sim z$ nos da $[x] = [z]$, que es una contradicción. Dejamos para el lector probar que si los subconjuntos A_i con $i \in I$ son todas las clases de X, la relación $x \sim y$ si x e y están en un mismo A_i, da una relación de equivalencia

Nosotros vamos a estudiar aquí solo una relación de equivalencia. Tomaremos $X = \mathbb{Z}$, los números enteros, y fijaremos en adelante otro entero $n \geq 2$. Definiremos $P(x, y)$ como que n divida a $x - y$, así que $x \sim y$, que se representará con una notación más usual con $x \equiv y$, significa por definición que n divide a $x - y$ o, con otras palabras, que $x - y$ es múltiplo de n. A veces, para resaltar el papel de n, se escribe $x \equiv y \bmod n$. Es muy fácil comprobar que $\equiv$ es una relación de equivalencia, llamado **congruencia módulo** n. Si, por ejemplo, $n = 5$ se tiene $22 \equiv 7$ pues $(22 - 7) = 15 = 3 \cdot 5$, también $-2 \equiv 3$ porque $-2 - 3 = (-1) \cdot 5$ y $1000 \equiv 100 \equiv 0$ pues cualquier diferencia entre 1000, 100 y 0 es divisible por 5. Para $n = 2$ las clases son 2 que tradicionalmente se llaman la clase de los números pares y la de los números impares, pudiéndose escribir $[2] = [94] = [-12]$ o $[21] = [9]$ y $x \equiv y$ es decir que x e y tienen la misma paridad. Obsérvese que se pone $x \equiv y$, que no es una igualdad porque casi siempre $x \neq y$, pero sí se escribe $[x] = [y]$ porque aquí se está diciendo que dos *conjuntos* son iguales.

¿Cuántas clases tiene la relación de congruencia módulo n? Necesitamos una cuestión previa, el **teorema de la división euclidiana**, que no probaremos pero que el lector sabe utilizar.

Teorema 4 *Para cada $a \in \mathbb{Z}$, existen q y $r \in \mathbb{Z}$ tales que $a = qn + r$ y $0 \leq r < n$, siendo además q y r únicos y llamándose el **cociente** y el **resto** (de la división de a por n).*

Para entender el teorema ayuda *imaginar* los puntos de $\mathbb{Z}$ en una recta, marcando de modo especial los puntos qn con $q \in \mathbb{Z}$. Entonces $\mathbb{R}$ es unión *disjunta* de los subconjuntos

$$I_q = \{x \in \mathbb{Z} \mid qn \leq x < (q+1)\, n\} = \{qn, qn+1, qn+2, \ldots, qn+(n-1)\}.$$

Al ser disjunta la unión, cada $a \in \mathbb{Z}$ está en un único I_q y si $a = qn + r$, q es el cociente y r el resto. Decimos esto porque si $n = 5$ y $q = 13$ el lector dirá enseguida que $13 = 2 \cdot 5 + 3$ luego $q = 2$ y $r = 3$. Puede que diga igual de rápido para $a = -13$ que $-13 = (-2) \cdot 5 - 3$ luego $q = -2$ y $r = -3$, *pero esto es falso*. Si releemos el teoremos, vemos que se exige $0 \leq r < 5$ luego hay que descomponer $-13 = (-3) \cdot 5 + 2$, siendo $q = -3$ y $r = 2$. Viene esto a cuento de la pregunta ¿cuántas clases tiene la relación de congruencia módulo n? La respuesta es que hay n y son $[0], [1], [2], \ldots, [n-2], [n-1]$. No hay nada incorrecto al escribir $[-13]$ pero si se quiere escribir $[-13] = [r]$ con $0 \leq r < n = 5$ este r ha de ser el resto de la división, *con las condiciones exactas* del **teorema de la división euclidiana.**

Denotaremos por $\mathbb{Z}_n$ al conjunto de las clases módulo n, que es un conjunto con n elementos $[0], [1], [2], \ldots, [n-2], [n-1]$ y cada uno de ellos es un subconjunto de $\mathbb{Z}$. Si es, por ejemplo, $n = 5$, uno de los elementos de $\mathbb{Z}_5$ es el subconjunto de $\mathbb{Z}$ formado por todos los números que dejan resto 3 al ser divididos por 5. El que lo denotemos por $[3], [13]$ o $[-2]$ es una libre elección y esa libertad tiene consecuencias positivas. La principal es que *las clases de $\mathbb{Z}_n$ se pueden sumar y multiplicar como si fueran números*; digamos que $\mathbb{Z}_n$ es "casi" como un cuerpo. Las definiciones son

$$[a] + [b] = [a + b], \qquad [a] \cdot [b] = [a \cdot b].$$

Dicho con palabras, se toman dos clases, se eligen representantes a y b de ellas, y la clase suma o producto es la que tiene como representante la suma $a+b$ o el producto $a \cdot b$ de los representantes. Hay un problema y es que una clase tiene muchos representantes. Si es $[a] = [a']$ y $[b] = \left[b'\right]$, ¿podemos asegurar que $a + b$ y $a' + b'$ están en la misma clase y lo mismo para $a \cdot b$ y $a' \cdot b'$? La respuesta es que sí. Por ejemplo,

$$a \cdot b - a' \cdot b' = a \cdot b - a \cdot b' + a \cdot b' - a' \cdot b' = a \cdot (b - b') + (a - a') \cdot b'$$

y como n divide a cada sumando, divide a la suma. Es muy fácil coger práctica para operar en $\mathbb{Z}_n$ y descubrir *que es más fácil que hacerlo en* $\mathbb{Z}$, al menos para n pequeño. Por ejemplo, en $\mathbb{Z}_5$,

$$([12]+[11])^2 = ([12]+[11])\cdot([12]+[11]) = [23]\cdot[23] = [3]\cdot[3] = [9] = [4]\,,$$

$$[101]+[52] = [153] = [3]\,, \qquad [101]+[52] = [1]+[2] = [3]\,.$$

De modo rutinario se comprueba que $\mathbb{Z}_n$ es un "casi" cuerpo con estas operaciones pero puede pasa que exista $[a]\neq 0$ tal que no exista $[b]\neq 0$ de modo que $[a]\cdot[b]=1$, así que habría elementos sin inverso multiplicativo. Por ejemplo, para [2] tenemos en $\mathbb{Z}_4$,

$$[2]\cdot[0]=[0]\,, \qquad [2]\cdot[1]=[2]\,, \qquad [2]\cdot[2]=[0]\,, \qquad [2]\cdot[3]=[2]\,,$$

y nunca aparece $[2]\cdot[a]=[1]$ para ningún $[a]\in\mathbb{Z}_4$. Si pasamos a $\mathbb{Z}_6$ vemos que a veces hay inverso (por ejemplo $[5]\cdot[5]=[25]=[1]$ luego $[5]^{-1}=[5]$) pero no existe $[2]^{-1}$. Sintetizamos lo obtenido en el siguiente teorema.

Teorema 5 *Las clases de restos $\mathbb{Z}_n$ con la suma y producto definidos cumplen todos las propiedades que tienen la suma y el producto en un cuerpo tomando $0=[0]$ y $1=[1]$, excepto la existencia de inverso multiplicativo cuando $[a]\neq 0$. Es necesario y suficiente para que todos los elementos no nulos de $\mathbb{Z}_n$ tengan inverso multiplicativo que n sea un número primo.*

Admitiremos sin demostración el último aserto. *Sabemos por tanto que $\mathbb{Z}_2,\mathbb{Z}_3,\mathbb{Z}_5,\mathbb{Z}_7,\ldots$ son cuerpos.* Si queremos conocer $[a]^{-1}$ no queda más remedio que hacerlo por tanteo, cosa fácil en $\mathbb{Z}_5$ (pues $[2]^{-1}=[3]$) pero que cuesta determinar en $\mathbb{Z}_{31}$ (¿cuánto vale $[17]^{-1}$?)[3]. Escribiremos $\mathbb{Z}_p$ en vez de $\mathbb{Z}_n$ para recordar que $n=p$ debe ser primo. Por cierto, $p=2$ es primo y $\mathbb{Z}_2$ es el cuerpo más simple en cuanto a su definición. *El cuerpo $\Bbbk=\mathbb{Z}_p$ aparecerá casi siempre en ejemplos con $p\leq 7$ para probar aspectos curiosos o poner a prueba la claridad de conceptos. Puede considerarse material adicional para los lectores más ambiciosos, pero advertimos que tales cuerpos aparecen muchísimo en Matemáticas y que es bueno ir estableciendo un contacto con ellos. No deben dejarse sistemáticamente de lado como una "curiosidad".*

Hemos sido cuidadosos en el desarrollo de la teoría distinguiendo entre $a\in\mathbb{Z}$ y $[a]\in\mathbb{Z}_p$ pero si se trata de echar cuentas conviene prescindir de los corchetes [] y, una vez fijado p, acostumbrarse a escribir cosas como $2^2=1$ (si $p=3$) o $2^2=-1$ (si $p=5$) por mucho que al principio suenen raras. No hay problema si p está claro. Por ejemplo, si nos piden resolver la ecuación de primer grado $2x+4=0$ en $\mathbb{Z}_5$ damos los pasos

$$2x+4=0, \qquad 2x=-4=1, \qquad x=2^{-1}=3.$$

Efectivamente $2\cdot 3+4=10=0$. El lector observará que, al saber que $\mathbb{Z}_p$ es un cuerpo, todas las propiedades abstractas y aparentemente inútiles vistas al final de la sección anterior, aparecen como muy útiles. Si sé que una propiedad tipo "$xy=0$ implica $x=0$ o $y=0$" (con la que estoy familiarizado en $\mathbb{R}$) deriva de los nueve axiomas de cuerpo, puedo asegurar que también valdrá en $\mathbb{Z}_p$.

Problema 7 *Resolver en $\mathbb{Z}_5$ el sistema de ecuaciones*

$$\begin{cases} x+2y=1 \\ 3x+4y=2 \end{cases}.$$

En cuanto haga algunas cuentas el lector observará que a veces compensa cambiar primero los representantes y luego operar y otras veces se opera y luego se cambian los representantes. Por ejemplo, en $\mathbb{Z}_5$,

$$[101]^4=[1]^4=\left[1^4\right]=[1]\,, \quad [101]^4=\left[(101)^4\right]=[104\,060\,401]=[1]$$

y más vale seguir el primer camino que el segundo.

Dijimos hace bastante tiempo que íbamos a definir el concepto de **anillo**. La definición es análoga a la de cuerpo pero con axiomas menos restrictivos. Definimos un **anillo unitario** como un conjunto $\mathbb{A}$ donde para cada par de elementos a,b se definen otros dos elementos denotados por $a+b$ y $a\cdot b$, llamados la suma y producto de a y b. Estas dos asignaciones u operaciones en $\mathbb{A}$ cumplen las siguientes propiedades

[3] La herramienta esencial es el **teorema de Bezout** que el lector verá en cursos de Álgebra.

1. *Asociatividad de la suma.* Para todo $a, b, c \in \mathbb{A}$ se tiene $a + (b + c) = (a + b) + c$.

2. *Conmutatividad de la suma.* Para todo $a, b \in \mathbb{A}$ se tiene $a + b = b + a$.

3. *Existencia de cero.* Existe un elemento, que denotaremos por 0, tal que para todo $a \in \mathbb{A}$ se tiene $a + 0 = 0 + a = a$.

4. *Existencia de opuesto o inverso aditivo.* Para todo $a \in \mathbb{A}$ existe un elemento que denotaremos por $-a$ tal que $a + (-a) = (-a) + a = 0$.

5. *Asociatividad del producto.* Para todo $a, b, c \in \mathbb{A}$ se tiene $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

6. *Existencia de unidad.* Existe un elemento, que denotaremos por 1 y es distinto de 0, tal que para todo $a \in \Bbbk$ se tiene $a \cdot 1 = 1 \cdot a = a$.

7. *Distributividad.* Para todo $a, b, c \in \mathbb{A}$ se tiene $a \cdot (b + c) = a \cdot b + a \cdot c$ y $(a + b) \cdot c = a \cdot c + b \cdot c$.

Puede verse que hemos copiado literalmente los axiomas de cuerpo pero quitando la conmutatividad del producto, y la existencia de inversos multiplicativos. Hablamos de anillos *unitarios* porque la definición general de anillo no exige la existencia de 1, pero todos los que vamos a ver lo son. Si el anillo tiene un producto conmutativo se llama **anillo conmutativo**. Por supuesto, todo cuerpo es un anillo unitario y conmutativo y $\mathbb{Z}_n$ también lo es aunque n no sea primo. Los ejemplos más importantes de anillos unitarios y conmutativos son $\mathbb{Z}$, los números enteros, y los polinomios con coeficientes en un cuerpo $\Bbbk$, que se denotan por $\Bbbk[X]$. Incluso $\mathbb{Z}[X]$ es un ejemplo de anillo unitario y conmutativo aunque $\mathbb{Z}$ no es un cuerpo. Pronto veremos que las matrices $n \times n$ con su suma y producto forman un anillo unitario pero no conmutativo y son el ejemplo principal de este tipo de anillos. (Por supuesto, no suponemos en este punto un dominio de las matrices.)

Advertimos hace tiempo que los cuerpos tienen ciertas propiedades que no tienen $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$. He aquí una lista de propiedades vista sin duda como ciertas en $\mathbb{R}$, incluso a veces en $\mathbb{Q}$, pero falsas o sin sentido en cuerpos arbitrarios:

1. En $\mathbb{Q}$ y $\mathbb{R}$ hay una relación de orden $\leq$ y por tanto está el concepto de positivo o negativo. Sin embargo $\mathbb{C}$ no tiene la posibilidad de ser ordenado de modo "útil". Otro tanto se puede decir de $\mathbb{Z}_p$ y en general de los cuerpos: no tienen definidas relaciones de orden.

2. Una suma de cuadrados nula $(a_1)^2 + \ldots + (a_k)^2 = 0$ implica $a_1 = \ldots = a_n = 0$ en $\mathbb{Q}$ y $\mathbb{R}$. Hemos visto que esto es falso en $\mathbb{C}$. En $\mathbb{Z}_5$, $[1]^2 + [2]^2 = [1] + [4] = [5] = [0]$ pero $[1]$ y $[2]$ son no nulos.

3. Hay cuerpos finitos, siendo $\mathbb{Z}_p$ un ejemplo.

4. La posibilidad de que haya raíces cuadradas es muy variable según el cuerpo. A veces hay criterios sencillos como en $\mathbb{C}$ (todo elemento tiene raíz cuadrada) o en $\mathbb{R}$ (debe ser $a \geq 0$). En otros cuerpos es más difícil y en general no se puede garantizar su existencia.

Hay un punto final que tiene cierta sutileza. Si se le pide al lector que desarrolle $(a + b)^2$ hará un cálculo más o menos como este

$$(a + b)^2 = (a + b)(a + b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2.$$

La pregunta es ¿qué es 2 si $\Bbbk$ es arbitrario? La respuesta es $2 = 1 + 1$. La última igualdad del cálculo previo es

$$a^2 + ab + ba + b^2 = a^2 + ab + ab + b^2 = a^2 + b^2 + (1 + 1)ab = a^2 + b^2 + 2ab$$

Para *cualquier* $\Bbbk$ está definido $1 + 1 + \ldots + 1$ (n veces) que se escribe como n. Entonces ¿podemos decir que $\mathbb{N} \subset \Bbbk$? No, porque puede ser $1 + 1 + \ldots + 1 = 0$. (En $\Bbbk$ el producto de n números no nulos es no nulo, pero la suma de n números no nulos bien puede ser nula.) En efecto, en $\mathbb{Z}_p$, $[1] + [1] + \ldots + [1]$ (p veces) es $[p] = [0]$. En $\mathbb{Z}_2$ se tiene $2 = 0$ luego $(a + b)^2 = a^2 + b^2$ (el cuadrado de la suma es la suma de los cuadrados) y en $\mathbb{Z}_3$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 = a^3 + b^3.$$

Si el lector conoce el binomio de Newton, en $\mathbb{Z}_p$, $(a + b)^p = a^p + b^p$. En un cuerpo abstracto $\Bbbk$ se dice que es de **característica cero** si nunca puede ser $1 + 1 + 1 + \cdots + 1 = 0$ y se dice que tiene **característica**

p si es posible que sea $1+1+1+\cdots+1=0$ (n veces) y p es el número más pequeño con el que esto se consigue. Usamos p porque se puede probar que tal número debe ser primo. Claramente $\mathbb{Q},\mathbb{R}$ o $\mathbb{C}$ tienen característica cero y $\mathbb{Z}_p$ tiene característica p. Ser de característica p simplifica muchos cálculos o los complica porque ciertas operaciones no son válidas. Si por ejemplo es $p=2$, no se puede dividir por $2a$ aunque sea $a\neq 0$ (en realidad $2=0=2a$). Esto crea problemas porque la famosa fórmula de las raíces de la ecuación de segundo grado no vale en este caso. Pronto lo veremos.

1.1.3. La ecuación de segundo grado

Ya hemos visto la definición de **raíz cuadrada** de $a\in\Bbbk$ en $\Bbbk$, que es un $r\in\Bbbk$ tal que $r^2=a$. No hay garantía de que r exista, pero si existe hay otra raíz cuadrada s, esta debe ser r (problema 3). Dijimos que en $\Bbbk=\mathbb{R}$ se denota por $\sqrt{a}$ a la raíz ≥ 0 pero si estamos en un $\Bbbk$ arbitrario, $\sqrt{a}$ será una de las dos raíces cuadradas (si existen). ¿Cuál? En $\mathbb{Z}_5$ tenemos $3^2=2^2=4$ y no parece que haya nada que haga más "distinguida" una raíz que otra. Afortunadamente, vamos a una fórmula donde aparecerá $\pm\sqrt{\ }$ luego la elección que hagamos no tendrá consecuencias.

Fijemos un cuerpo $\Bbbk$ arbitrario *con característica distinta de* 2, luego $1+1\neq 0$ y se puede dividir por $2a$ cuando $a\neq 0$. Ya advertimos que $na=(1+\cdots+1)\,a=a+\cdots+a$ con n sumandos. Así pues, en cualquier cuerpo $2a=a+a$. Decimos esto porque pronto usaremos que $(2a)^2=4a^2$. Esto se sigue de

$$(2a)^2=(a+a)^2=a^2+a^2+a^2+a^2=4a^2.$$

Consideremos la ecuación de segundo grado $ax^2+bx+c=0$ con $a\neq 0$. Nos preguntamos si existen raíces de la ecuación. Aquí **raíz** es sinónimo de "solución"; es decir, un $r\in\Bbbk$ tal que al sustituir x por r sale cero. (¡No confundir conceptos!). Piense el lector al principio que $\Bbbk=\mathbb{R}$ si está más a gusto.

Teorema 6 *Consideremos la ecuación de segundo grado $ax^2+bx+c=0$ con $a,b,c\in\Bbbk$ y $a\neq 0$ en un cuerpo $\Bbbk$ con característica $\neq 2$. Sea $D=b^2-4ac$, que llamamos el **discriminante**. Es necesario y suficiente para que la ecuación tenga raíces el que D tenga una raíz cuadrada. Si representamos por $\sqrt{D}$ a una de ellas, las raíces de la ecuación son*

$$\frac{-b\pm\sqrt{D}}{2a}=\frac{-b\pm\sqrt{b^2-4ac}}{2a}.$$

Hay pues dos raíces distintas si $D\neq 0$ y tiene raíz cuadrada, y una sola si $D=0$.

Demostración. Se hace el **cambio de variable** $x=y-(b/2a)$.[4] Tenemos equivalencias

$$\begin{aligned}ax^2+bx+c=0 &\iff a\left(y-\frac{b}{2a}\right)^2+b\left(y-\frac{b}{2a}\right)+c=0\\ &\iff a\left(y^2+\frac{b^2}{4a^2}-\frac{by}{a}\right)+\left(by-\frac{b^2}{2a}\right)+c=0 \iff ay^2+\frac{b^2}{4a}-\frac{b^2}{2a}+c=0\\ &\iff ay^2=\frac{b^2}{4a}-c=\frac{b^2-4ac}{4a} \iff y^2=\frac{b^2-4ac}{4a^2}=\frac{D}{4a^2}.\end{aligned}$$

Se ha usado la notación u/v en lugar de uv^{-1} y $\iff$ significa que cada ecuación equivale a la siguiente (que son ciertas o falsas a la vez). La cadena de equivalencias nos dice que r es raíz de $ax^2+bx+c=0$ si y solo si t es raíz de $y^2=D/4a^2$ con la relación $r=t-(b/2a)$. Las posibles raíces de $y^2=D/4a^2$ son (ver comentario previo al teorema)

$$t=\pm\sqrt{\frac{D}{4a^2}}=\pm\frac{\sqrt{D}}{2a}$$

en el caso de que exista una raíz $\sqrt{D}$ de D. Por tanto, deshaciendo el cambio de variable obtenemos las raíces anunciadas. ♣

Problema 8 *Tomamos $\Bbbk=\mathbb{C}$ y $p,q\in\mathbb{R}$ con $q>0$. Probar que las raíces de $(X-p)^2+q^2=0$ son $p\pm qi$.*

Problema 9 *Calcular las raíces de $X^2+iX+1=0$ y $iX^2+X+i=0$. ¿Son iguales?*

[4] Si la característica es 2 el cambio de variable es imposible.

Problema 10 *Probar que en $\mathbb{Z}_3$ la ecuación $x^2+bx+\left(b^2-2\right)=0$ no tiene solución para ningún b pero $x^2+bx+\left(b^2-1\right)=0$ la tiene para todo b. ¿Son distintas las raíces en este caso?*

Problema 11 *En $\mathbb{Z}_5$ dar un ejemplo de ecuación de segundo grado $X^2+\alpha X+\beta$ con α,β ambos no nulos que no tenga raíces.*

Los números complejos no están ordenados, aunque $\mathbb{R}\subset\mathbb{C}$ sí lo está. Se definió $\mathbb{C}$ con la idea de poder resolver ecuaciones arbitrarias de segundo grado, y lo hemos conseguido a la vista del teorema 6. Sin embargo los complejos ofrecen mucho más de lo que podíamos imaginar.

Teorema 7 (Fundamental del Álgebra) *Sean $a_0,a_1,\ldots,a_n$ números complejos, con $a_n\neq 0$ y $n\in\mathbb{N}$. La ecuación $a_0+a_1x+a_2x^2+\ldots+a_nx^n=0$ tiene siempre solución (raíz) en $\mathbb{C}$.*

No vamos a demostrar el teorema, aunque sí lo usaremos. La demostración no da una fórmula para la solución como pasa con $n=2$, e inicialmente es poco satisfactoria, porque se hace por reducción al absurdo. Si una ecuación como la descrita no tuviera solución se llegaría a contradicción. En todo caso, hay algoritmos para aproximar una solución tanto como se quiera (léase, tanto como se esté dispuesto a trabajar). Cualquiera de las muchas demostraciones usa la llamada **completitud de** $\mathbb{R}$. Digamos para una descripción muy tosca, que $\mathbb{R}$ es un cuerpo con una relación de orden pero "tiene más", ya que $\mathbb{Q}$ es también un cuerpo ordenado. Lo que distingue $\mathbb{Q}$ de $\mathbb{R}$ es lo que llamamos completitud y son esas propiedades difíciles que hemos advertido al lector que no iba a necesitar de momento.

1.2. Una notación esencial: los sumatorios

El lector ha tratado funciones, casi siempre numéricas y dadas por una fórmula; por ejemplo, $f:\mathbb{R}\to\mathbb{R}$ y $f(x)=x\left(x^2+1\right)$. Son igualmente importantes las funciones $f:\{1,2,\ldots,n\}\to A$, siendo A un conjunto arbitrario y $n\in\mathbb{N}$. Se llaman **sucesiones** y se representan muchas veces en la forma $(x_1,\ldots,x_n)$, indicando esta notación que a 1 le corresponde $x_1\in A$, a 2 le corresponde x_2, ... y a n le corresponde x_n. Si $n=3$ y A es el conjunto de todos los números primos, $(5,5,11)$ es un ejemplo de sucesión. Obsérvese que $x_1=x_2=5$ está permitido por la definición y que, en general, la sucesión no tiene por qué ser inyectiva ni suprayectiva. Aclaramos todo esto porque si se trata un conjunto A con n elementos, es necesario o conveniente escribir los *distintos* elementos de A como $a_1,a_2,\ldots,a_n$. Lo que se hace es elegir una sucesión $(a_1,\ldots,a_n)$; es decir, una *función*, de modo que sea una *biyección* entre $\{1,2,\ldots,n\}$ y A. Con esta función, con esta sucesión, numeramos u ordenamos los elementos de A. Hay múltiples maneras de elegir esta sucesión, de hecho hay $n!$ sucesiones biyectivas de $\{1,2,\ldots,n\}$ en A cuando A tiene n elementos. En principio $A=\{a_1,\ldots,a_n\}$ y $(a_1,\ldots,a_n)$ son entes distintos pues uno es un *conjunto* y otro una *función,* pero sucede a menudo que lo que se va a decir sobre A no depende de la manera en que se hayan numerado u ordenado sus elementos. Esto hace que aparezca la notación $A=\{a_1,\ldots,a_n\}$ que, estrictamente hablando es incorrecta, pues a la derecha tenemos no solo cuáles son los elementos de A sino cómo se han numerado.

Este tipo de exceso en la notación es admisible cuando A es un conjunto de números y lo que se hace con los elementos de A es sumarlos o multiplicarlos. Es técnicamente difícil de formalizar, pero admitiremos que el resultado final no depende de si se han agrupado los números en operaciones intermedias o el orden de ejecución de las operaciones. Se pone por tanto la suma o producto como $a_1+a_2+\ldots+a_n$ o $a_1a_2\cdots a_n$.

El lector debe acostumbrarse a la notación

$$a_1+a_2+\ldots+a_{n-1}+a_n=\sum_{i=1}^{n}a_i,\qquad a_1\cdot a_2\cdot\ldots\cdot a_{n-1}\cdot a_n=\prod_{i=1}^{n}a_i,$$

con símbolos derivados de las letras griegas mayúsculas sigma y pi.[5] Este tipo de abreviatura es absolutamente necesario para escribir razonamientos y fórmulas generales. Podemos sustituir i por cualquier

[5] Usaremos más las sumas que los productos y será en esas fórmulas donde nos vamos a centrar.

otra letra sin que varíe el significado.[6] Por ejemplo,

$$a_1 + a_2 + \ldots + a_{n-1} + a_n = \sum_{i=1}^{n} a_i = \sum_{p=1}^{n} a_p = \sum_{H=1}^{n} a_H.$$

Hay veces que los índices no van desde 1 hasta n sino que se mueven en otro conjunto. Por ejemplo,

$$\sum_{i=-n}^{n} a_i = a_{-n}+a_{-(n-1)}+\ldots+a_{-1}+a_0+a_1+a_2+\ldots+a_{n-1}+a_n, \qquad \sum_{i=0}^{n} a_i = a_0+a_1+a_2+\ldots+a_{n-1}+a_n.$$

Más generalmente, podemos suponer que A es un conjunto *finito* de elementos de $\Bbbk$ y $\sum_{a\in A} a$ o $\prod_{a\in A} a$ representará la suma o producto de los elementos de A. Insistimos en que se da por cierto que las operaciones pueden hacerse de cualquier modo sin que varíe el resultado final. Nos vamos a centrar en sumas. Supongamos que los elementos de A están en una tabla de m filas y n columnas. Digamos que la tabla es

$$\begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \qquad \text{y un ejemplo,} \qquad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

Si estudiamos primero el ejemplo, la suma S de sus elementos se puede calcular de varias maneras

$$(1+4)+(2+5)+(3+6), \quad (1+2+3)+(4+5+6), \quad 3+5+2+6+1+4.$$

Lo que queremos resaltar es que en la primera suma lo hacemos por columnas, en la segunda por filas y en la tercera no nos imponemos un orden definido de operaciones. Esto, expresado con generalidad, es que la suma de elementos de la tabla A se puede expresar de tres modos iguales

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} = \sum_{\substack{1\le i\le m\\ 1\le j\le n}} a_{ij}. \tag{1.2}$$

Cuesta ver un poco que $\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}$ representa la suma de la tabla por filas, pero así es. Con palabras todo está más claro. Elegimos i en $\{1,\ldots,m\}$ y, variando j en $\{1,\ldots,n\}$ calculamos $F_i = \sum_{j=1}^{n} a_{ij}$ que es la suma de los términos de la fila i. Entonces, $F_1+\ldots+F_m = \sum_{i=1}^{m} F_i = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}$ es la suma de todos los números de la tabla por filas. El caso $\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}$ tiene una descripción análoga, y el último no informa sobre cómo se hacen las operaciones.

Hay un caso particular de 1.2 muy frecuente. Tenemos sucesiones $(x_1,\ldots,x_m)$ e $(y_1,\ldots,y_n)$ en $\Bbbk$ y definimos $a_{ij} = x_i y_j$. Entonces

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j = \sum_{j=1}^{n}\sum_{i=1}^{m} x_i y_j = \sum_{\substack{1\le i\le m\\ 1\le j\le n}} x_i y_j = \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{j=1}^{n} y_j\right). \tag{1.3}$$

El cuarto elemento, el nuevo, se obtiene por distributividad (vulgo, sacando factor común) como se indica

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j &= \sum_{i=1}^{m}\left(x_i y_1 + x_i y_2 + \ldots + x_i y_n\right) = \sum_{i=1}^{m} x_i\left(y_1 + y_2 + \ldots + y_n\right) \\ &= x_1\left(y_1 + y_2 + \ldots + y_n\right) + x_2\left(y_1 + y_2 + \ldots + y_n\right) + \ldots + x_m\left(y_1 + y_2 + \ldots + y_n\right) \\ &= \left(x_1 + \ldots + x_m\right)\left(y_1 + y_2 + \ldots + y_n\right) = \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{j=1}^{n} y_j\right).\end{aligned}$$

[6] El lector sabe que $\int x^2 dx = \int y^2 dy$ sin que la elección de la letra de la variable altere el significado. Estas variables, o índices en el caso de sumatorios, se llaman **variables** o **índices mudos** (en inglés, *dummy indexes*). La mejor traducción sería variable o índice *ficticio*. pues no representan una variable sino una operación que se hace con la variable.

Insistimos en que hay que irse familiarizando con estas fórmulas porque son de uso continuo en el Cálculo Matricial y Cálculo Tensorial. No hay debajo ningún teorema profundo, sino tan solo una necesidad de expresar fórmulas con sumas y productos con un número finito pero indeterminado de elementos que en ellas van a aparecer.

Son más fáciles fórmulas como

$$\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} (a_i + b_i)\,, \qquad \lambda \left(\sum_{i=1}^{n} a_i \right) = \sum_{i=1}^{n} \lambda a_i. \tag{1.4}$$

La segunda generaliza el axioma de distributividad y la primera se demuestra con

$$\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i = (a_1 + a_2 + \ldots + a_n) + (b_1 + b_2 + \ldots + b_n) = a_1 + b_1 + a_2 + b_2 \ldots + a_n + b_n = \sum_{i=1}^{n} (a_i + b_i)\,.$$

Problema 12 *Sean $a_1, \ldots, a_n$ elementos de $\Bbbk$, cuerpo arbitrario. Probar la fórmula*

$$\begin{aligned}(a_1 + \cdots + a_n)^2 &= \left(\sum_{i=1}^{n} a_i \right)^2 \\ &= (a_1)^2 + \cdots + (a_n)^2 + 2\left(a_1a_2 + \cdots + a_1a_n + a_2a_3 + \cdots + a_2a_n + \cdots + a_{n-1}a_2\right) \\ &= \sum_{i=1}^{n} (a_i)^2 + 2 \sum_{1 \le i < j \le n} a_i a_j.\end{aligned}$$

Como caso particular, si $\Bbbk = \mathbb{Z}_2$ se obtiene que "el cuadrado de la suma es la suma de los cuadrados" pues

$$\left(\sum_{i=1}^{n} a_i \right)^2 = (a_1)^2 + \cdots + (a_n)^2 = \sum_{i=1}^{n} (a_i)^2\,. \blacklozenge$$

Solución. Como caso particular de (1.3),

$$\begin{aligned}\left(\sum_{i=1}^{n} a_i \right)^2 &= \left(\sum_{i=1}^{n} a_i \right) \left(\sum_{i=1}^{n} a_i \right) = \sum_{i,j=1}^{n} a_i a_j \\ &= \sum_{h=1}^{n} (a_h)^2 + \sum_{i \ne j} a_i a_j = \sum_{h=1}^{n} (a_h)^2 + \sum_{1 \le i < j \le m} (a_i a_j + a_j a_i) \\ &= \sum_{h=1}^{n} (a_h)^2 + (1+1) \sum_{1 \le i < j \le m} a_i a_j = \sum_{h=1}^{n} (a_h)^2 + 2 \sum_{1 \le i < j \le m} a_i a_j.\end{aligned}$$

El caso de $\mathbb{Z}_2$ resulta de ser $1 + 1 = 0$. $\blacklozenge$

1.3. Matrices

1.3.1. Conceptos generales

En adelante $\Bbbk$ denotará un cuerpo arbitrario, aunque en una primera lectura se puede suponer que $\Bbbk = \mathbb{R}$. Una **matriz con coeficientes en $\Bbbk$ con m filas y n columnas** es una tabla de $m \times n$ números a_{ij} con m filas y n columnas, denotando i la fila y j la columna donde se halla a_{ij}. El conjunto de estas matrices se representa por $\Bbbk^{m \times n}$ y los a_{ij} son los **coeficientes** o **términos de la matriz**.[7] Las matrices se llaman **reales o complejas** si $\Bbbk$ es $\mathbb{R}$ o $\mathbb{C}$ respectivamente. En los ejemplos

$$a = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}, \qquad b = \begin{pmatrix} x & 0 & 0 & -x \\ -x & 1 & 1 & x \end{pmatrix},$$

[7] En inglés se usa sobre todo "entry", con muchos significados, pero que aquí corresponde al "apunte contable" en un libro de contabilidad.

tenemos para a que $m = 3$ y $n = 2$ y para b que $m = 2$ y $n = 4$. Además $a_{31} = 5$, $a_{22} = 4$, $b_{12} = b_{13} = 0$ y $b_{24} = x$. Hemos empezado dando la definición con la posición más usual de los índices, ambos como subíndices, en la forma a_{ij}. Sin embargo tiene sus ventajas la notación a^i_j con i índice de fila y j índice de columna. Con esta notación $a^3_1 = 5$, $a^2_2 = 4$, $b^1_2 = b^1_3 = 0$ y $b^2_4 = x$. Damos para a la fila i, la columna j, y toda la matriz

$$a^i = \left(a^i_1, \ldots, a^i_j, \ldots, a^i_n\right), \qquad a_j = \begin{pmatrix} a^1_j \\ \vdots \\ a^i_j \\ \vdots \\ a^m_j \end{pmatrix}, \qquad a = \begin{pmatrix} a^1_1 & \cdots & a^1_j & \cdots & a^1_n \\ \vdots & & \vdots & & \vdots \\ a^i_1 & \cdots & a^i_j & \cdots & a^i_n \\ \vdots & & \vdots & & \vdots \\ a^m_1 & \cdots & a^m_j & \cdots & a^m_n \end{pmatrix}$$

Como se ve, a^i es la fila i y a_j la columna j de $a \in \Bbbk^{m\times n}$. En un ejemplo,

$$a = \begin{pmatrix} 1 & 0 & h & -h \\ 2 & 0 & 0 & -1 \\ 3 & -3 & 1 & h \end{pmatrix}, \qquad a^3 = (3, -3, 1, h), \qquad a_2 = \begin{pmatrix} 0 \\ 0 \\ -3 \end{pmatrix}.$$

La **traspuesta** de a es la matriz b cuyas filas son las columnas de a; es decir $b^j_i = a^i_j$. Suele decirse que las filas de a se convierten en la columnas de b; más concretamente, la fila i de a es la columna i de b. La traspuesta de a se denotará por $a^\top$ y si $a \in \Bbbk^{m\times n}$ entonces $a^\top \in \Bbbk^{n\times m}$. La matrices a y b de los primeros ejemplos tienen como traspuestas

$$a^\top = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \qquad b^\top = \begin{pmatrix} x & -x \\ 0 & 1 \\ 0 & 1 \\ -x & x \end{pmatrix}.$$

Las matrices se llaman **cuadradas** si $m = n$. Una **matriz fila** es la que solo tiene una fila (o sea, $m = 1$) y una **matriz columna** la que solo tiene una columna (o sea, $n = 1$). Se suele prescindir del subíndice o superíndice ya que siempre será 1. El lector estará acostumbrado a representar los puntos del plano o del espacio como matrices fila, digamos $(1, 3)$ o $(1, -1, 0)$. Sin embargo, por razones técnicas, al representar transformaciones del plano y del espacio con matrices, es más conveniente que estos puntos sean *vectores columna* y representarlos a su vez como $(1, 3)^\top$ o $(1, -1, 0)^\top$. La matriz de $\Bbbk^{m\times n}$ con todos los coeficientes cero es la **matriz cero** y se denota por $0_{m\times n}$. La matriz *cuadrada* $m \times m$ con todo ceros excepto $a^1_1 = a^2_2 = \ldots = a^m_m = 1$ se llama la **matriz unidad** y se denota por I_m, prescindiéndose de m si se sobrentiende. Debíamos denotar los coeficientes de I por I^p_q pero una larga tradición pone δ^p_q en vez de I^p_q. De modo más formal, las **deltas de Kronecker** se definen por $\delta^i_j = 1, 0$ según sea $i = j$ o $i \neq j$ para $1 \leq i, j \leq m$. Se generaliza la definición a δ_{ij} o δ^{ij}, siempre con los valores 1 o 0 según sea $i = j$ o $i \neq j$. Las deltas de Kronecker son muy importantes y aparecerán con frecuencia.

Fijamos en adelante $m, n \in \mathbb{N}$ y trabajamos con matrices en $\Bbbk^{m\times n}$. Estas matrices se pueden sumar y multiplicar por elementos de $\Bbbk$, los llamados **escalares.** Para $a, b \in \Bbbk^{m\times n}$ y $\lambda \in \Bbbk$ definimos la **suma** $a + b$ y el **producto por escalares** λa dando su coeficiente general en el lugar (i, j) en la forma

$$(a + b)^i_j = a^i_j + b^i_j, \qquad (\lambda a)^i_j = \lambda a^i_j, \qquad 1 \leq i \leq m,\ 1 \leq j \leq n.$$

A veces se pondrá $\lambda \cdot a$ en vez de λa por claridad o por énfasis. Es muy fácil dar ejemplos,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ 2 & -2 \\ 3 & -3 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 5 & 2 \\ 8 & 3 \end{pmatrix}, \qquad 7\begin{pmatrix} x & 0 & 0 & -x \\ -x & 1 & 1 & x \end{pmatrix} = \begin{pmatrix} 7x & 0 & 0 & -7x \\ -7x & 7 & 7 & 7x \end{pmatrix}.$$

Teorema 8 *La suma y el producto por escalares tienen las siguientes propiedades*

1. Asociatividad de la suma. *Para todo* $a, b, c \in \Bbbk^{m\times n}$ *se tiene* $a + (b + c) = (a + b) + c$.
2. Conmutatividad de la suma. *Para todo* $a, b \in \Bbbk^{m\times n}$ *se tiene* $a + b = b + a$.
3. Existencia de cero. *Existe un elemento, que denotaremos por* 0, *tal que para todo* $a \in \Bbbk^{m\times n}$ *se tiene* $a + 0 = 0 + a = a$.

4. Existencia de opuesto o inverso aditivo. *Para todo* $a \in \Bbbk$ *existe un elemento que denotaremos por* $-a$ *tal que* $a + (-a) = (-a) + a = 0$.

5. *Para todo* $\lambda, \mu \in \Bbbk$ *y* $a \in \Bbbk^{m\times n}$ *se tiene* $(\lambda\mu)\, a = \lambda\,(\mu a)$

6. *El elemento unidad* 1 *de* $\Bbbk$ *verifica para todo* $a \in \Bbbk^{m\times n}$ *que* $1 \cdot a = a$.

7. Doble distributividad. *Para todo* $\lambda, \mu \in \Bbbk$ *y* $a, b \in \Bbbk^{m\times n}$ *se tiene* $(\lambda + \mu)\, a = \lambda a + \mu a$ *y* $\lambda\,(a + b) = \lambda a + \lambda b$.

El cero en ***3*** *es* $0_{m\times n}$ *y* $-a$ *en* ***4*** *es la matriz con coeficientes* $-a_j^i$; *o sea,* $(-a)_j^i = -a_j^i$.

El lector puede ver este teorema como un mero recordatorio de propiedades obvias. Desde luego son muy fáciles de probar, pero no hay que perder el sentido de lo que indican. Por ejemplo, $(\lambda\mu)\, a = \lambda\,(\mu a)$ indica que las operaciones se pueden hacer en orden diferente pero llegando al mismo resultado, teniendo en cuenta que el paréntesis indica prioridad en la operación. La matriz $(\lambda\mu)\, a$ se obtiene *multiplicando primero* λ y μ, lo que da otro número $\lambda\mu$, y luego multiplicando $\lambda\mu$ por a. Por otra parte, $\lambda\,(\mu a)$ se obtiene *multiplicando primero* μa y luego, multiplicando μa por el escalar λ. Para comprobar que $(\lambda\mu)\, a = \lambda\,(\mu a)$ debemos observar que $(\lambda\mu)\, a$ y $\lambda\,(\mu a)$ son matrices $m \times n$ luego debe probarse que los coeficientes son iguales. Calculamos,

$$((\lambda\mu)\, a)_j^i = (\lambda\mu)\, a_j^i, \qquad (\lambda\,(\mu a))_j^i = \lambda\,(\mu a)_j^i = \lambda\left(\mu a_j^i\right).$$

Efectivamente, $(\lambda\mu)\, a_j^i = \lambda\left(\mu a_j^i\right)$ pues λ, μ y a_j^i están en $\Bbbk$ y ahí se tiene la propiedad asociativa del producto en $\Bbbk$.

Problema 13 *Demostrar el teorema 8; o sea, verificar las propiedades enunciadas.*

Adelantamos que el teorema 8 nos dice que $\Bbbk^{m\times n}$ es un espacio vectorial sobre el cuerpo $\Bbbk$, aunque el concepto *espacio vectorial* aún no se ha definido.

Las matrices se pueden también multiplicar entre ellas, aunque para que esté definido el producto ab se necesita que $a \in \Bbbk^{m\times n}$ y $b \in \Bbbk^{n\times p}$; es decir, que el número de columnas de a (primer factor) sea igual al de filas de b (segundo factor). A veces, por claridad o por énfasis pondremos $a \cdot b$ en vez de ab. Vamos a dar primero la fórmula en el caso $m = p = 1$ donde se multiplica una matriz fila por una matriz columna con n arbitrario. En tal caso

$$a = (a_1, \ldots, a_n), \qquad b = \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}, \qquad ab = (a_1, \ldots, a_n) \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix} = a_1 b^1 + a_2 b^2 + \cdots + a_n b^n.$$

Por ejemplo,

$$(1, 2) \begin{pmatrix} 3 \\ 4 \end{pmatrix} = 1 \cdot 3 + 2 \cdot 4 = 11, \qquad (1, 0, -1) \begin{pmatrix} 2 \\ -2 \\ -6 \end{pmatrix} = 1 \cdot 2 + 0 \cdot (-2) + (-1)\,(-6) = 8,$$

$$(a, b, 0, \cos\alpha) \begin{pmatrix} -a \\ -b \\ a \\ b \end{pmatrix} = -\left(a^2 + b^2\right) + b\cos\alpha, \qquad (1, 2, \ldots, n) \begin{pmatrix} 1 \\ 2 \\ \vdots \\ n \end{pmatrix} = \sum_{i=1}^{n} (i)^2.$$

Como se ve, en estos casos, $ab \in \Bbbk^{1\times 1} = \Bbbk$. Describimos como fórmula de dificultad intermedia el caso $a \in \Bbbk^{m\times n}$ y $b \in \Bbbk^{n\times 1}$, donde a es arbitraria pero b es una matriz columna con tantos términos como columnas tiene a. Entonces $ab \in \Bbbk^{m\times 1}$ y va a ser otro vector columna con m filas. La definición es

$$ab = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix} = \begin{pmatrix} a^1 \cdot b \\ \vdots \\ a^m \cdot b \end{pmatrix} \in \Bbbk^m, \quad a^i \cdot b = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \end{pmatrix} \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}.$$

Como hemos visto antes, $a^i \cdot b \in \Bbbk^{1\times 1} = \Bbbk$ y podemos detallar ab por

$$a^i \cdot b = a^i_1 b^1 + a^i_2 b^2 + \cdots + a^i_n b^n = \sum_{k=1}^{n} a^i_k b^k, \text{ luego } ab = \begin{pmatrix} \sum_{k=1}^n a^1_k b^k \\ \vdots \\ \sum_{k=1}^n a^m_k b^k \end{pmatrix}.$$

Aquí ya aparece lo fundamental de la regla del producto: *el coeficiente del producto en el lugar (i,j) es el producto de la fila i del primer factor por la columna j del segundo.* La fórmula general para $a \in \Bbbk^{m\times n}$ y $b \in \Bbbk^{n\times p}$ está dada imponiendo como coeficiente en el lugar (i,j) el producto de fila i por columna j. Con símbolos,

$$(ab)^i_j = a^i \cdot b_j = \begin{pmatrix} a^i_1 & \cdots & a^i_n \end{pmatrix} \begin{pmatrix} b^1_j \\ \vdots \\ b^n_j \end{pmatrix} = a^i_1 b^1_j + a^i_2 b^2_j + \cdots + a^i_n b^n_j = \sum_{k=1}^{n} a^i_k b^k_j.$$

Como se ve, la fórmula es laboriosa porque hay que calcular mp productos $a^i \cdot b_j$ y cada uno de ellos requiere n multiplicaciones cuyos productos se suman. Hay que practicar haciendo muchos ejemplos. Para empezar,

$$(1,2)\begin{pmatrix} 1 & 0 & h \\ -1 & 1 & 0 \end{pmatrix} = (1\cdot 1 + 2\cdot(-1), 1\cdot 0 + 2\cdot 1, 1\cdot h + 2\cdot 0) = (-1, 2, h),$$

$$\begin{pmatrix} p & q \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ p & q \end{pmatrix} = \begin{pmatrix} qp & q^2 \\ 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} p & q \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & h & h \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} q & q & hp & hp \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Para cálculos concretos como estos se puede eludir la fórmula $(ab)^i_j = a^i \cdot b_j = \sum_{k=1}^n a^i_k b^k_j$. Sin embargo, para m, n o p arbitrarios hay que usarla inevitablemente. Un ejemplo: si nos preguntan si $ab = ba$ podemos decir que es falsa (en general) tomando ejemplos con $a, b \in \Bbbk^{2\times 2}$, pero si piden probar que $a\,(bc) = (ab)\,c$ no queda más remedio que empezar escribiendo

$$(a\,(bc))^i_j = \sum_k a^i_k\,(bc)^k_j = \ldots\ldots$$

y precisar donde se mueve k. Se llega a una fórmula final que coincide con la que sale para $(ab)\,c$, luego $a\,(bc) = (ab)\,c$. Daremos un teorema (teorema 9 más abajo) con las principales propiedades del producto de matrices. De momento, algunos problemas que calculan productos.

Problema 14 *Verificar los siguientes productos de matrices, suponiendo $\Bbbk = \mathbb{R}$,*

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} x & 0 & 0 & -x \\ -x & 1 & 1 & x \end{pmatrix} = \begin{pmatrix} -x & 2 & 2 & x \\ -x & 4 & 4 & x \\ -x & 6 & 6 & x \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^3 = \begin{pmatrix} 1 & 3a \\ 0 & 1 \end{pmatrix},$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad (1,2,\ldots,n)\begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \sum_{i=1}^{n} i^2, \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+2b \\ 2a+b \\ a+2b \\ 2a+b \end{bmatrix}.$$

*Los casos **2** y **3** prueban que el producto de matrices no es, en general, conmutativo.*

Problema 15 *Calcular suponiendo $\Bbbk = \mathbb{C}$ para $a \in \mathbb{C}$ arbitrario,*

$$\begin{pmatrix} i & i \\ 0 & i \end{pmatrix}\begin{pmatrix} i & 0 \\ i & i \end{pmatrix}, \quad \begin{pmatrix} i & 0 \\ i & i \end{pmatrix}\begin{pmatrix} i & i \\ 0 & i \end{pmatrix}, \quad \begin{pmatrix} i & a \\ 0 & i \end{pmatrix}^3,$$

Problema 16 *Calcular para $\Bbbk = \mathbb{Z}_2$ en el primer caso y $\Bbbk = \mathbb{Z}_3$ en el segundo los productos matriciales.*

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 2 & 2 & 2 & 2 \end{pmatrix}$$

Tenemos pendiente, el teorema que recoge las principales propiedades del producto matricial. Advertencia: si no se sigue la demostración es que no se han asimilado las notaciones de sumatorios.

Teorema 9 *El producto de matices es asociativo y distributivo. Cuando tiene sentido,*

$$a\,(bc) = (ab)\,c, \qquad a\,(b+c) = ab + ac, \qquad (a+b)\,c = ac + bc$$

y también $Ia = a$, $aI = a$ (¡ojo! I representa en las dos igualdades la matriz identidad pero de tamaños diferentes).

Demostración. Para que tenga sentido $a\,(bc) = (ab)\,c$ debe ser $a \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^{n\times p}$, $c \in \mathbb{R}^{p\times q}$. Hemos de ver para $1 \le i \le m$ y $1 \le k \le q$ se tiene $(a\,(bc))^i_k = ((ab)\,c)^i_k$. Fijemos i en $\{1,\dots,m\}$ y k en $\{1,\dots,q\}$. Constataremos que en $(a\,(bc))^i_k$ aparecen todos los productos $a^i_u b^u_v c^v_k = P\,(u,v)$ con $1 \le u \le n$ y $1 \le v \le p$ y que lo mismo pasa con $((ab)\,c)^i_k$. *El trabajo a realizar para ver que $(a\,(bc))^i_k = ((ab)\,c)^i_k$ es mostrar que los productos P se suman en un orden distinto, pero como la suma en $\Bbbk$ es conmutativa, el resultado será el mismo.* Calculamos

$$(a\,(bc))^i_k = a^i \cdot (bc)_k = \sum_{u=1}^{n} a^i_u\,(bc)^u_k = \sum_{u=1}^{n} a^i_u \left(\sum_{v=1}^{p} b^u_v c^v_k\right) \overset{1}{=} \sum_{u=1}^{n}\sum_{v=1}^{p} a^i_u b^u_v c^v_k \overset{2}{=} \sum_{\substack{1\le u\le n\\ 1\le v\le p}} a^i_u b^u_v c^v_k.$$

En $\overset{1}{=}$ se usa que, por distributividad,

$$a^i_u \left(\sum_{v=1}^{p} b^u_v c^v_k\right) = \sum_{v=1}^{p} a^i_u b^u_v c^v_k = \sum_{v=1}^{p} P\,(u,v)\,,$$

Y en $\overset{2}{=}$ se usa (1.2) para la tabla de los $P\,(u,v)$. Ahora tenemos por otra parte

$$((ab)\,c)^i_k = (ab)^i\,c_k = \sum_{r=1}^{p} (ab)^i_r\,c^r_k = \sum_{r=1}^{p} \left(\sum_{s=1}^{n} a^i_s b^s_r\right) c^r_k \overset{1}{=} \sum_{r=1}^{p}\sum_{s=1}^{n} a^i_s b^s_r c^r_k \overset{2}{=} \sum_{\substack{1\le s\le n\\ 1\le r\le p}} a^i_s b^s_r c^r_k.$$

Como hace un momento, en $\overset{1}{=}$ se usa la distributividad y en $\overset{2}{=}$ se usa (1.2) para la tabla de los $P\,(s,r)$. Independientemente de las letras elegidas para denotar a los $P\,(\bullet,\bullet)$, lo que dicen los cálculos precedentes es que tanto $(a\,(bc))^i_k$ como $((ab)\,c)^i_k$ son la suma de todos estos P, luego son iguales como queríamos demostrar.

Probamos una de las formas de distributividad, digamos $a\,(b+c) = ab + ac$ donde $a \in \mathbb{R}^{m\times n}$, $b, c \in \mathbb{R}^{n\times p}$. Tenemos para $1 \le i \le m$ y $1 \le k \le p$ que

$$(a\,(b+c))^i_k = a^i\,(b+c)_k = \sum_{u=1}^{n} a^i_u\,(b+c)^u_k = \sum_{u=1}^{n} a^i_u\,(b^u_k + c^u_k)\,,$$

$$(ab)^i_k = a^i b_k = \sum_{u=1}^{n} a^i_u b^u_k, \qquad (ac)^i_k = a^i c_k = \sum_{u=1}^{n} a^i_u c^u_k.$$

Obtenemos que $a\,(b+c) = ab + ac$ pues coinciden los coeficientes dado que

$$(a\,(b+c))^i_k = \sum_{u=1}^{n} a^i_u\,(b^u_k + c^u_k)\,, \qquad (ab+ac)^i_k = (ab)^i_k + (ac)^i_k = \sum_{u=1}^{n} a^i_u b^u_k + \sum_{u=1}^{n} a^i_u c^u_k.$$

y para cada i, k es $a^i_u\,(b^u_k + c^u_k) = a^i_u b^u_k + a^i_u c^u_k$. Queda para el lector probar que $(a+b)\,c = ac + bc$.

Si $a \in \Bbbk^{m\times n}$ e $I = I_n$ veremos que $aI_n = a$. En efecto, para $1 \le i \le m$ y $1 \le k \le n$ se calcula

$$(aI_n)^i_k = a^i\,(I_n)_k = \sum_{j=1}^{n} a^i_j \delta^j_k \overset{1}{=} a^i_k.$$

Se justifica $\overset{1}{=}$ porque por definición de las deltas de Kronecker todas son cero excepto cuando el índice móvil j es k que vale 1. En el sumatorio solo queda a^i_k. Queda para el lector probar que $I_m a = a$ cuando $a \in \Bbbk^{m\times n}$. ♣

Señalamos con más énfasis algo que está en la demostración precedente pero que aparecerá con mucha frecuencia en cálculos con deltas de Kronecker. Si tenemos $\sum_{i=1}^{n} \delta^p_i a^{i*}_{\blacklozenge}$,donde $*$ y $\blacklozenge$ representan índices o grupos de índices de los que puede depender a, sabemos que hay n sumandos, pero como $\delta^p_i = 0$ cuando $p \neq i$, vemos que la suma se reduce a $a^{p*}_{\blacklozenge}$. Las deltas de Kronecker son grandes "destructoras de sumatorios".

Corolario 1 *Consideremos las matrices cuadradas $\Bbbk^{m\times m}$. Con las operaciones suma y producto, $\Bbbk^{m\times m}$ es un anillo no conmutativo pero sí unitario. Se verifican por tanto las propiedades*

1. Asociatividad de la suma. *Para todo $a, b, c \in \Bbbk^{m\times m}$ se tiene $a + (b + c) = (a + b) + c$.*
2. Conmutatividad de la suma. *Para todo $a, b \in \Bbbk^{m\times m}$ se tiene $a + b = b + a$.*
3. Existencia de cero. *Existe un elemento, que denotaremos por 0 (de hecho, $0 = 0_{m\times m}$) tal que para todo $a \in \Bbbk^{m\times m}$ se tiene $a + 0 = 0 + a = a$.*
4. Existencia de opuesto o inverso aditivo. *Para todo $a \in \Bbbk^{m\times m}$ existe un elemento que denotaremos por $-a$ (de hecho, $(-a)^i_j = -a^i_j$) tal que $a + (-a) = (-a) + a = 0$.*
5. Asociatividad del producto. *Para todo $a, b, c \in \Bbbk^{m\times m}$ se tiene $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.*
6. Distributividad. *Para todo $a, b, c \in \Bbbk^{m\times m}$ se tiene $a \cdot (b + c) = a \cdot b + a \cdot c$ y $(a + b) \cdot c = a \cdot c + b \cdot c$.*
7. Existencia de unidad. *Existe un elemento, que denotaremos por 1 y es distinto de 0 (de hecho, $1 = I_m$) tal que para todo $a \in \Bbbk^{m\times m}$ se tiene $a \cdot 1 = 1 \cdot a = a$.*

Una advertencia importante. Si se comparan las propiedades del *anillo* $\Bbbk^{m\times m}$ y del cuerpo $\Bbbk$, se ve que tienen las mismas propiedades excepto la existencia de inverso multiplicativo. Las fórmulas que dimos de sumatorios usan para suma y producto la conmutatividad, asociatividad y distributividad de estas operaciones, pero nunca la existencia de inverso multiplicativo. Por tanto, *las fórmulas de sumatorios que hemos dado siguen siendo válidas para matrices cuadradas.*

Problema 17 *Sean $a \in \Bbbk^{m\times n}$ y $b \in \Bbbk^{n\times p}$. Probar que*

1. *La columna j de ab es el producto de la matriz a por la columna j de b.*
2. *La fila i de ab es el producto de la fila i de a por la matriz b.*

Con símbolos

$$\Bbbk^{m\times 1} \ni (ab)_j = ab_j, \quad \Bbbk^{1\times p} \ni (ab)^i = a^i b. \blacklozenge$$

Solución. El coeficiente i del vector columna ab_j es $\sum_{k=1}^{n} a^i_k\,(b_j)^k = \sum_{k=1}^{n} a^i_k b^k_j$, que es precisamente el coeficiente i del vector columna $(ab)_j$. Quizás guste más

$$(ab)_j = \begin{pmatrix} (ab)^1_j \\ \vdots \\ (ab)^1_j \end{pmatrix} = \begin{pmatrix} a^1 \cdot b_j \\ \vdots \\ a^m \cdot b_j \end{pmatrix} = ab_j,$$

o examinar el ejemplo,

$$\left(\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 2 & 2 \\ 2 & 2 \end{pmatrix} \right)_2 = \begin{pmatrix} 4 & 6 \\ 2 & 2 \end{pmatrix}_2 = \begin{pmatrix} 6 \\ 2 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}\left(\begin{pmatrix} 0 & 2 \\ 2 & 2 \\ 2 & 2 \end{pmatrix}\right)_2 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}.$$

El otro caso queda para el lector. ♦

Una matriz *cuadrada* de $\mathbb{k}^{m\times m}$ se llama **invertible** *si existe* otra matriz b en $\mathbb{k}^{m\times m}$ tal que $ab = ba = I_m$. Se verá más adelante que una sola de las condiciones $ab = I$ o $ba = I_m$ sirve para garantizar la otra, pero de momento, para verificar que a es invertible se necesita verificar $ab = ba = I_m$. La matriz b es la **inversa de** a y es fácil probar que es única porque si b_1 y b_2 fuesen inversas de a; o sea, si $ab_1 = b_1a = I$ y $ab_2 = b_2a = I$, se deduciría que

$$b_1 = b_1I = b_1\left(ab_2\right) = \left(b_1a\right)b_2 = Ib_2 = b_2.$$

La inversa de a (si existe) se denota por a^{-1}. (¡Cuidado! en un cuerpo a^{-1} y $1/a$ son sinónimos, pero con matrices *jamás* se usa $1/a$ para la matriz inversa de a.) Las matrices **no invertibles** se llaman también **singulares** y, coherentemente, las invertibles son matrices **no singulares** o **regulares**. Usaremos poco la terminología "singular", "no singular" y "regular". Es muy importante conocer si a es o no invertible y, si lo es, cómo es a^{-1}. Esto llevará tiempo pero sí podemos deducir algunas propiedades de la invertibilidad.

Problema 18 *Probar que si a y b son invertibles, ab lo es también, siendo $(ab)^{-1} = b^{-1}a^{-1}$. Más generalmente, si $a_1, a_2, \ldots, a_k$ son matrices invertibles*[8] *el producto $p = a_1 \cdot a_2 \cdot \cdots \cdot a_k$ es invertible y*

$$p^{-1} = (a_k)^{-1} \cdot (a_{k-1})^{-1} \cdot \cdots \cdot (a_2)^{-1} \cdot (a_1)^{-1}$$

(la inversa del producto es el producto de las inversas en orden inverso*). Pruébese, admitiendo provisionalmente que para que a sea invertible basta una de las condiciones $ab = I$ o $ba = I$, que si el producto uv de dos matrices es invertible, u y v tienen que ser invertibles.*

El que $\mathbb{k}^{m\times m}$ sea un anillo pero no un cuerpo y, sobre todo, que falle la conmutatividad, trae algunas sorpresas, pues ciertas fórmulas familiares como $(a+b)^2 = a^2+b^2+2ab$ o $(a+b)(a-b) = a^2-b^2$ *dejan de ser ciertas*, aunque sí lo son si añadimos la condición de que sea $ab = ba$. Por ejemplo, en $\mathbb{k}^{2\times 2}$, para

$$a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad a+b = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (a+b)^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$a^2 = b^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad ab = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad a^2+b^2+2ab = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Problema 19 *Buscar un contraejemplo de $(a+b)(a-b) = a^2-b^2$ y probar esta fórmula y la del cuadrado del binomio con la hipótesis $ab = ba$.*

Otra sorpresa tiene que ver con las raíces cuadradas. Las hemos definido en un cuerpo $\mathbb{k}$ pero la definición tiene sentido en un anillo $\mathbb{A}$ aunque no sea conmutativo. Vimos que en un cuerpo hay *como máximo* dos raíces cuadradas. Sin embargo en un anillo de matrices puede haber un *número infinito de raíces cuadradas.* Por ejemplo,

$$\begin{pmatrix} 0 & \lambda^{-1} \\ \lambda & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{para todo } \lambda \neq 0.$$

Las matrices nos interesan no solo como objetos algebraicos que pueden ser sumados y multiplicados y serían por tanto una generalización de los números. Interesan también desde un punto de vista "geométrico" porque sirven para definir *funciones.* Si $a \in \mathbb{k}^{m\times n}$, esta matriz induce una función $L_a : \mathbb{k}^n \to \mathbb{k}^m$ (ojo a la inversión de posición de m y n) dada por $L_a(x) = ax$. Se llama la **función inducida por la matriz** a. Por ejemplo, para $\mathbb{k} = \mathbb{R}$,

$$a = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad L_a\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \begin{pmatrix} x^1+2x^2+3x^3 \\ 4x^1+5x^2+6x^3 \end{pmatrix}.$$

Estas funciones son el paradigma de las llamadas **funciones lineales**, concepto central en este curso. Se caracterizan por preservar sumas y productos por escalares,

$$L_a(x+y) = a(x+y) = ax+ay = L_a(x)+L_a(y), \quad L_a(\lambda x) = a(\lambda x) = \lambda(ax) = \lambda L_a(x).$$

Trataremos esto con mucho detalle más adelante.

[8] En este caso queda claro que a_j no es la fila j de a.

1.3.2. Tipos especiales de matrices

Ya hemos visto las matrices $0_{m\times n}$ e I_m. Vamos a suponer en esta sección, aunque lo recalcaremos, que todas las matrices son cuadradas $m \times m$. Las matrices más sencillas son las llamadas **matrices escalares,** que son cuadradas, con un mismo $\lambda \in \Bbbk$ para todos los a_i^i, y $a_j^i = 0$ si $i \neq j$. Más breve: $a = \lambda I$, con I la matriz unidad. De este modo se puede considerar $\Bbbk \subset \Bbbk^{m\times m}$ identificando $\lambda \in \Bbbk$ con λI. Si b es otra matriz, $ab = (\lambda I)\, b = \lambda\,(Ib) = \lambda b$. Queda más claro con palabras: si nos piden la matriz λb con $\lambda \in \Bbbk$ y $b \in \Bbbk^{n\times n}$, se llega a lo mismo si entendemos la pregunta como que se multiplica la matriz b por el *escalar* λ, que si se multiplica la matriz b por la *matriz* λI. Obsérvese que es *falsa* la fórmula

$$\begin{pmatrix} \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda \end{pmatrix} + \begin{pmatrix} b_1^1 & \cdots & b_m^1 \\ \vdots & \ddots & \vdots \\ b_1^m & \cdots & b_m^m \end{pmatrix} = \begin{pmatrix} \lambda + b_1^1 & \cdots & \lambda + b_m^1 \\ \vdots & \ddots & \vdots \\ \lambda + b_1^m & \cdots & \lambda + b_m^m \end{pmatrix}.$$

Una matriz *cuadrada* se llama **diagonal** si los términos a_j^i con $i \neq j$ son nulos. Los términos $a_1^1, a_2^2, \ldots, a_m^m$ pueden ser distintos y tomar cualquier valor, incluso 0. Las matrices escalares son un caso particular de matrices diagonales con todo $a_i^i = \lambda$. Hemos dicho que no se tiene en general la conmutatividad $ab = ba$.

Problema 20 *Probar que si a es escalar, $ab = ba$ para todo b. Si a y b son ambas diagonales también ab es diagonal. ¿Cuánto valen los $(ab)_i^i$? Mostrar con un ejemplo que puede ser a diagonal (pero no escalar) y existir b tal que $ab = ba$ sea falso.*

Si se tiene un polinomio $P(X) = \lambda_0 + \lambda_1 X + \lambda_2 X^2 + \cdots + \lambda_n X^n$ con coeficientes en $\Bbbk$ y una matriz $a \in \Bbbk^{m\times m}$, se puede sustituir X por a y tener una matriz

$$P(a) = \lambda_0 + \lambda_1 a + \lambda_2 a^2 + \cdots + \lambda_n a^n = [P(X)]_{X=a}$$

(Lo último es la notación que indica la sustitución de X por a.) Obsérvese que λ_0 es lo mismo que $\lambda_0 I = \lambda_0 I_m$ y que a^k no es la fila k de a sino la *potencia* k de a; o sea $a \cdot a \cdot \cdots \cdot a$ (k veces).

Veremos mucho más adelante aplicaciones resultantes de poder escribir una matriz b como $P(a)$. Vamos a probar que al sustituir a en sumas y productos de polinomios aparecen las sumas y productos de las sustituciones. Si se prefiere otro enunciado informal, es lo mismo operar primero (con polinomios) y sustituir después (la matriz) que sustituir primero (en los polinomios) y operar después (con matrices). Con símbolos,

$$[P(X) + Q(X)]_{X=a} = P(a) + Q(a), \qquad [P(X)\,Q(X)]_{X=a} = P(a)\,Q(a).$$

Problema 21 *Probar las afirmaciones anteriores. Obtener como corolario que $P(a)$ y $Q(a)$ siempre conmutan.* ♦

Solución. Hacemos la parte del producto que es la más difícil. Tenemos con (1.3),

$$P(X) = \sum_{i=0}^{p} \lambda_i X^i, \quad Q(X) = \sum_{j=0}^{q} \mu_j X^j, \quad P(X)\,Q(X) = \sum_{\substack{1\leq i\leq p \\ 1\leq j\leq q}} \lambda_i \mu_j X^{i+j}$$

A la derecha los sumandos van en el orden que se quiera. Por otra parte,

$$P(a) = \sum_{i=0}^{p} \lambda_i a^i, \quad Q(a) = \sum_{j=0}^{q} \mu_j a^j, \quad P(a)\,Q(a) = \left(\sum_{i=0}^{p} \lambda_i a^i\right)\left(\sum_{j=0}^{q} \mu_j a^j\right) \overset{*}{=} \sum_{\substack{1\leq i\leq p \\ 1\leq j\leq q}} \lambda_i \mu_j a^{i+j}$$

Se justifica $\overset{*}{=}$ por la fórmula (1.3) junto con el comentario tras el corolario 1. Es ya inmediato que $[P(X)\,Q(X)]_{X=a} = P(a)\,Q(a)$.

La última afirmación es obvia puesto que $P(X)\,Q(X) = Q(X)\,P(X)$. ♦

Problema 22 *Sea $P(X) = \lambda_0 + \lambda_1 X + \lambda_2 X^2 + \cdots + \lambda_n X^n$ un polinomio de grado ≥ 1 verificando $P(a) = 0$. Probar que si $\lambda_0 \neq 0$, a es invertible.*

Tras las matrices escalares y diagonales vienen las matrices **triangulares**. (Se supone también que a es cuadrada $m \times m$.) Hay matrices **triangulares superiores** y **triangulares inferiores**. Diremos que a es **triangular superior** (**inferior**) si todos los términos bajo (sobre) la diagonal principal son cero. Con símbolos,

$$a_j^i = 0 \text{ si } i > j \text{ (triangular superior)}, \qquad a_j^i = 0 \text{ si } i < j \text{ (triangular inferior)}$$

Las matrices escalares y diagonales son triangulares de los dos tipos y ser triangular de los dos tipos equivale a ser diagonal. Las matrices

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

son todas triangulares superiores y ninguna es triangular inferior salvo la última.

Problema 23 *Probar que la suma y el producto de dos matrices triangulares superiores es triangular superior.*

Ya definimos la **traspuesta** de a. Aquí no se necesita que a sea cuadrada y se define la traspuesta $a^\top$ por $\left(a^\top\right)_j^i = a_i^j$. Si a es cuadrada, también lo es $a^\top$; en general, si $a \in \Bbbk^{m\times n}$, entonces $a^\top \in \Bbbk^{n\times m}$.

Problema 24 *Probar que para $a, b \in \Bbbk^{m\times n}$ y $\lambda \in \Bbbk$ se tiene $(a+b)^\top = a^\top + b^\top$ y $(\lambda a)^\top = \lambda a^\top$ (da lo mismo operar y trasponer que trasponer y operar). Respecto al producto, si $a \in \Bbbk^{m\times n}$ y $b \in \Bbbk^{n\times p}$ se tiene $(ab)^\top = b^\top a^\top$ (la traspuesta del producto es el producto de las traspuestas* en orden inverso*).* ♦

Solución. Hacemos el producto. Sabemos que $ab \in \Bbbk^{m\times p}$ luego $ba \in \Bbbk^{p\times m}$. Fijamos i en $\{1, 2, \ldots, p\}$ y j en $\{1, 2, \ldots, m\}$. Calculamos

$$\left((ab)^\top\right)_j^i = (ab)_i^j = \sum_{k=1}^n a_k^j b_i^k = \sum_{k=1}^n \left(a^\top\right)_j^k \left(b^\top\right)_k^i = \sum_{k=1}^n \left(b^\top\right)_k^i \left(a^\top\right)_j^k = \left(b^\top\right)^i \left(a^\top\right)_j = \left(b^\top a^\top\right)_j^i . \blacklozenge$$

Aparte de esto hay ciertas propiedades obvias, pero que se utilizan mucho como $\left(a^\top\right)^\top = a$, $\left(I_m\right)^\top = I_m$ y, en general, que si a es diagonal, $a^\top = a$.

Problema 25 *Probar que si a es cuadrada e invertible, $\left(a^{-1}\right)^\top = \left(a^\top\right)^{-1}$; o sea, la traspuesta de la inversa es la inversa de la traspuesta. Indicación: con el trabajo previo es una comprobación directa.*

La trasposición sirve para definir muchos espacios de matrices. Supongamos $m = n$. Diremos que a es **simétrica** si $a^\top = a$ y es **antisimétrica** si $a^\top = -a$.

Problema 26 *Sea $a \in \mathbb{R}^{n\times n}$. Consideramos $a^{(k)} = a \cdot a \cdot \ldots \cdot a$, a por si misma k veces (ponemos el exponente entre paréntesis para no confundir con la fila k). Si a es simétrica (antisimétrica), determinar si lo es $a^{(k)}$.*

Una función muy importante que se define exclusivamente sobre *matrices cuadradas* $a \in \Bbbk^{n\times n}$ es la **traza**. Por definición, $\operatorname{tr}(a) = a_1^1 + a_2^2 + \ldots + a_n^n = \sum_{i=1}^n a_i^i$ es la suma de los términos de la diagonal principal.

Problema 27 *La traza tiene las siguientes propiedades elementales*

$$\operatorname{tr}(a+b) = \operatorname{tr}(a) + \operatorname{tr}(b), \quad \operatorname{tr}(\lambda a) = \lambda \operatorname{tr}(a), \quad \operatorname{tr}\left(a^\top\right) = \operatorname{tr}(a), \quad \operatorname{tr}(ab) = \operatorname{tr}(ba).$$

La propiedad $\operatorname{tr}(ab) = \operatorname{tr}(ba)$ se generaliza con importantes fórmulas como

$$\operatorname{tr}(abc) = \operatorname{tr}(cab) = \operatorname{tr}(bca), \quad \operatorname{tr}(abcd) = \operatorname{tr}(cdab).$$

Hay una definición particular si $a \in \mathbb{C}^{n\times n}$ es una matriz *compleja.* Se define la **adjunta (hermitiana)** como $a^* \in \mathbb{C}^{n\times n}$ y $(a^*)_j^i = \overline{a_i^j}$. Dicho con palabras: se traspone primero la matriz y luego se conjugan los términos; o bien, se conjuga primero la matriz (término a término) y luego se traspone.

Las operaciones $a \to a^\top$ y $a \to a^*$ tienen propiedades similares y $a^\top = a^*$ si a es una matriz real.

Problema 28 *Probar las propiedades semejantes a la trasposición del problema 24. Probar además que tanto para* $\top$ *como para* $*$ *se tiene que*

$$\operatorname{tr}\left(aa^{\top}\right)=\sum_{i,j=1}^{n}\left(a_j^i\right)^2,\quad \operatorname{tr}\left(aa^{*}\right)=\sum_{i,j=1}^{n}\left|a_j^i\right|^2$$

y que por tanto para que sea $a=0$ *es necesario y suficiente que sea* $\operatorname{tr}\left(aa^{\top}\right)=0$ *(caso* $\Bbbk=\mathbb{R}$*) y* $\operatorname{tr}\left(aa^{*}\right)=0$ *(caso* $\Bbbk=\mathbb{C}$*).*

Aunque ahora no tengamos aplicaciones para este problema, hay que decir que es muy importante porque da una forma alternativa de ver que una matriz es nula.

1.3.3. Un importante comentario sobre la notación

Ya advertimos al lector que al denotar al coeficiente de la matriz a en fila i y columna j por a_j^i no seguíamos la notación más corriente y que predomina, que es a_{ij}. El producto de matrices según se elija notación es

$$(ab)_j^i=a_1^ib_j^1+a_2^ib_j^2+\ldots+a_{n-1}^ib_j^{n-1}+a_n^ib_j^n=\sum_{k=1}^{n}a_k^ib_j^k,$$

o bien

$$(ab)_{ij}=a_{i1}b_{1j}+a_{i2}b_{2j}+\ldots+a_{i(n-1)}b_{(n-1)j}+a_{in}b_{nj}=\sum_{k=1}^{n}a_{ik}b_{kj}.$$

De momento, la única razón para nuestra elección que hemos dado es la facilidad de distinguir entre la fila i de a, que es a^i, de la columna j, que es a_j. Si solo se tratara de esto, el cambio no merecería la pena. Las razones son más profundas aunque con la base conceptual que tenemos es difícil ser precisos. Las hay de dos tipos.

1. En Álgebra Lineal hay muchísimas fórmulas que son sumas de productos y los elementos que se multiplican dependen ambos de un mismo índice, digamos i, y quizás de otros más. Se podrían poner los índices arriba o abajo (superíndices o subíndices) y escribir $\sum_i a_ib_i$, $\sum_i a_ib^i$ o incluso $\sum_i a^ib^i$. Sin embargo, al desarrollarse el Cálculo Tensorial, base matemática de la Teoría de la Relatividad, se constató que las fórmulas *se manejaban y recordaban mejor si se disponían los índices en los objetos que aparecían de modo que un índice de suma en factores diferentes apareciera una vez como subíndice y otra como superíndice*. Esto es la llamada **convención de Einstein** que permite incluso prescindir de los signos de sumatorios y simplificar tipográficamente las fórmulas.[9] Dejemos claro que no es esto más que un convenio de notación (y como tal, sustituible por muchos otros) cuyas ventajas no se ven a corto plazo.

2. La segunda razón es que una matriz representa muchas cosas que matemáticamente son muy diferentes. Puede representar puntos del plano y del espacio, funciones lineales, productos escalares, etc. Todos estos objetos son casos particulares del concepto inclusivo de **tensor**. Desde luego, el lector no conoce estos conceptos, pero debe saber que la forma como en el desarrollo inicial del Álgebra Lineal se sabía qué tipo de objeto representaba una matriz, era por la posición de sus índices. Incluso un físico de nuestros días (los científicos más en contacto con el Cálculo Tensorial) sabe que al hablarle de los tensores a_{ij}, a^{ij} y a_j^i le están nombrando cosas diferentes aunque vendrán representadas todas ellas por una misma matriz al "introducir coordenadas".

1.4. Sistemas de ecuaciones lineales

En toda la sección $\Bbbk$ es un cuerpo, y en una primera lectura y en los ejemplos se puede suponer $\Bbbk=\mathbb{R}$.

[9] Hoy día, con la posibilidad de elaborar textos matemáticos con LaTeX, tendemos a olvidar el enorme trabajo que suponía para los tipógrafos, más todavía que para los autores, el pasar un manuscrito matemático a letra impresa.

1.4.1. Cuestiones generales

En general, una **ecuación** en el conjunto arbitrario X es una manera de describir un subconjunto S (quizás vacío) de puntos de X. Primero un ejemplo: supongamos $X=\mathbb{R}^2$ y sea la ecuación $x-2y=5$. Se sabe que los puntos que la verifican forman una recta S, que $(1,-2)\in S$ puesto que $1\cdot 1-2\cdot(-2)=5$, y que $(0,1)\notin S$ al ser $1\cdot 0-2\cdot(1)\neq 5$. Se sabe también que los puntos de S pueden describirse por

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \lambda \in \mathbb{R};$$

es decir, que al elegir λ en $\mathbb{R}$ de todos los modos posibles y operar, obtenemos todos los puntos de S. El primer tipo de ecuaciones de S se llaman **ecuaciones implícitas** y el segundo, **ecuaciones paramétricas**.

Ahora la exposición general y abstracta: $S\subset X$ se puede describir de dos maneras: **(a)** Dando una función $f: X\to Y$ y un $y_0\in Y$ fijo y de modo que $S=\{x\in X \mid f(x)=y_0\}$; y **(b)** Tomando un conjunto Z y una función $P: Z\to X$ de modo que sea $S=\{x\in X \mid \exists z\in Z,\ x=P(z)\}$. Las ecuaciones $f(x)=y_0$ son las implícitas de S y las $x=P(z)$ con $z\in Z$ son las paramétricas de S. En los ejemplos del párrafo anterior, $X=\mathbb{R}^2$, $Y=Z=\mathbb{R}$, $f\left((x,y)^\top\right)=x-2y$, $y_0=5$, $P(\lambda)=(5,0)^\top+\lambda(2,1)^\top$. Frecuentemente, resolver una ecuación $f(x)=y_0$ consiste en considerar que define un subconjunto S de X y expresar S en paramétricas, aunque podría entenderse también en el otro sentido, que sería obtener a partir de ecuaciones paramétricas para S otras ecuaciones implícitas.

Nos vamos a centrar en las **ecuaciones lineales**.[10] El conjunto X es $\Bbbk^n$, Y es $\Bbbk^m$ y la función $f:\Bbbk^n\to\Bbbk^m$ viene representada por una matriz $a\in\Bbbk^{m\times m}$. La ecuación, de uno u otro modo es

$$\begin{pmatrix} a_1^1 & \cdots & a_j^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots & & \vdots \\ a_1^i & \cdots & a_j^i & \cdots & a_n^i \\ \vdots & & \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_j^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^j \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} y^1 \\ \vdots \\ y^i \\ \vdots \\ y^m \end{pmatrix}, \quad \begin{cases} a_1^1x^1+\cdots+a_j^1x^j+\cdots+a_n^1x^n=y^1 \\ \vdots \\ a_1^ix^1+\cdots+a_j^ix^j+\cdots+a_n^ix^n=y^i \\ \vdots \\ a_1^mx^1+\cdots+a_j^mx^j+\cdots+a_n^mx^n=y^m \end{cases} \tag{1.5}$$

Si se quiere un ejemplo concreto (ponemos (x,y,z) en vez de $\left(x^1,x^2,x^3\right)$ porque es más familiar)

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 8 \end{pmatrix}, \text{ o bien } \begin{cases} x+2y+3z=7 \\ 4x+5y+6z=8 \end{cases}$$

Suponemos que el lector podrá demostrar (aunque no se requiere ahora mismo que sepa hacerlo) que $S\in\mathbb{R}^3$ es una recta, expresable en paramétricas de muchas formas, de las que damos dos,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1+z \\ 2-2z \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ -2x \\ x+1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

y hay que entender que al variar, según el caso, z,λ,x,μ en $\mathbb{R}$ se tienen todos los puntos de S.

Formulamos el problema general. Tenemos una matriz $a\in\Bbbk^{m\times n}$ e $y\in\Bbbk^m$. Una **ecuación lineal** es la que tiene la forma $ax=y$, y es pues expresable en las formas de (1.5). En $ax=y$ hay m ecuaciones, que son las de la llave de (1.5), n **incógnitas** $x^1,\dots,x^n$y m **términos independientes** (o **constantes)** $y^1,\dots,y^m$. Tanto los x^j como los y^i son escalares (o sea, elementos de $\Bbbk$) pero podemos decir también que $x\in\Bbbk^n$ es la **incógnita (vectorial)** e y el **término independiente (vectorial)**. El sistema es **compatible** si tiene al menos una solución e **incompatible** en caso contrario (otros dicen **consistente** e **inconsistente**). Un sistema es **determinado** si solo hay una solución e **indeterminado** si hay más de una. Un sistema es **homogéneo** si $y=0$, luego $ax=0$ es compatible pues $0\in\Bbbk^n$ es solución. Si tenemos $ax=y$, llamaremos a $ax=0$ el **sistema homogéneo asociado** a $ax=y$. Como es laborioso escribir (1.5) usaremos siempre que podamos la notación $ax=y$ y en vez de usar las llaves se trabajará con dos matrices: la matriz a y la matriz $(a\mid y)\in\Bbbk^{m\times(n+1)}$ que se obtiene añadiendo a la derecha de a

[10] Aunque hablamos de un cuerpo $\Bbbk$ arbitrario, el lector puede, si se siente más cómodo, suponer que $\Bbbk=\mathbb{R}$ en una primera lectura. En los ejemplos, si no se dice lo contrario, se supondrá $\Bbbk=\mathbb{R}$.

la columna formada por y. A $(a \mid y)$ se la llama la **matriz ampliada del sistema** y a a simplemente la **matriz del sistema**. Para la teoría general conviene usar la primera formulación del sistema en (1.5) en vez de la segunda, por lo laborioso que resulta incluir los signos $+, =$ y las letras x^j y y^i.

Queremos un procedimiento sistemático para resolver el sistema $ax = y$, dando todas sus soluciones y conociendo qué ha de suceder para que no existan. Esencialmente hay tres pasos:

1. Probar que si las incógnitas están dispuestas en las ecuaciones de un modo especial, que luego precisaremos, es sencillo resolver el sistema. Estos sistemas "especiales" son aquellos en donde la matriz a está en forma escalonada, reducida o escalonada reducida, que definiremos. Preferimos abordar este paso primero, porque marca el objetivo de los pasos **2** y **3**.

2. Mostrar que se pueden sustituir unas ecuaciones del sistema por otras de modo que, si bien cambia el sistema, no se altera el conjunto de soluciones.

3. Probar que para el sistema $ax = y$ se pueden elegir las operaciones de **2** adecuadadamente, de modo que se transforma en otro sistema $bx = z$ con b en forma escalonada, reducida o escalonada reducida. Y puesto que por **1** ya sabemos cómo resolver $bx = z$, tendremos por **2** las soluciones de $ax = y$

Hemos sido un poco ambiguos al hablar de "dar las soluciones". Nos referimos a que, si el conjunto S de soluciones es no vacío, se dará en forma paramétrica e intentaremos "afinar" (no podemos ser más precisos) sobre cómo son las mejores. Las tareas de **1-3** se hacen en las secciones siguientes.

1.4.2. Matrices y sistemas en forma escalonada y reducida

Consideramos en lo sucesivo el sistema $ax = y$ con $a \in \Bbbk^{m\times n}$. Evidentemente, si a tiene muchos coeficientes que sean 0 o 1, el sistema será más fácil de resolver. Vamos a describir en abstracto una estructura de a con la que $ax = y$ tiene fácil solución. En cada fila $a^i = \left(a^i_1, \ldots, a^i_n\right)$ definimos el **pivote** como el primer coeficiente no nulo. Se entiende que si $a^i_1 = \ldots = a^i_n = 0$, la fila i no tiene pivote. Las **columnas pivotales** son aquellas donde están los pivotes. Entonces, a está en **forma escalonada** si

1E Las filas nulas son las últimas; digamos $a^{r+1}, \ldots, a^m$ para cierto r con $1 \leq r \leq m$.

2E Si $a^1, \ldots, a^r$ son las filas no nulas, y $a^1_{p_1}, a^2_{p_2}, \ldots, a^r_{p_r}$ sus pivotes, se tiene $1 \leq p_1 < p_2 < \ldots < p_r \leq n$. (Al ir bajando de fila, el pivote se mueve, como mínimo, un lugar a la derecha.)

Damos como ejemplos, con columnas pivotales respectivas 2 y 3; 1 y 3; 3; y $1, 2$ y 3,

$$\begin{pmatrix} 0 & 3 & 1 & 1 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 4 & 6 & 8 \\ 0 & 2 & 4 & 6 \\ 0 & 0 & 2 & 4 \end{pmatrix}.$$

Se dice que a está en **forma reducida** si

1R Los pivotes (por definición, no nulos) valen todos 1.

2R Si un pivote está en la columna i, este es el único término no nulo de esa columna.

Damos como ejemplos

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 2 \end{pmatrix}.$$

Si se cumplen las cuatro condiciones, a está en forma **escalonada reducida**. Como ejemplos,

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 4 \end{pmatrix}.$$

Pretendemos mostrar cómo si a está en forma escalonada o escalonada reducida, la solución es sencilla. Obsérvese que no decimos que la matriz ampliada $(a \mid y)$ esté en forma escalonada o reducida, sino solo que lo está a. El proceso de solución se llama **sustitución ascendente** y consiste en ir resolviendo las ecuaciones desde la última a la primera aprovechando la forma escalonada de la matriz.[11] Veamos un ejemplo, con las letras más familiares $x, y, z, \ldots$ en vez de $x^1, x^2, x^3, \ldots$. El sistema

$$\left\{\begin{array}{r} 2x+3y+z=3 \\ y-z=-1 \\ 2z=4 \end{array}\right. \quad \text{tiene matriz ampliada} \quad \left(\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 2 & 4 \end{array}\right). \tag{1.6}$$

La solución es $(x, y, z)^\top = (-1, 1, 2)$ porque

$$z = \frac{4}{2} = 2, \;\; y = -1 + z = -1 + 2 = 1 \;\; , x = \frac{1}{2}(3 - 3y - z) = \frac{1}{2}(3 - 3 \cdot 1 - 2) = -1.$$

Es posible que haya menos pivotes que incógnitas, en cuyo caso, las variables no pivotales pasan primero a la derecha y luego se resuelve el sistema. Por ejemplo,

$$\left\{\begin{array}{r} 2x-y-z=-1 \\ 5z-t=4 \\ 4t=4 \end{array}\right. \quad \text{con matriz ampliada} \quad \left(\begin{array}{cccc|c} 2 & -1 & -1 & 0 & -1 \\ 0 & 0 & 5 & -1 & 4 \\ 0 & 0 & 0 & 4 & 4 \end{array}\right), \tag{1.7}$$

se pone en forma equivalente para ser resuelto como

$$\left\{\begin{array}{r} 2x-z=y-1 \\ 5z-t=4 \\ 4t=4 \end{array}\right. \quad \text{con matriz ampliada} \quad \left(\begin{array}{ccc|c} 2 & -1 & 0 & y-1 \\ 0 & 5 & -1 & 4 \\ 0 & 0 & 4 & 4 \end{array}\right),$$

y se obtienen por sustitución ascendente las soluciones

$$t = \frac{4}{4} = 1, \; z = \frac{1}{5}(4 + t) = 1, \; y = y, \; x = \frac{1}{2}(-1 + y + z) = \frac{1}{2}(-1 + y + 1) = \frac{1}{2}y.$$

Si se pone con notación vectorial, la solución es (se usa $\bullet^\top$ para ahorrar espacio)

$$(x, y, x, t)^\top = \left(\frac{1}{2}y, y, 1, 1\right)^\top = (0, 0, 1, 1)^\top + \lambda\left(\frac{1}{2}, 1, 0, 0\right)^\top.$$

Puede darse un caso extremo, que es que haya una ecuación del tipo $0 = \beta$ con $\beta \neq 0$, y en ese caso el sistema es incompatible y la falta de solución es, sin más, la respuesta del problema.

Describimos el proceso general para $ax = y$ con a en forma escalonada.

1. Las variables del sistema se dividen en **variables pivotales**, que son las que multiplican a un pivote, siendo las otras las variables **no pivotales**. Se pasan las variables no pivotales al lado de los términos independientes y jugarán en adelante el papel de parámetros. Si hay alguna ecuación incompatible del tipo $0 = \beta$ con $\beta \neq 0$, el sistema no tiene solución y recíprocamente.

2. Suponiendo compatibilidad, nos ha quedado un sistema escalonado para las variables pivotales, que se resuelve por sustitución ascendente, con fórmulas para las variables pivotales en función de las no pivotales (si las hay).

3. Es posible que alguna variable x^j no aparezca en el sistema. Se la considera no pivotal y en la solución se entiende que puede tomar cualquier valor de $\Bbbk$.

Las variables no pivotales se llaman también **libres** y las pivotales **determinadas** o **dependientes**. Supongamos que $ax = y$ es compatible. Al sustituir las variables libres de todas las formas posibles, obtenemos dependiendo de ellas (valga la redundancia) las dependientes que ellas determinan y así todas las soluciones de $ax = y$. Si, como caso extremo, no hay variables libres, todas las variables

[11] En inglés el proceso se llama *back substitution* porque se sustituye empezando por la ultima ecuación y, en orden inverso hacia la primera. "Retrosustitución" nos parece muy exacto pero excesivamente complejo. Si imaginamos las ecuaciones escritas en un papel, empezamos resolviendo la de más abajo, y vamos subiendo resolviéndolas correlativamente hasta llegar a la primera, hacemos sustituciones en orden ascendente en el texto. Por eso hemos elegido "sustitución ascendente".

son determinadas y la solución es única. Si hay variables libres, hay múltiples soluciones. En resumen, habiendo solución, esta es única si y solo si no hay variables libres. Aunque todo esto queda claro, vamos a hacer una descripción con símbolos para $ax = y$.

Sean $x^{p_1}, \ldots, x^{p_r}$ las variables dependientes y $x^{q_1}, \ldots, x^{q_s}$ las libres, verificando los índices

$$1 \leq p_1 < p_2 < \ldots < p_r \leq n, \quad 1 \leq q_1 < q_2 < \ldots < q_s \leq n, \quad r + s = n.$$

Tenemos unas funciones $f^\bullet$ que expresan las x^{p_i} en términos de las x^{q_j} con fórmulas generales

$$\left\{ \begin{array}{c} x^{p_1} = f^{p_1}\left(x^{q_1}, \ldots, x^{q_s}\right) \\ \vdots \\ x^{p_r} = f^{p_{r1}}\left(x^{q_1}, \ldots, x^{q_s}\right) \end{array} \right. \quad \text{o también} \quad P \begin{pmatrix} x^{q_1} \\ \vdots \\ x^{q_s} \end{pmatrix} = \begin{pmatrix} \vdots \\ x^{q_j} \\ \vdots \\ f^{p_j}\left(x^{q_1}, \ldots, x^{q_s}\right) \\ \vdots \end{pmatrix}.$$

Con símbolos queda más clara una importante propiedad: la función P de $\Bbbk^s$ en $S \subset \Bbbk^n$ que parametriza el conjunto de soluciones S de $ax = y$ es *biyectiva*. Esto es como decir que a cada sucesión de variables libres $\left(x^{q_1}, \ldots, x^{q_s}\right)^\top$ le corresponde una solución y solo una de $ax = y$. En cierto modo estamos dando la solución de modo óptimo porque no hay parámetros sobrantes ni posible confusión al haber parámetros diferentes que la determinen. El que P es biyectiva es inmediato. Es suprayectiva ya que, por construcción, su imagen es S. Por otra parte, si tenemos $P\left(x^{q_1}, \ldots, x^{q_s}\right)^\top = P\left(u^{q_1}, \ldots, u^{q_s}\right)^\top$, observamos los huecos $q_1 < q_2 < \ldots < q_s$ de estos vectores, que están en $\Bbbk^n$ y obtenemos $x^{q_1} = u^{q_1}, \ldots, x^{q_s} = u^{q_s}$, que es la condición de que P sea inyectiva. Ilustramos esto con

$$\left\{ \begin{array}{r} x + 2y + 3z + 3v + 4w = 11 \\ u - 4w = 12 \end{array} \right. \tag{1.8}$$

cuya solución es $u = 12 + 4w$, $x = 11 - 2y - 3z - 3v - 4w$ en forma tradicional. Con la función P,

$$P \begin{pmatrix} y \\ z \\ v \\ w \end{pmatrix} = \begin{pmatrix} 11 - 2y - 3z - 3v - 4w \\ y \\ z \\ 12 + 4w \\ v \\ w \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \\ u \\ v \\ w \end{pmatrix}.$$

Nos hemos centrado en el caso en el que a está en forma escalonada. Si el sistema está en forma reducida, se puede alterar el orden de las ecuaciones para que pasen a estar *además* en forma escalonada, poniendo primero la que tenga el pivote 1 más a la izquierda, luego la que tenga en pivote 1 más a la izquierda entre las restantes,... y así colocarlas en forma escalonada. Hay que observar que la definición de matriz reducida impide que en la columna de un pivote haya otro término no nulo, luego los pivotes de las diversas ecuaciones ocupan diferentes columnas y al reorganizar las ecuaciones aparece la forma de escalera. Un sistema en forma escalonada reducida es más fácil de resolver que si solo está escalonado, pues cada variable pivotal solo aparece en una ecuación y es fácil expresarla con las variables no pivotales.

Problema 29 *Resolver el sistema escalonado*

$$\left\{ \begin{array}{r} x^1 + 2x^2 - x^3 - 2x^4 = 1 \\ 4x^3 - 6x^4 = 8 \end{array} \right.$$

1.4.3. Operaciones fila en matrices y sistemas lineales

El lector habrá hecho problemas donde se resuelven sistemas lineales sustituyendo unas ecuaciones por otras, sumándolas o multiplicándolas por números no nulos. El objeto es reducir lo más posible el número de coeficientes no nulos de las variables, de modo estas que aparezcan convenientemente "aisladas". Estos procesos se llaman de eliminación, y los básicos son los de **eliminación de Gauss** y **eliminación de Gauss-Jordan**. Pretenden modificar la matriz ampliada $(a \mid y)$ de $ax = y$ para llegar a otra $(b \mid z)$ de modo que b sea "sencilla" y que $bx = z$ tenga las mismas soluciones que $ax = y$.

Dado un sistema $ax = y$ como en (1.5) vamos a describir tres operaciones que transformaran $ax = y$ en otro con el mismo número de ecuaciones y sin modificar sus soluciones (incluso si este conjunto es vacío). Estas operaciones, llamadas **operaciones elementales (de fila)**, modifican el sistema de dos maneras, bien sea alterando el orden de las ecuaciones o bien sustituyendo *una sola* ecuación por una combinación de otras. Las operaciones son

1. Permutar dos ecuaciones, digamos la u y la v, dejando invariables las demás. Esta operación se denotará por F_{uv}.
2. Sustituir la ecuación u por la suma de la ecuación u y el producto de $\lambda \in \Bbbk$ por la ecuación v, dejando invariables las demás. Esta operación se denotará por $F_{uv}[\lambda]$. Se supone $u \neq v$ pero no necesariamente $u < v$. Si es $\lambda = 0$ el sistema no se modifica.
3. Sustituir la ecuación w por la propia ecuación w multiplicada por $\mu \in \mathbb{R}$ *no nulo,* dejando invariables las demás. Esta operación se denotará por $F_w[\mu]$.

Por ejemplo, aplicar F_{12} al sistema es intercambiar las dos primeras ecuaciones; aplicar $F_{2,5}[7]$ es sustituir la ecuación 2 por la suma de la ecuación 2 y la 5 multiplicada por 7; y aplicar $F_3[-1]$ es multiplicar por -1 la ecuación 3; o sea, cambiarla de signo. Siempre, las ecuaciones $3, \ldots, m$ con F_{12} quedan invariantes; con $F_{2,5}[7]$ solo se cambia la ecuación 2; y con $F_3[-1]$ solo la ecuación 3.

Si tenemos una matriz $a \in \Bbbk^{m\times n}$ podemos hacer con ella tres **operaciones elementales** que se denotarán también con F_{uv}, $F_{uv}[\lambda]$ y $F_w[\mu]$ y tienen definiciones análogas:

1. La matriz $b = F_{uv}(a)$ se obtiene permutando las filas u y v dejando invariables las demás. Con símbolos,
$$b^u = a^v, \qquad b^v = a^u, \qquad b^w = a^w \quad \text{si } w \neq u, v.$$
2. La matriz $b = F_{uv}[\lambda](a)$ se obtiene sustituyendo la fila u por la fila u más λ por la fila v, siendo $\lambda \in \Bbbk$, y dejando invariables las demás, incluida la propia fila v. Con símbolos,
$$b^u = a^u + \lambda a^v, \qquad b^w = a^w \quad \text{si } w \neq u.$$
3. La matriz $b = F_w[\mu](a)$ se obtiene sustituyendo la fila w por la fila w multiplicada por $\mu \in \Bbbk$ *no nulo* y dejando invariables las demás. Con símbolos,
$$b^w = \mu a^w, \qquad b^u = a^u \quad \text{si } w \neq u.$$

Ilustraremos las propiedades con matrices por tener notaciones más sencillas. *Las operaciones elementales tienen operaciones inversas.* Esto quiere decir que si $F : \Bbbk^{m\times n} \to \Bbbk^{m\times n}$ es una de estas operaciones, hay otra operación G tal que para toda matriz a se tiene $G(F(a)) = a$ y $F(G(a)) = a$. Informalmente, la operación G "deshace" la operación F, y formalmente, como funciones que son de $\Bbbk^{m\times n}$ en $\Bbbk^{m\times n}$, F y G son funciones inversas una de la otra. Precisando,

$$(F_{uv})^{-1} = F_{vu} = F_{uv}, \quad (F_{uv}[\lambda])^{-1} = F_{uv}[-\lambda], \quad (F_w[\mu])^{-1} = F_w[1/\mu]. \tag{1.9}$$

Debemos probar, por ejemplo, que para toda $a \in \Bbbk^{m\times n}$ se tiene $F_{uv}[\lambda] \circ F_{uv}[-\lambda](a) = a$. Veámoslo. La matriz $b = F_{uv}[-\lambda](a)$ viene dada por

$$b^u = a^u - \lambda a^v, \qquad b^w = a^w \quad \text{si } w \neq u.$$

La matriz $c = F_{uv}(b) = F_{uv}[\lambda] \circ F_{uv}[-\lambda](a)$ cumple

$$c^u = b^u + \lambda b^v = (a^u - \lambda a^v) + \lambda a^v = a^u, \qquad c^w = b^w = a^w \quad \text{si } w \neq u.$$

En definitiva, $c = a$ como queríamos demostrar. Es análogo ver que $F_{uv}[-\lambda] \circ F_{uv}[\lambda](a) = a$.

Problema 30 *Hacer alguna comprobación más sobre las inversas de las funciones* $F_\bullet[\bullet]$.

Diremos que dos sistemas $ax = y$ y $bx = z$ son **equivalentes** si se puede pasar de uno a otro realizando un número finito de operaciones elementales. Es una comprobación pesada pero muy sencilla la que verifica que ser equivalentes con esta definición es relación de equivalencia. La importancia de la relación se basa en este teorema.

Teorema 10 *Los sistemas equivalentes* $\mathbf{S}: ax = y$ *y* $\mathbf{T}: bx = z$ *tienen el mismo conjunto de soluciones.*

Demostración. Sean $\Sigma_{\mathbf{S}}$ y $\Sigma_{\mathbf{T}}$ los conjuntos de soluciones. Basta mostrar que si F es una operación elemental y $\mathbf{T} = F(\mathbf{S})$ entonces $\Sigma_{\mathbf{S}} \subset \Sigma_{\mathbf{T}}$ porque, como se supone probado esto con $F, \mathbf{S}, \mathbf{T}$ arbitrarios, podemos invertir los papeles y usando que $\mathbf{S} = F^{-1}(\mathbf{T})$ llegar a que $\Sigma_{\mathbf{T}} \subset \Sigma_{\mathbf{S}}$, el otro contenido. Verifiquemos por ejemplo que $\Sigma_{\mathbf{S}} \subset \Sigma_{\mathbf{T}}$ con $F = F_{uv}[\lambda]$. Sea $s \in \Bbbk^n$ solución de $\mathbf{S}$. Tiene que verificar todas las ecuaciones de $\mathbf{S}$ y (¡evidentemente!) s verificará todas las de $\mathbf{T}$, salvo quizás la $u-$ésima que es diferente. Pero esta ecuación es

$$\sum_{i=1}^{n} \left(a_i^u + \lambda a_i^v\right) x^i = y^u + \lambda y^v$$

y se tiene

$$\sum_{i=1}^{n} \left(a_i^u + \lambda a_i^v\right) s^i = \sum_{i=1}^{n} a_i^u s^i + \sum_{i=1}^{n} \lambda a_i^v s^i = y^u + \lambda y^v$$

porque $\sum_{i=1}^{n} a_i^u s^i = y^u$ y $\sum_{i=1}^{n} a_i^v s^i = y^v$ son las ecuaciones u y v que $\mathbf{S}$, que por hipótesis verifica s. Quedan para el lector las verificaciones con $F = F_{uv}$ y $F_w[\mu]$. ♣

1.4.4. Transformación de matrices

Dejamos de lado un tiempo los sistemas lineales y vamos a centrarnos en la posibilidad de transformar a forma escalonada, reducida, o escalonada reducida con operaciones elementales.

Teorema 11 *Toda matriz* $a \in \Bbbk^{m\times n}$ *se puede transformar en escalonada.*

Demostración. Si a no estuviera en forma escalonada, permutaríamos sus filas a de modo que en la primera, el pivote, quedara lo más a la izquierda posible. Con estas operaciones tipo $F_{u,v}$, hemos transformado $a = a(0)$ en $a(1)$ con pivote $a(1)_j^1 \neq 0$ y, por la construcción, con las columnas $a(1)_1, \ldots, a(1)_{j-1}$ nulas. Se trata ahora de eliminar en $a(1)$ los términos $a(1)_j^i$ con $i \geq 2$; o sea, los términos de esa primera columna j no nula que están debajo de $a(1)_j^1$. Para eliminar $a(1)_j^i$ se resta a la fila i la fila 1 multiplicada por $a(1)_j^i / a(1)_j^1$, haciendo una de estas operaciones cada vez que $a(1)_j^i \neq 0$. Obsérvese que en la nueva matriz $a(2)$,

$$a(2)_j^i = a(1)_j^i - \frac{a(1)_j^i}{a(1)_j^1} a(1)_j^1 = 0, \quad i = 2, 3, \ldots, n,$$

y que las $j-1$ primeras columnas siguen siendo nulas, aunque las columnas de $a(2)$ a la derecha de la columna j habrán sido modificadas. Ignoramos en lo que sigue la fila 1 de $a(2)$, que no tocaremos más, y permutamos las restantes de modo que pase a la fila 2 una con el pivote lo más a la izquierda posible. En la nueva matriz $a(3)$, este pivote será $a(3)_k^2$ con $k > j$ obligatoriamente. Repetimos operaciones análogas a las anteriores para transformar en cero los coeficientes $a(3)_k^h$ y $h \geq 3$ que no lo sean; es decir, los que no sean nulos y estén debajo del pivote $a(3)_k^2$. Esto nos lleva a una matriz $a(4)$. Ignoramos las dos primeras filas de $a(4)$ y permutamos las restantes de modo que en la nueva matriz $a(5)$ el pivote $a(5)_\ell^3$ esté lo más a la izquierda posible. Con las ideas de antes, eliminaremos los coeficientes no nulos de $a(5)$ bajo el pivote $a(5)_\ell^3$, $\ell > k > j$, llegando a una matriz $a(6)$. Repitiendo el proceso acabamos en una matriz $a(2t)$ en forma escalonada. Hemos dado siempre por cierto que tras llegar a una matriz $a(2s)$ e ignorar las s primeras filas, íbamos a encontrar en las restantes alguna fila con pivote. Esto puede no suceder, pero es porque en tal caso las últimas filas son todas nulas y esa matriz ya está en forma escalonada como queríamos.

Llamamos la atención sobre no haber necesitado operaciones $F_w[\mu]$. ♣

Problema 31 *Aplicar los procedimientos del teorema 11 a la matriz*

$$a = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 4 & 2 & 1 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \end{pmatrix}. \blacklozenge$$

Solución. Los pasos son (quizás no seguimos la vía mejor)

$$a(0)=\begin{pmatrix}0&0&0&1\\0&4&2&1\\0&2&0&3\\0&0&3&0\end{pmatrix}\overset{1}{\to}\begin{pmatrix}0&4&2&1\\0&2&0&3\\0&0&3&0\\0&0&0&1\end{pmatrix}=a(1),$$

pasándose la primera fila al final. A continuación se resta a la fila 2 la mitad de la fila 1 y

$$a(1)=\begin{pmatrix}0&4&2&1\\0&2&0&3\\0&0&3&0\\0&0&0&1\end{pmatrix}\overset{2}{\to}\begin{pmatrix}0&4&2&1\\0&2-\frac{2}{4}\cdot 4&0-\frac{2}{4}\cdot 2&3-\frac{2}{4}\cdot 1\\0&0&3&0\\0&0&0&1\end{pmatrix}=\begin{pmatrix}0&4&2&1\\0&0&-1&\frac{5}{2}\\0&0&3&0\\0&0&0&1\end{pmatrix}=a(2).$$

Ahora se suma a la fila 3 el triple de la fila 2 porque no hay que permutar filas, resultando

$$a(2)=a(3)\overset{3}{\to}\begin{pmatrix}0&4&2&1\\0&0&-1&\frac{5}{2}\\0&0&3+3\cdot(-1)&0+3\cdot\frac{5}{2}\\0&0&0&1\end{pmatrix}=\begin{pmatrix}0&4&2&1\\0&0&-1&\frac{5}{2}\\0&0&0&\frac{15}{2}\\0&0&0&1\end{pmatrix}=a(4),$$

No se necesita permutar filas para eliminar el 1 al final, sino basta restar a la fila 4 la fila 3 multiplicada por $\frac{2}{15}$ y

$$a(4)=a(5)\overset{4}{\to}\begin{pmatrix}0&4&2&1\\0&0&-1&\frac{5}{2}\\0&0&0&\frac{15}{2}\\0&0&0&1-\frac{2}{15}\cdot\frac{15}{2}\end{pmatrix}=\begin{pmatrix}0&4&2&1\\0&0&-1&\frac{5}{2}\\0&0&0&\frac{15}{2}\\0&0&0&0\end{pmatrix}=a(6).\ \blacklozenge$$

El algoritmo que permite transformar a en b de forma escalonada se llama el **algoritmo de Gauss**. Pronto veremos que adaptado al sistema $ax=y$ permite transformar ese sistema en otro $bx=z$ con las mismas soluciones pero en forma escalonada que ya sabemos resolver. Hemos advertido al final de la demostración del teorema 11 que no se han usado operaciones elementales del tipo $F_w[\mu]$, pero si se hace, se puede llegar a la forma escalonada reducida, cosa poco conveniente para los ordenadores porque incrementa en un 50 % aproximadamente las multiplicaciones a realizar. El **algoritmo de Gauss-Jordan** que se describe en el siguiente teorema tiene no solo la posibilidad de llegar a la forma escalonada reducida más conveniente a veces para las personas, que han de llevar la cuenta de numerosas operaciones y sustituciones.

Teorema 12 *Con operaciones elementales, toda matriz $a\in\Bbbk^{m\times n}$ se puede transformar en forma escalonada reducida.*

Demostración. Se trata de refinar el algoritmo de Gauss. Una vez que se ha llegado desde $a=a(0)$ hasta $a(1)$ con el pivote de la primera fila lo más a la izquierda posible, se divide esta fila 1 por el pivote, llegándose a una matriz $a'(1)$ con pivote 1 en la fila 1; es decir, $a'(1)^1_j=1$. Es sencillo eliminar los coeficientes $a'(1)^i_j$ debajo del $a'(1)^1_j$ restando a la fila i la fila 1 de $a'(1)$ multiplicada por $a'(1)^i_j$. Hasta aquí es como completar el algoritmo de Gauss para que los pivotes sean 1, pero sin tener aún forma reducida. No obstante, esto se obtiene fácilmente con operaciones tipo $F_{u,v}[\lambda]$ que den ceros *no solo debajo de los pivotes, sino también encima de ellos*, de modo que el 1 del pivote sea el único término no nulo de la columna. ♣

Problema 32 *Aplicar los procedimientos del teorema 12 a la matriz*

$$a=\begin{pmatrix}0&0&0&1\\0&4&2&1\\0&2&0&3\\0&0&3&0\end{pmatrix}.\ \blacklozenge$$

Solución. Los pasos son

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 4 & 2 & 1 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \end{pmatrix} \overset{1}{\to} \begin{pmatrix} 0 & 4 & 2 & 1 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{2}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{3}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & -1 & 3-\frac{2}{4} \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & -1 & \frac{5}{2} \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{4}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{5}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 & 0-3\cdot\frac{-5}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 & \frac{15}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{6}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \overset{7}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hasta aquí hemos llegado a la forma escalonada con pivotes 1, pero la matriz no está reducida y hay que eliminar $\frac{1}{2}$, $\frac{1}{4}$ y $-\frac{5}{2}$. Lo mejor es restar la fila 3 a las filas 1 y 2 tras multiplicarlas por $\frac{1}{4}$ y $-\frac{5}{2}$ quedando

$$\begin{pmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{8}{\to} \begin{pmatrix} 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

y luego restar a la fila 1 la fila 2 por $\frac{1}{2}$ y se llega a

$$\begin{pmatrix} 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{9}{\to} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \blacklozenge$$

Como se puede comprobar, estos procedimientos son muy tediosos pero hay que practicarlos lo suficiente para resolver casos sencillos y entender las ideas teóricas subyacentes. Advertimos al lector que los teoremas prueban que siempre se puede llegar a estas formas, pero la práctica dice que puede haber otra lista de operaciones elementales para llegar a ellas más cómodamente. El teorema da un procedimiento "infalible" pero no necesariamente el mejor. Al resolver un ejercicio una persona y no un ordenador, puede convenir poner primero la matriz en forma reducida; o sea, con pivotes 1 y ellos como único término no nulo en su columna, y dejar para el final el paso a forma escalonada.

Problema 33 *Poner en forma reducida y escalonada reducida la matriz*

$$\begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 2 & 2 & 0 \\ -1 & 1 & 0 & 1 \\ 2 & 4 & 4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} . \blacklozenge$$

Solución. Ya hay pivotes que valen 1, luego empezamos en el paso 2,

$$\begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 2 & 2 & 0 \\ -1 & 1 & 0 & 1 \\ 2 & 4 & 4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \overset{1}{\to} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 2 & 2 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & 4 & 4 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \overset{2}{\to} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \overset{3}{\to} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & 4 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

En $\overset{1}{\to}$ se deja la columna 4 solo con su pivote; en $\overset{2}{\to}$ se hace lo mismo con la columna 1; y en $\overset{3}{\to}$ lo mismo con la columna 2. Solo queda poner en condiciones la columna 3. Se busca primero que la *fila* 6 tenga pivote 1 y luego se opera para anular los términos restantes de esa columna. En concreto,

$$\overset{4}{\to}\begin{pmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & 4 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \overset{5}{\to} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Reordenando las filas se llega a la forma reducida. ♦

Problema 34 *Pasar las siguientes matrices a forma escalonada*

$$a = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \qquad c = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \qquad d = \begin{pmatrix} p & 1 & 1 \\ q & 1 & 1 \\ r & 1 & 1 \end{pmatrix}.$$

En d se suponen p, q, r no nulos y distintos entre sí.

Problema 35 *Escribir en forma escalonada o reducida (según se prefiera)*

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & h & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix}$$

con $h \neq 0$ en el tercer caso.

Diremos para tranquilidad del lector inquieto con tanto cálculo, que si bien hay que llegar en muchos casos hasta el final para resolver sistemas lineales, hay otras aplicaciones importantísimas fuera de este capítulo donde no hace falta saber con precisión cómo es la forma (del tipo que sea) de a, sino el *número de filas no nulas que tiene la b* (escalonada, etc.) en que se puede trasformar a. Puede pensarse, dada la diversidad de métodos y operaciones que a transformada en b escalonada, en c reducida y en d escalonada reducida, podía ser tal que b, c y d tuviesen distinto número de filas no nulas. Es de suma importancia que esto no puede suceder, aunque su demostración (que se verá mucho más adelante) es compleja. Si lo admitimos, ese número de filas no nulas, el mismo para cualquier forma (escalonada, etc.) en que transformemos a, se llamará en **rango de** a. Es imposible ver aquí la importancia y significado geométrico del concepto, pero cuando deba calcularse de modo efectivo sí que será necesario conocer, siquiera de modo aproximado, la forma escalonada, reducida, o escalonada reducida de a.

1.4.5. Solución de sistemas lineales

Una vez que se sabe cómo escribir matrices en forma escalonada, reducida, o escalonada reducida podemos resolver cualquier sistema lineal o probar que no tiene solución. Damos primero un ejemplo, aunque quizás haya lectores que prefieran ir directamente al teorema 13

Se trata de resolver con $\Bbbk = \mathbb{R}$ el sistema lineal

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ h \end{pmatrix}, \quad h \in \mathbb{R} \quad \text{con matriz ampliada} \quad \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 1 \\ 7 & 8 & 9 & h \end{array}\right).$$

El lector probablemente intentará combinar las ecuaciones de modo que desaparezcan en algunas de ellas las incógnitas. Por ejemplo,

$$\left\{\begin{array}{c} x+2y+3z=1 \\ 4x+5y+6z=1 \\ 7x+8y+9z=h \end{array}\right. \to \left\{\begin{array}{c} x+2y+3z=1 \\ -3y-6z=-3 \\ -6y-12z=h-7 \end{array}\right. \to \left\{\begin{array}{c} x+2y+3z=1 \\ -3y-6z=-3 \\ 0=h-1 \end{array}\right.$$

Deducirá que si $h-1\neq 0$ el sistema es incompatible. Supuesto $h=1$ resolverá

$$\left\{\begin{array}{l} x+2y=1-3z \\ -3y=-3+6z \end{array}\right.$$

y puede poner la solución, según la notación que prefiera como

$$\left\{\begin{array}{l} x=-1+z \\ y=1-2z \end{array}\right. \quad \text{o bien} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1+z \\ 1-2z \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Puede ahorrarse mucho trabajo si hace las siguientes operaciones, con la matriz ampliada, paralelas a las que acaba de realizar

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 1 \\ 7 & 8 & 9 & h \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -3 & -6 & -3 \\ 0 & -6 & -12 & h-7 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -3 & -6 & -3 \\ 0 & 0 & 0 & h-1 \end{array}\right)$$
$$\to \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & h-1 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & -1 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & h-1 \end{array}\right)$$

El sistema inicial tiene las mismas soluciones que

$$\left\{\begin{array}{l} x-z=-1 \\ y+2z=1 \\ 0=h-1 \end{array}\right. \quad \text{equivalente a} \quad \left\{\begin{array}{l} x=-1+z \\ y=1-2z \\ 0=h-1 \end{array}\right.$$

Si $h\neq 1$, el sistema es incompatible. Si $h=1$, la solución es

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1+z \\ 1-2z \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

El procedimiento para $ax=y$ es este. Aplicamos a la matriz ampliada $(a \mid y)$ operaciones fila de modo que la matriz a, que está dentro de $(a \mid y)$, quede en forma escalonada o escalonada reducida. ¡Cuidado! No decimos que transformemos $(a \mid y)$ en $(b \mid z)$ con $(b \mid z)$ escalonada o escalonada reducida sino que b es la que ha de estarlo. Al hacer las operaciones, que sí se hacen sobre toda $(a \mid y)$, naturalmente se modifica la última columna siendo casi siempre $z\neq y$, salvo si $y=0$ en cuyo caso $z=0$. Esto se puede hacer por los teoremas 11 o 12. Por el teorema 10 se sabe que $bx=z$ tiene las mismas soluciones que $ax=y$, y ya vimos en la sección *Matrices y sistemas en forma escalonada y reducida* cómo saber si hay solución y encontrarla en su caso para $bx=z$. Hemos probado un teorema fundamental.

Teorema 13 *El sistema $ax=y$ es compatible si y solo si al reducir $(a \mid y)$ a $(b \mid z)$ con b en la forma escalonada o escalonada reducida, no existen ecuaciones de la forma $0=z^k$ con $z^k\neq 0$. Si es compatible, $ax=y$ tiene las mismas soluciones que $bx=z$ y las soluciones se pueden determinar por sustitución ascendente. Estas soluciones dependen de forma biyectiva del número de variables libres de $bx=z$.*

Dos observaciones importantes si el sistema es homogéneo: **(a)** podemos ahorrarnos ampliar la matriz pues si la ampliamos, siempre la última columna será cero sean cuales sean las operaciones hechas; y **(b)** cuando hay estrictamente más incógnitas que ecuaciones (o sea, $m<n$) el sistema tiene al menos una solución que no es la solución 0.

Si $ax=y$ no es homogéneo, podemos preguntarnos para que valores y tendrá solución $ax=y$. El teorema 13 nos dice como responder, si bien no da una fórmula directa. Al reducir $(a \mid y)$ a $(b \mid z)$ los nuevos z aparecen como expresiones de $y^1,\ldots,y^m$. Si las filas cero de b son $b^{r+1},\ldots b^m$ las ecuaciones

$$0=z^{r+1}\left(y^1,\ldots,y^m\right),\ldots,0=z^m\left(y^1,\ldots,y^m\right)$$

nos dan exactamente los y para los que $ax=y$ es compatible. Nos dan las y en implícitas. Estas ecuaciones son banales si $y=0$ (sistema homogéneo) pues quedan $0=0$. Hacemos un ejemplo con

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \quad \text{con matriz ampliada} \quad \left(\begin{array}{ccc|c} 1 & 2 & 3 & p \\ 4 & 5 & 6 & q \\ 7 & 8 & 9 & r \end{array}\right).$$

Nos preguntamos para que valores de $(p,q,r)^\top$ el sistema tiene solución. Operamos

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & p \\ 4 & 5 & 6 & q \\ 7 & 8 & 9 & r \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -3 & -6 & q-4p \\ 0 & -6 & -12 & r-7p \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & -3 & -6 & q-4p \\ 0 & 0 & 0 & r-7p+2(q-4p) \end{array}\right)$$

Los $(p,q,r)^\top$ para los que el sistema es compatible son los que cumplen

$$r-7p+2(q-4p) = -15q+2q+r = 0 \qquad (\text{vale } (1,1,13) \text{ en particular}).$$

Como no nos piden soluciones no hay que operar más sobre la matriz.

Problema 36 *Resolver los sistemas de ecuaciones* [12]

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ 0 & i & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1+i \\ 0 \end{pmatrix},$$

suponiendo en el primero $\Bbbk = \mathbb{R}, \mathbb{Z}_5$ *(luego hay dos sistemas) y en el segundo* $\Bbbk = \mathbb{C}$.

Problema 37 *Nos dan los sistemas con* $\Bbbk = \mathbb{R}$ *y donde* h *es un parámetro*

$$\begin{pmatrix} 1 & h \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ h \end{pmatrix}, \qquad \begin{pmatrix} h & 0 & 1 \\ 0 & h & 0 \\ 1 & 0 & h \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} h \\ 0 \\ h \end{pmatrix}.$$

Determinar en qué casos es compatible y determinado.

Problema 38 *Resolver los sistemas con* $\Bbbk = \mathbb{C}$ *en el primer caso y* $\Bbbk$ *arbitrario en el segundo*

$$\begin{pmatrix} 1+i & \sqrt{2} \\ \sqrt{2} & 1-i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1-i \\ h \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 1 & h \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} h \\ 1 \end{pmatrix}.$$

Discutir en el segundo si la respuesta depende de cómo sea el cuerpo $\Bbbk$.

Problema 39 *Determinar para qué valores* p,q,r,h *tienen solución las ecuaciones* [13]

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 4 & 6 & 1 \\ 3 & 6 & 9 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \quad \begin{pmatrix} 1 & \cdots & 1 & 1 & 1 \\ 2 & \cdots & 2 & 2 & 1 \\ 3 & \cdots & 3 & 3 & 1 \\ \vdots & & \vdots & \vdots & \vdots \\ n & \cdots & n & n & 1 \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^{n-1} \\ x^n \\ x^{n+1} \end{pmatrix} = \begin{pmatrix} h \\ \vdots \\ h \\ h \\ h \end{pmatrix}.$$

1.4.6. La información que dan los sistemas lineales

Hasta ahora no hemos preocupado solamente de obtener soluciones, pero hay que añadir algunas cuestiones de importancia. Lo primero es que si el sistema es homogéneo $ax = 0$ y x_1 y x_2 son soluciones y $\lambda \in \Bbbk$, se tiene $a(x_1+x_2) = a(\lambda x_1) = 0$, luego x_1+x_2 y λx_1 son soluciones también. Un subconjunto S de $\Bbbk^n$ no vacío cuyos elementos cumplen que si $x_1, x_2 \in S$ y $\lambda \in \Bbbk$, entonces $x_1+x_2 \in S$ y $\lambda x_1 \in S$ es un caso particular de **subespacio vectorial**, concepto fundamental que se verá más adelante. Basta decir que para $\mathbb{R}^2$ y $\mathbb{R}^3$ se puede visualizar S como el origen o una recta o un plano a través del origen, pero un subespacio de $\mathbb{R}^n$ con $n \geq 4$ es imposible de visualizar. Es aún mayor la dificultad para $\Bbbk^n$ si $\Bbbk \neq \mathbb{R}$. Si el sistema no es homogéneo, digamos que es $ax = y$ con $y \neq 0$, y t es una solución *particular*, sus soluciones resultan al sumar a t todas las soluciones del sistema homogéneo asociado $ax = 0$. La consecuencia es que en $\mathbb{R}^2$ y $\mathbb{R}^3$ las soluciones de $ax = y$ forman un punto, recta o plano que no para por el origen. En los demás casos este espacio es el trasladado del espacio de soluciones de $ax = 0$, pero no

[12] En este problema y muchos más ponemos como incógnitas letras como x, y, z en vez de $x^1, x^2, \ldots$

[13] Estrictamente hablando no se pide resolverlas, pero puede y debe intentarse.

se puede visualizar. Al tratar más adelante los **subespacios afines** estudiaremos todo esto con mucho más detalle. De momento, hay que quedarse con la idea de que los espacios de soluciones son "figuras interesantes" de $\mathbb{k}^n$.

Otro concepto que comentamos muy informalmente es el del "tamaño" del espacio de soluciones si es no vacío. Sabemos que las soluciones se expresan de modo biunívoco con todas la elecciones de variables libres. Se supone que cuántas más variables libres tengamos, más grande es el espacio de soluciones. Definamos el "tamaño" del espacio de soluciones como el número de variables libres que se necesitan para expresarlo. Si busca el lector un poco más atrás el concepto de rango de a, recordará que r, el rango de a es el número de variables pivotales o dependientes que aparecen a transformar $ax = y$ en $bx = z$ con b en forma escalonada, reducida o escalonada reducida, sin que r varíe por el camino seguido para transformar a en b. El número de variables libres es n menos el número de variables pivotales; es decir, $n-r$. Conocer el rango r de a permite conocer el "tamaño" (es lo que luego llamaremos **dimensión**) del espacio de soluciones de $ax = y$ (insistimos, supuesto no vacío). El lector puede tomar como a una de las matrices que han aparecido en los problemas anteriores, consultar el cálculo hecho para transformar a en b escalonada, reducida o escalonada reducida, y cuando esté seguro de las filas no nulas que b tendrá (¡quizás lo vea sin conocer b en detalle!), conoce su rango r y la dimensión, por ejemplo, del espacio de soluciones de $ax = 0$. Puestos a especular, podríamos pensar que si el espacio de soluciones tiene dimensión o "tamaño" $n - r$, porque sus elementos se expresan con $n - r$ parámetros, cabría la posibilidad de hacer lo mismo con todavía menos parámetros. Se mostrará en su momento que esto es imposible con "fórmulas razonables".

Hasta ahora las matrices eran objetos algebraicos semejantes en muchos aspectos a los números y que servían para estudiar las ecuaciones lineales. Hay otra utilidad de las matrices: definir **funciones lineales**. Si $a \in \mathbb{k}^{m\times n}$ definimos $L : \mathbb{k}^n \to \mathbb{k}^m$ por $L(x) = ax$. Obsérvese que se da la vuelta al orden de las letras (a es $m \times n$ y L va de $\mathbb{k}^n$ en $\mathbb{k}^m$) y que a y x son multiplicables pues $x \in \mathbb{k}^n = \mathbb{k}^{n\times 1}$. [14] Dada esta L nos podemos preguntar si L es inyectiva, suprayectiva o biyectiva. Lo primero equivale a que si se tiene $L(x) = y = ax$ solo un x *como máximo* pueda cumplir la condición; o sea, $ax = y$ es para todo y un sistema determinado si es compatible. Lo segundo equivale a que para todo y haya *como mínimo* un x tal que $L(x) = y = ax$; o sea, que el sistema sea para todo y compatible, aunque quizás indeterminado. Lo tercero equivale a que para todo y el sistema $ax = y$ sea compatible y determinado.

Todo esto vale como anuncio de que hay mucho que ganar con una estructura teórica más amplia que la vista hasta ahora, quizás más difícil de comprender, pero más poderosa. Es la materia de los dos capítulos siguientes. Si solo nos preocupara el rigor lógico, se podría exponer antes de nada los conceptos de espacio vectorial, función lineal, dimensión, etc. y obtener muchos de los teoremas de este capítulo como corolarios de teoremas generales para estos entes.

1.5. Matrices elementales e inversión de matrices

Nos va a ser muy útil determinar que si $a \in \mathbb{k}^{m\times n}$ es una matriz y $b = F(a)$, siendo F una operación elemental, la matriz b se obtiene a partir de a en la forma $b = pa$, siendo $p \in \mathbb{k}^{m\times m}$ una matriz invertible que dependerá, como es natural, de F. Definimos las **matrices elementales** $m \times m$ como las que se obtienen al aplicar una operación elemental F a la matriz unidad I_m. Obsérvese que son matrices *cuadradas*. Tenemos las matrices elementales ϕ definidas por

$$F_{uv}(I_m) = \phi_{uv}, \qquad F_{uv}[\lambda](I_m) = \phi_{uv}[\lambda], \qquad \phi_w[\mu](I_m) = \phi_w[\mu].$$

Así, en el segundo caso, $\phi_{uv}[\lambda]$ se obtiene sustituyendo en la matriz unidad I su fila u por la suma a esta fila u de la fila v multiplicada por λ y dejando las otras invariables. Por ejemplo, para $m = 5$,

$$\phi_{25} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad \phi_{13}[\lambda] = \begin{pmatrix} 1 & 0 & \lambda & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \phi_4(1/2) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

El siguiente teorema técnico es fundamental.

[14] Quizás recuerde el lector un antiguo comentario en que se dijo que los vectores de $\mathbb{R}^2$ y $\mathbb{R}^3$ y de $\mathbb{k}^n$ en general se iban a escribir como vectores *columna*. Aquí está la razón, que es poder escribir cómodamente las funciones lineales.

Teorema 14 *Sean $a \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{n\times p}$ y $E = F_h \circ F_{h-1} \circ \cdots \circ F_2 \circ F_1$ una composición de operaciones elementales sobre matrices con m filas. Se tiene que $E(ab) = E(a)\, b$; o sea, es lo mismo aplicar E sobre el producto que aplicar E sobre el primer factor y multiplicar la matriz que resulte por el segundo factor.*

Demostración. Probamos primero el teorema si $E = F$; o sea si en E consiste en una sola transformación elemental. Supongamos (suele ser el caso más complicado) que $E = F = F_{uv}[\lambda]$. Sea i un índice de fila. Según sea $i \neq u$ o $i = u$ tenemos

$$[E(ab)]_j^i \overset{*}{=} (ab)_j^i = \sum_{k=1}^{n} a_k^i b_j^k \overset{*}{=} \sum_{k=1}^{n} [E(a)]_k^i\, b_j^k = [E(a)\cdot b]_j^i,$$

$$[E(ab)]_j^u \overset{*}{=} (ab)_j^u + \lambda (ab)_j^v = \sum_{k=1}^{n} a_k^u b_j^k + \lambda \sum_{k=1}^{n} a_k^v b_j^k = \sum_{k=1}^{n} \left(a_k^u + \lambda a_k^v\right) b_j^k \overset{*}{=} \sum_{k=1}^{n} [E(a)]_k^u\, b_j^k = [E(a)\cdot b]_j^u.$$

En cada $\overset{*}{=}$ se ha usado la definición de $E = F_{uv}[\lambda]$ que dice que para toda matriz c (se aplicará a a y ab) se tiene $E(c)^i = c^i$ si $i \neq c$ y $E(c)^u = c^u + \lambda c^v$ si $i = u$. Los otros casos de F quedan para el lector.

Probamos ahora el teorema si $E = F_h \circ \cdots \circ F_1$. El cálculo usa lo anterior y es

$$\begin{aligned} E(ab) &= (F_h \circ \ldots \circ F_2)(F_1(ab)) = (F_h \circ \ldots \circ F_2)(F_1(a)\cdot b) \\ &= (F_h \circ \ldots \circ F_3)(F_2(F_1(a)\cdot b)) = (F_h \circ \ldots \circ F_3)((F_2 \circ F_1)(a)\cdot b) \\ &= \ldots\ldots = F_h((F_{h-1} \circ \cdots \circ F_2 \circ F_1)(a)\cdot b) = (F_h \circ F_{h-1} \circ \cdots \circ F_2 \circ F_1)(a)\cdot b = E(a)\cdot b \end{aligned}$$

Lo que se ha hechos es para $c = a, F_1(a), \ldots, (F_{h-1} \circ \cdots \circ F_2 \circ F_1)(a)$ aplicar $F(cb) = F(c)\, b$. ♣

De este teorema se sigue otro de gran importancia.

Teorema 15 *Si E es composición de operaciones elementales, $E = F_h \circ \cdots \circ F_1$, la matriz $E(a)$ se obtiene multiplicando a la izquierda de a por las matrices elementales $\phi_i = F_i(I_n)$; más concretamente,*

$$E(a) = (F_h \circ \cdots \circ F_1)(a) = \phi_h \cdot \cdots \cdot \phi_1 \cdot a.$$

Demostración. Evidentemente, $E(a) = E(I_m \cdot a) = E(I_m)\cdot a$ y basta ver que $E(I_m) = \phi_h \cdot \cdots \cdot \phi_1$. Esto es inmediato porque

$$\begin{aligned} E(I_m) &= (F_h \circ \cdots \circ F_2)(F_1(I)) = (F_h \circ \cdots \circ F_2)(\phi_1) = (F_h \circ \cdots \circ F_3)(F_2(\phi_1)) \\ &= (F_h \circ \cdots \circ F_3)(\phi_2 \cdot \phi_1) = (F_h \circ \cdots \circ F_4)(F_3(\phi_2 \cdot \phi_1)) \\ &= (F_h \circ \cdots \circ F_4)(\phi_3 \cdot \phi_2 \cdot \phi_1) = \ldots\ldots = F_h(\phi_{h-1} \cdot \cdots \cdot \phi_2 \cdot \phi_1) = \phi_h \cdot \cdots \cdot \phi_1, \end{aligned}$$

como queríamos demostrar. ♣

Si F_1 y F_2 son operaciones elementales inversa una de otra, se tiene

$$I = (F_2 \circ F_1)(I) = \phi_2 \cdot \phi_1, \quad I = (F_1 \circ F_2)(I) = \phi_1 \cdot \phi_2$$

luego ϕ_1 y ϕ_2 son inversa una de otra y, en particular, *toda matriz elemental es invertible*. Hemos probado en el problema 18 que el producto de matrices invertibles también es invertible, luego *todo producto de matrices elementales es invertible*.

Para resolver el sistema $ax = y$ lo que hemos hecho ha sido aplicar una sucesión $F_h, \cdots, F_2, F_1$ de operaciones elementales a a y, por el teorema 15, cada una supone premultiplicar a por una matriz elemental ϕ_j. Si $p = \phi_h \cdot \cdots \cdot \phi_2 \cdot \phi_1$ es su producto, tenemos $pa = py$, y cuando hemos dicho que se transformaba $ax = y$ en $bx = z$, con b en forma escalonada, reducida o escalonada reducida, lo que hacíamos era llegar a $b = pa$ y $z = py$. Hemos probado otro teorema fundamental.

Teorema 16 *Para cualquier $a \in \Bbbk^{m\times n}$ existe $p = \phi_h \cdot \cdots \cdot \phi_2 \cdot \phi_1$ producto de matrices elementales, tales que $b = pa$ está, según queramos, en forma escalonada, reducida o escalonada reducida.*

Al resolver $ax = y$ no nos ha preocupado conocer p, pero tiene su interés en bastantes casos. Supongamos con más generalidad que nos interesa aplicar la operación fila F a a para llegar a b, luego $b = F(a)$. Sabemos que $F(a) = F(Ia) = F(I)\, a = pa$ y que $p = F(I)$. Para no "perder" p lo que hacemos es

poner juntas a e I en $(a \mid I)$ y hacer en esta matriz más ancha la operación F. Pasamos de $(a \mid I)$ a $(F(a) \mid F(I)) = (b \mid p) = (pa \mid p)$. Por ejemplo, si $F = F_{21}[-4]$,

$$(a \mid I) = \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4-4 & 5-4\cdot 2 & 6-4\cdot 3 & 0-4\cdot 1 & 1-4\cdot 0 & 0-4\cdot 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right),$$

$$(b \mid p) = (F(a) \mid F(I)) = (pa \mid p) = \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right).$$

Para mayor seguridad el ordenador comprueba que

$$pa = \left(\begin{array}{ccc} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)\left(\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}\right) = \left(\begin{array}{ccc} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{array}\right) = b.$$

Evidentemente, si se hacen operaciones $F_1, \ldots, F_h$, se tiene

$$(a \mid I) \overset{F_1}{\to} \left(a_{(1)} \mid p_{(1)}\right) \overset{F_2}{\to} \left(a_{(2)} \mid p_{(2)}\right) \overset{F_3}{\to} \ldots\ldots \overset{F_h}{\to} \left(a_{(h)} \mid p_{(h)}\right)$$

y $p_{(h)} \cdot a = a_{(h)}$.

Nos interesa sobre todo el caso en el que a es cuadrada $m \times m$. En este caso, el teorema 11 nos dice que se puede transformar a a forma escalonada reducida e con operaciones fila, que es como decir que existe p, producto de matrices elementales tal que $pa = e$. Esta matriz p va a ser a^{-1}, la inversa de a, si existe. Esto y más dice otro teorema fundamental.

Teorema 17 *Para $a \in \Bbbk^{m\times m}$ son equivalentes las siguientes propiedades*

1. *La condición $av = 0$ con $v \in \Bbbk^m$ solo es posible si $v = 0$.*
2. *Al transformarla en e escalonada reducida con $pa = e$ y p producto de matrices elementales, tiene que ser $e = I_m$.*
3. *a es producto de matrices elementales*
4. *a es invertible.*
5. *Existen matrices $b, c \in \Bbbk^{m\times m}$ tales que o bien $ab = I_m$ o bien $ca = I_m$.*

*Si b o c cumplen **5**, ha de ser $b = c = a^{-1}$.*

Demostración. El extraño orden de las condiciones es para facilitar una demostración circular.

1 $\Longrightarrow$ **2**. Si $pa = e$ y $e \neq I_m$, al aplicar e a $v = (0, \ldots, 0, 1)^\top$, se tiene $ev = 0$. Obtenemos $pav = ev = 0$ y, como p es invertible, $av = p^{-1}pav = p^{-1}0 = 0$. Hay contradicción porque $v \neq 0$.

2 $\Longrightarrow$ **3**. Como $pa = I$ y $p = \phi_h \cdots \phi_1$, obtenemos que $a = p^{-1} = \phi_1^{-1} \cdots \phi_h^{-1}$, pues las matrices elementales son invertibles. Además con sus inversas también elementales, y resulta **3**.

3 $\Longrightarrow$ **4**. Como el producto de matrices invertibles es invertible y las elementales lo son, a es invertible.

4 $\Longrightarrow$ **5**. Es obvio que a^{-1} sirve como b y c.

5 $\Longrightarrow$ **1**. Si se tiene $ca = I$ y $av = 0$, resulta $0 = c0 = cav = Iv = v$. Si se tiene $ab = I$, acabamos de probar que b cumple **5** $\Longrightarrow$ **1** $\Longrightarrow$ **2** $\Longrightarrow$ **3** $\Longrightarrow$ **4**, luego b es invertible. De $ab = I$ resulta $a = abb^{-1} = b^{-1}$, y a tiene como inversa b.

Si $ca = I$, como a es invertible, $c = caa^{-1} = Ia^{-1} = a^{-1}$. Análogamente, si $ab = I$, $b = a^{-1}$. ♣

Lo que aprovechamos del teorema 17 para el cálculo efectivo de la inversa de a, o para saber que no existe, se explica ahora. Hacemos en a operaciones fila hasta transformarla en escalonada reducida e. Es esencial conocer la matriz p tal que $pa = e$. El teorema dice que $pa = I$ equivale a que exista a^{-1} y sea p. Suele ser más fácil ver que a no es invertible porque si al hacer operaciones fila para transformarla en escalonada reducida vemos que no podrá ser $e = I$ (por ejemplo, porque aun sin concluir el cálculo aparece una fila de ceros), ya sabemos que a no será invertible.

Problema 40 *Calcular la inversa, según el valor de h, de*

$$a = \begin{pmatrix} 1 & 0 & 1 \\ 0 & h & 0 \\ -1 & 0 & 1 \end{pmatrix}. \blacklozenge$$

Solución. Las operaciones, suponiendo $h \neq 0$, son

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & h & 0 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & h & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{h} & 0 \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right) \to$$

.

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1-\frac{1}{2} & 0 & 0-\frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{h} & 0 \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right) = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{h} & 0 \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right).$$

Efectivamente, el ordenador verifica que

$$pa = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & \frac{1}{h} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & h & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

El caso $h = 0$ no es invertible porque la segunda fila de a es nula. $\blacklozenge$

Problema 41 *Sean $c = \cos\theta$ y $s = \sin\theta$. Determinar los casos en los que*

$$a = \begin{pmatrix} c & s \\ s & c \end{pmatrix}$$

no tiene inversa. $\blacklozenge$

Solución. En el caso $\cos\theta = 0$ es $\sin\theta = \pm 1$. Entonces, aplicando a a la permutación de sus filas queda una matriz diagonal (s, s) no nula, así que el caso $\cos\theta = 0$ da a invertible. Sea en adelante $c = \cos\theta \neq 0$. Entonces

$$\begin{pmatrix} c & s \\ s & c \end{pmatrix} \to \begin{pmatrix} c & s \\ s - \frac{s}{c}c & c - \frac{s}{c}s \end{pmatrix} = \begin{pmatrix} c & s \\ 0 & \frac{c^2 - s^2}{c} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ 0 & \frac{\cos 2\theta}{\cos\theta} \end{pmatrix}.$$

Si $\cos 2\theta = 0$ hay una fila nula y aunque sigamos trabajando, e tendrá su segunda fila nula, luego $e \neq I$ y a no es invertible. Si $\cos 2\theta \neq 0$ las operaciones llevan a $e = I$ luego a es invertible. El resumen es que solo los casos con $\cos 2\theta = 0$ pero $\cos\theta \neq 0$ da a no invertible. El cálculo se ha abreviado porque no pedían a^{-1}, pero el lector puede hacerlo y verá que a^{-1} es sencilla. $\blacklozenge$

El lector se puede poner todos los problemas que quiera de cálculo de inversas y de soluciones de sistemas lineales. Puede elegir matrices elementales $\phi_1, \ldots, \phi_h$, calcular $a = \phi_1 \cdot \cdots \cdot \phi_h$ y preguntar o preguntarse cuánto vale a^{-1}. La respuesta es $a^{-1} = \phi_h^{-1} \cdot \cdots \cdot \phi_1^{-1}$. Evidentemente es muy pesado multiplicar matrices, pero tiene la ventaja de que se sabe de antemano la solución. Damos como ejemplo

$$\phi_3 = F_{12}(I) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \phi_2 = F_{31}[3](I) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}, \quad \phi_1 = F_1[2](I) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$a = \phi_3\phi_2\phi_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 0 \\ 6 & 0 & 1 \end{pmatrix},$$

$$a^{-1} = \phi_1^{-1}\phi_2^{-1}\phi_3^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 1 & 0 & 0 \\ 0 & -3 & 1 \end{pmatrix}.$$

Se puede hacer algo parecido con sistemas lineales. Se pone un sistema "fácil" $bx = z$ con b y z conocidos. Si $b \in \Bbbk^{m\times n}$ se toma $c \in \Bbbk^{m\times m}$ invertible. Claramente $bx = z$ equivale a $cbx = cz$ así que definiendo $a = cb$ e $y = cz$ tenemos un sistema $ax = y$ "difícil" cuyas soluciones ya sabemos pues son las mismas de $bx = z$.

Problema 42 *Dadas las matrices*

$$a=\begin{pmatrix}1&2&3\\0&1&2\\0&0&1\end{pmatrix},\quad b=\begin{pmatrix}0&1&1\\1&0&1\\1&0&0\end{pmatrix},\quad c=\begin{pmatrix}1&h&1\\0&1&h\\1&0&1\end{pmatrix},$$

calcular las inversas de a y b y los valores de h para los que c es invertible.

El siguiente problema no debe olvidarse, pues hace que las matrices invertibles más fáciles de detectar sean las triangulares[15] con términos no nulos en la diagonal principal.

Problema 43 *Sea a una matriz triangular superior. Probar que a es invertible si y solo si todos los a_i^i son no nulos. Si esto sucede la inversa será también triangular superior y los términos de la diagonal principal de a^{-1} serán $\left(a_1^1\right)^{-1},\left(a_2^2\right)^{-1},\dots\left(a_n^n\right)^{-1}$. No se pide calcular la inversa sino dar condiciones suficientes, y también necesarias, para su existencia más algunos aspectos de a^{-1}.*

Si solo nos interesa saber si a es invertible pero no se necesita la forma exacta de a^{-1}, se puede evitar mucho trabajo en los cálculos anteriores. Supongamos que hay operaciones fila $F_1,F_2,\dots,F_h$ tales que $F_h\circ\cdots\circ F_1(a)=t$, siendo t triangular superior. Si las ϕ_j son las matrices elementales $F_j(I)$ hemos obtenido

$$t=\phi_h\cdot\cdots\cdot\phi_1\cdot a=p\cdot a,\qquad p=\phi_h\cdot\cdots\cdot\phi_1$$

con p invertible ya que es producto de matrices elementales. Claramente, a es invertible si y solo si lo es t. En efecto. $t=pa$, luego si p es invertible, lo es t como producto de dos matrices invertibles. Recíprocamente. si t es invertible, $a=p^{-1}t$ y a es invertible como producto de dos matrices invertibles.

Problema 44 *Calcular la inversa de*

$$a=\begin{pmatrix}1&1&\cdots&1&1\\0&1&\cdots&1&1\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&1&1\\0&0&\cdots&0&1\end{pmatrix}\in\mathbb{k}^{n\times n}.\ \blacklozenge$$

Solución. Primeramente (no lo hacemos) se hacen unos tanteos para $n=2,3$ y quizás 4, y se *intuye* la respuesta

$$b=a^{-1}=\begin{pmatrix}1&-1&\cdots&0&0\\0&1&\cdots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&1&-1\\0&0&\cdots&0&1\end{pmatrix}$$

En b hay 1 en la diagonal, -1 sobre la diagonal, y el resto 0. Ha de verificarse $ab=I_n$ pero no es una verificación trivial y hay que organizarse. Recordamos que $(ab)_j^i=a^i\cdot b_j$ y debemos probar que es δ_j^i. Por propiedades *generales* de matrices triangulares, ab es triangular y $(ab)_i^i=a_i^i\cdot b_i^i$, luego en nuestro caso *particular* solo queda ver que $a^i\cdot b_j=0$ para $i<j$ (los términos sobre la diagonal de ab son nulos). Indicando posición con un índice entre paréntesis escribimos para $i<j$

$$a^i=\left(0,\dots,0,\overset{(i)}{1},1,\dots,\overset{(j-1)}{1},\overset{(j)}{1},1,\dots,1\right),\quad (b_j)^\top=\left(0,\dots,0,0,\dots,0,\overset{(j-1)}{-1},\overset{(j)}{1},0,\dots,0\right),$$

y queda $a^i\cdot b_j=1\cdot(-1)+1\cdot 1=0$. $\blacklozenge$

Problema 45 *Calcular la inversa de*

$$a=\begin{pmatrix}1&\lambda_2&\lambda_3&\cdots&\lambda_{n-1}&\lambda_n\\0&1&0&\cdots&0&0\\0&0&1&\cdots&0&0\\&&&&&\\0&0&0&\ddots&0&0\\0&0&0&\cdots&1&0\\0&0&0&\cdots&0&1\end{pmatrix}.$$

[15] Tratamos las triangulares superiores, pero hay resultados análogos para las triangulares inferiores.

La matriz a es como la matriz unidad solo que en la primera fila, en lugar de poner ceros tras el uno van los λ. ♦

Solución. Si se quiere atacar el problema con el algoritmo, es fácil porque

$$\left(\begin{array}{cccccc|cccccc} 1 & \lambda_2 & \lambda_3 & \cdots & \lambda_{n-1} & \lambda_n & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array}\right)$$

restando a la fila 1 las filas $2, \ldots, n$ multiplicadas por $\lambda_2, \ldots, \lambda_n$, nos da

$$\left(\begin{array}{cccccc|cccccc} 1 & 0 & 0 & \cdots & 0 & 0 & 1 & -\lambda_2 & -\lambda_3 & \cdots & -\lambda_{n-1} & -\lambda_n \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 1 \end{array}\right)$$

La inversa es la última matriz.

Hay sin embargo una solución más interesante en la que no se piensa al principio y es observar que

$$a = \begin{pmatrix} 1 & \lambda_2 & \lambda_3 & \cdots & \lambda_{n-1} & \lambda_n \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} = \left(F_{12}\left[\lambda_2\right] \circ F_{13}\left[\lambda_3\right] \circ \cdots \circ F_{1n}\left[\lambda_n\right]\right)(I)$$

y por consiguiente,

$$a^{-1} = \left(F_{1n}\left[-\lambda_n\right] \circ \cdots \circ F_{13}\left[-\lambda_3\right] \circ F_{12}\left[-\lambda_2\right]\right)(I) = \begin{pmatrix} 1 & -\lambda_2 & -\lambda_3 & \cdots & -\lambda_{n-1} & -\lambda_n \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \blacklozenge$$

Problema 46 *Sea $p \in \Bbbk$. Calcular la inversa de*

$$a = \begin{pmatrix} 1 & p & p^2 & p^3 & p^4 \\ 0 & 1 & p & p^2 & p^3 \\ 0 & 0 & 1 & p & p^2 \\ 0 & 0 & 0 & 1 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

1.6. Operaciones con columnas

Las operaciones elementales hechas con las filas de las matrices tienen una teoría paralela con las columnas. De momento no es necesaria pero sí hará falta al diagonalizar formas cuadráticas. Para no interrumpir demasiado la exposición damos solo un bosquejo y el lector decidirá con qué grado de detalle, demostraciones incluidas, quiere completarlo. Todo es análogo excepto que el operar con filas tiene un reflejo en sistemas de ecuaciones lineales y el operar con columnas, de momento, no tiene aplicación.

Si tenemos una matriz $a \in \Bbbk^{m\times n}$ podemos hacer con ella tres **operaciones elementales** *con sus columnas*, que se denotarán con K_{uv}, $K_{uv}\left[\lambda\right]$ y $K_w\left[\mu\right]$ y tienen definición análoga a las operaciones con filas:

1. La matriz $b = K_{uv}(a)$ se obtiene permutando las columnas u y v dejando invariables las demás. Con símbolos,
$$b_u = a_v, \qquad b_v = a_u, \qquad b_w = a_w \quad \text{si } w \neq u, v.$$

2. La matriz $b = K_{uv}[\lambda](a)$ se obtiene sustituyendo la columna u por la columna u más λ por la columna v, siendo $\lambda \in \Bbbk$, y dejando invariables las demás, incluida la propia columna v. Con símbolos,
$$b_u = a_u + \lambda a_v, \qquad b_w = a_w \quad \text{si } w \neq u.$$

3. La matriz $b = K_w[\mu](a)$ se obtiene sustituyendo la columna w por la columna w multiplicada por $\mu \in \Bbbk$ *no nulo* y dejando invariables las demás. Con símbolos,
$$b_w = \mu a_w, \qquad b^u = a^u \quad \text{si } w \neq u.$$

Como ejemplos, para
$$a = \begin{pmatrix} 1 & 1 & -2 \\ 1 & -1 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix},$$
$$K_{13}(a) = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 2 \\ -1 & 1 & 0 \end{pmatrix}, \quad K_{31}[-2](a) = \begin{pmatrix} 1 & 1 & -4 \\ 1 & -1 & -1 \\ 2 & 1 & -3 \\ 0 & 1 & -1 \end{pmatrix}, \quad K_2[3](a) = \begin{pmatrix} 1 & 3 & -2 \\ 1 & -3 & 1 \\ 2 & 3 & 1 \\ 0 & 3 & -1 \end{pmatrix}$$

Tenemos pues en el espacio $\Bbbk^{m\times n}$ definidos tres tipos de *funciones* de él en si mismo, denotadas por $K_\bullet[\bullet]$. Cada una de estas funciones tiene una función inversa. Concretamente,
$$(K_{uv})^{-1} = K_{vu} = K_{uv}, \qquad (K_{uv}[\lambda])^{-1} = K_{uv}[-\lambda], \qquad (K_w[\mu])^{-1} = K_w[1/\mu].$$

Si K es una operación elemental columna, las matrices $K(I_n)$ que resultan al aplicar a la matriz unidad I_n esa operación, se llaman también **matrices elementales.** Obsérvese que son matrices cuadradas pero $n \times n$ (las ϕ eran $m \times m$). Tenemos las matrices elementales (denotadas con la letra κ "kappa")
$$K_{uv}(I_n) = \kappa_{uv}, \qquad K_{uv}[\lambda](I_n) = \kappa_{uv}[\lambda], \qquad K_w[\mu](I_m) = \kappa_w[\mu].$$

Así, en el segundo caso, $K_{uv}[\lambda]$ se obtiene sustituyendo en la matriz unidad I su *columna* u por la suma a esta *columna* u de la *columna* v multiplicada por λ y dejando las otras invariables. Tal como sucede con las filas, estas matrices son invertibles siendo la inversa de $\kappa = (I)K$ la matriz $(I)K^{-1}$ siendo K^{-1} la operación inversa de κ.

Para $a \in \Bbbk^{m\times n}$ diremos que a^p_j es el **pivote de la columna** j si es el primer coeficiente no nulo de esa columna (leyendo de arriba abajo). Se entiende que si $a^1_j = \ldots = a^m_j = 0$, la columna j no tiene pivote. Diremos que a está en **forma escalonada** si

1E Las columnas nulas son las últimas; digamos $a_{r+1}, \ldots, a_n$ para cierto r con $1 \leq r \leq n$.

2E Si $a_1, \ldots, a_r$ son las columnas no nulas, y $a^{q_1}_1, a^{q_2}_2, \ldots, a^{q_r}_r$ sus pivotes, se tiene que $1 \leq q_1 < q_2 < \ldots < q_r \leq m$. (Al ir hacia la derecha, el pivote está cada vez más bajo.)

Damos como ejemplos,
$$\begin{pmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ 1 & 4 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 5 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 \\ 4 & 2 & 0 \\ 6 & 4 & 2 \\ 8 & 6 & 4 \end{pmatrix}.$$

Se dice que a está en **forma reducida** si

1R Los pivotes (por definición, no nulos) valen todos 1.

2R Si un pivote está en la fila i, este es el único término no nulo de esa fila.

Damos como ejemplos

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \\ 1 & 0 & 0 \\ 4 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}.$$

Diremos que a está en forma **escalonada reducida** si se cumplen las cuatro condiciones. Como ejemplos,

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 8 & 6 & 4 \end{pmatrix}.$$

Sin embargo, solo uno de los siete ejemplos de matrices en forma sea escalonada sea reducida está en forma escalonada reducida. ¿Cuál es?

Es muy probable que el lector haya caído en la cuenta que las líneas anteriores son la trascripción de las que definen "escalonada", "reducida" y "escalonada reducida" por *filas*, cambiando "fila" por "columna" y que los ejemplos son los traspuestos de los que allí había. Se puede abreviar diciendo, por ejemplo, que a está escalonada por columnas si $a^\top$ lo está por filas. Esto ahorra papel, pero es mejor detallar definiciones. En todo caso, la idea de que las definiciones "salen por trasposición" es importante y no debe olvidarse.

Teorema 18 *Con operaciones elementales de columna toda matriz a puede transformarse en forma escalonada o escalonada reducida por columnas.*

Teorema 19 *Sean $a \in \mathbb{R}^{m\times n}$ y $b \in \mathbb{R}^{n\times p}$ y E una composición de operaciones elementales de columna sobre matrices con p columnas. Se tiene que $(ab)\,E = a\,(b)\,E$; o sea, es lo mismo aplicar E sobre el producto que aplicar E sobre el segundo factor y multiplicar la matriz que resulte por el primer factor.*

Teorema 20 *Si E es composición de operaciones elementales de columna $E = (K_1 \circ \cdots \circ K_h)$, la matriz $(a)\,E$ se obtiene multiplicando a la derecha de a por las matrices elementales $\kappa_i = K_i\,(I_m)$; más concretamente,*

$$(a)\,E = (a)\,(K_1 \circ \cdots \circ K_h) = a \cdot \kappa_1 \cdot \cdots \cdot \kappa_h$$

Si se prefiere, se puede calcular la inversa de a con operaciones columna. El fundamento es que si con operaciones $K_1, \ldots, K_h$ se llega a $I = (a)\,(K_1 \circ \cdots \circ K_h) = a\cdot\kappa_1\cdot\cdots\cdot\kappa_h = a\cdot p$ el producto $\kappa_1\cdot\cdots\cdot\kappa_h = p$ es la inversa de a. Ya dijimos al principio que se necesita también $p\cdot a = I$ pero esto se deriva de $a\cdot p = I$. Si la forma escalonada reducida $a\cdot p = r$ tiene *columnas* nulas, a no es invertible.

Problema 47 *Pasar las siguientes matrices a forma escalonada reducida por columnas*

$$a = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \qquad c = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \qquad d = \begin{pmatrix} p & 1 & 1 \\ q & 1 & 1 \\ r & 1 & 1 \end{pmatrix}.$$

En d se suponen p, q, r no nulos y distintos entre sí.

Problema 48 *Como en el problema anterior para las matrices* complejas

$$a = \begin{pmatrix} 1 & 0 & 0 \\ i & 1 & 0 \\ 1+i & i & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 & 1+i & 1-i \\ 0 & 1 & 1+i \end{pmatrix}, \qquad c = \begin{pmatrix} i & 1 & 0 & 0 \\ 1 & i & 1 & 0 \\ 0 & 1 & i & 1 \\ 0 & 0 & 1 & i \end{pmatrix}$$

excepto c, que como es largo, basta dejarla en forma escalonada.

1.7. El principio de inducción

El **principio de inducción** recoge una idea muy intuitiva como método de demostración. Si tenemos una sucesión de proposiciones $P_1, P_2, \ldots, P_n, \ldots$ (posiblemente infinita), sabemos que la proposición P_1 es cierta y somos capaces de probar que la veracidad de cada proposición implica la de la siguiente, pensaremos sin duda que P_n es cierta para cualquier valor de n. Al fundamentar las Matemáticas se consideró conveniente *definir como un axioma* (no como "sentido común") que esto era cierto. Hay que subrayar que lo que da el principio de inducción es un método de demostración e incluso de construcción, una herramienta, pero la destreza en el manejo de la herramienta requiere conocimiento del área a la que se va a aplicar, y esto no lo da el principio.

Principio de Inducción. *Supongamos que para cada $n \in \mathbb{N}$ hay una proposición, propiedad, condición, fórmula, etc. P_n* que puede ser cierta o falsa. *Se cumplirá que P_n es cierta para todo $n \in \mathbb{N}$ si se verifican las siguientes condiciones*

(H1) *P_1 es cierta.*

(H2) Para todo $n \in \mathbb{N}$ se cumple que si *P_n es cierta, P_{n+1} es cierta también.*

El caso $n = 1$ se suele llamar el **caso base** y la demostración de **H2** el **paso inductivo**, siendo en ese paso P_n la **hipótesis inductiva** y P_{n+1} la tesis o conclusión. Hay varios errores en la comprensión y aplicación de este principio y para evitarlos presentamos unos comentarios.

1. De un modo atípico hemos puesto en el enunciado la conclusión antes que las hipótesis. La conclusión es que P_n es cierta para todo n y las hipótesis son **H1** y **H2**. El trabajo está en verificar **H1** y **H2** y el principio de inducción no dice cómo. Si las P_n son, por ejemplo, de naturaleza algebraica, habrá que saber Álgebra.

2. Suele ser muy fácil comprobar que P_1 es cierta y se salta con frecuencia por obvio. Sin embargo es esencial. Si la proposición P_n es $n = n + 1$, y uno da por hecho que $1 = 2$ es cierta, como $a = b$ implica que $a + 1 = b + 1$, tendrá que P_n implica P_{n+1} y habrá "probado" que todos los números son iguales. La demostración falla no porque falle **H2**, sino porque falla **H1**.

3. En **H2** no hay una sola comprobación sino varias, quizás infinitas. Si en algún momento se rompe la cadena de comprobaciones, por ejemplo, P_4 no implica P_5, nos falla **H2** donde se dice muy claro "para todo $n \in \mathbb{N}\ldots$". Lo más que obtenemos es que P_1, P_2, P_3 y P_4 son ciertas. Sobre $P_5, P_6, \ldots$ no sabemos nada. Podrían ser incluso ciertas, pero lo que aquí cuenta es que no llegamos a ello por el principio de inducción.

4. Hay que observar **H2** no pide probar que P_n sea cierta sino que es cierta una *afirmación condicional,* y esta afirmación es que *si suponemos cierta P_n podemos deducir que P_{n+1} es cierta también.* En la "demostración" de que $n = n+1$ para todo n no se ha hecho nada incorrecto por lo que respecta a **H2** porque *si fuera cierto que $n = n+1$, se deduciría que $n+1 = n+2$.* El problema, repetimos, está en que falla **H1**.

5. La proposición a probar tiene que depender de un $n \in \mathbb{N} = 1, 2, 3, \ldots$, aunque en la formulación de P_n pueden aparecer variables que no sean números naturales. Puede incluso que haya en la formulación de una proposición $P(m, n, \ldots, x, y, \ldots)$ muchas variables y entre ellas varios números naturales, debiéndose decidir respecto a cual, o a una combinación obtenida de ellos, se va a hacer la inducción. Un ejemplo. Sea X un conjunto con n elementos y $0 \leq m \leq n$. El número de subconjuntos de X con m elementos es

$$\binom{m}{n} = \frac{n!}{m!\,(n-m)!}.$$

(Se cuentan $\emptyset$ y el propio X.) No entramos ahora en la demostración, que se puede hacer por inducción, pero ¿respecto a n o respecto a m? Recomendamos al lector que escriba P_n y Q_m y, antes de intentar demostrar nada, vea que el problema de la implicación de P_{n+1} por P_n no es el mismo que el de la implicación de Q_{m+1} por Q_m. Para ir a lo más fácil, que es **H1**, en la inducción sobre n solo hay que ver que un conjunto X con un elemento no tiene más que un subconjunto con cero elementos (que es $\emptyset$) y otro con un elemento, que es el propio X y que $\binom{1}{0} = \binom{1}{1} = 1$, así que

P_1 es cierta. En la inducción sobre m, la proposición Q_1 dice que fijado X con n elementos y n arbitrario, el número de subconjuntos de X con un elemento es n, que es precisamente $\binom{n}{1}$. Tanto P_1 como Q_1 son ciertas, pero ¿qué método inductivo conviene elegir?

6. Hay proposiciones P_r donde r no es un número natural, sino un número real. Demostrar que P_r es cierta para todo r no se puede hacer por inducción. Vale como ejemplo $r^2 \leq 2^r$, $r > 0$. Una excepción sería que P_n estuviera enunciada para $n = 0, 1, 2, \ldots$ o incluso para $n = -k, -(k-1), \ldots, -1, 0, 1, 2, \ldots$ con $k \in \mathbb{N}$, donde el caso base es P_0 o P_{-k}.

Damos un ejemplo. Recordamos que la **factorial de** $n \in \mathbb{N}$ es el número $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1) \cdot n$. Por ejemplo, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$. Por definición $0! = 1$ (*no olvidarlo*). Para probar por inducción que $2^n \leq (n+1)!$, que sería P_n, se ve primero que es cierta para $n = 1$ y se supone que es cierta $P_n : 2^n \leq (n+1)!$. Entonces

$$2^{n+1} = 2 \cdot 2^n \overset{1}{\leq} 2 \cdot (n+1)! \overset{2}{\leq} (n+2) \cdot (n+1)! = (n+2)! = ((n+1)+1)!$$

Se ha usado en $\overset{1}{\leq}$ la llamada **hipótesis inductiva**, que es la validez de $2^n \leq (n+1)!$ y en $\overset{2}{\leq}$ que $2 \leq n+2$. Una vez mostrado que se verifican **H1** y **H2**, se sabe que $2^n \leq (n+1)!$ vale para todo $n \in \mathbb{N}$. El lector habrá visto que la idea para demostrar que P_n implica P_{n+1} ha de buscarse por cuenta propia sin que el principio de inducción diga cómo hacerlo.

Supongamos que estudiamos si $P_n : 2^n \leq n!$ es cierta para todo n. Enseguida se ve que es falsa para $n \leq 3$ pero sí es cierta para $n = 4$ y conjeturamos que será cierta para $n \geq 4$. No podemos aplicar el principio, tal como enunciamos, pero se tiene el **Principio de Inducción Generalizado.**

Principio de Inducción Generalizado. *Supongamos que para cada $n \in \mathbb{N}$ hay una proposición, propiedad, condición, fórmula, etc. P_n para cada $n \geq m$* siendo $m \in \mathbb{N}$ fijo de antemano. *Se cumplirá que P_n es cierta para todo $n \geq m$ si se verifican las siguientes condiciones*

(H1) *P_m es cierta (la "primera" de las proposiciones es cierta).*

(H2) Para todo $n \geq m$ se cumple que si *P_n es cierta, P_{n+1} es cierta también.*

Para estudiar $P_n : 2^n \leq n!$ se constata primero que P_4 es cierta y luego se prueba, suponiendo $2^n \leq n!$ que

$$2^{n+1} = 2 \cdot 2^n \overset{1}{\leq} 2 \cdot n! \overset{2}{\leq} (n+1) \cdot n! = (n+1)!$$

con comentarios análogos al caso $2^n \leq (n+1)!$. Evidentemente hemos supuesto $m = 4$.

Es muy probable que una demostración donde se comprueban los casos de $n = 1, 2, 3$ y se dice "y así sucesivamente" es una demostración que, rigurosamente expuesta, requiere el principio de inducción.

Vamos a "demostrar" por inducción que todos los alumnos de una clase tienen la misma estatura. Se hará por inducción sobre n, el número de alumnos de la clase. El teorema es cierto si $n = 1$. Supongámoslo cierto para n, elijamos un alumno concreto a de estatura α y dividamos el conjunto A de alumnos en dos subconjuntos B y C con $n-1$ alumnos de modo que $a \in B \cap C$ y $A = B \cup C$. Por la hipótesis de inducción todos los alumnos de B tienen la misma estatura $\beta = \alpha$ y los de C la misma estatura $\gamma = \alpha$ por estar a en B y C, luego todos miden α.

Problema 49 *¿De verdad crees que todos los alumnos tienen la misma estatura?*

Problema 50 *Sea $a \in \mathbb{k}^{m \times m}$ con todo $a^i_j = 1$. ¿Cuál es la expresión de $a^n = a \cdot \ldots \cdot a$, la potencia $n-$ésima de a? (no es la fila n de a).*

Ciertos puntos de este capítulo quedan mejor si se hacen por inducción.[16] El teorema 15 prueba que si $a \in \mathbb{k}^{m \times n}$ y $b \in \mathbb{k}^{n \times p}$ y E una composición de operaciones elementales sobre matrices con m filas, se tiene que $E(ab) = E(a)\, b$. El lector puede revisar la demostración pero le vamos a decir cómo se redactaría con el formalismo de inducción. Se supone que $E = F_h \circ \cdots \circ F_1$ y se va a probar que $E(ab) = E(a)\, b$ por inducción sobre h. El caso $h = 1$, que corresponde a $E = F$, una única operación

[16] El lector debe tener en cuenta que no es una simple cuestión de estética o de ahorrar espacio. El dar una demostración por inducción evitando el "y así sucesivamente" es lo que aparece en cualquier libro que cuente con una mínima sofisticación por parte del lector.

fila, se hace como en ese teorema se vio. Es el caso base, la demostración de que se tiene la hipótesis **H1**. Para probar **H2** se supone que $E(ab) = E(a)\,b$ es cierta para h y se ha de probar para $h+1$. Supongamos $E = F_{h+1} \circ F_h \circ \cdots \circ F_1$, Sea $E' = F_h \circ \cdots \circ F_1$, luego $E = F_{h+1} \circ E'$. Entonces

$$E(ab) = F_{h+1} \circ (E'(ab)) \overset{1}{=} F_{h+1}(E'(a) \cdot b) \overset{2}{=} [F_{h+1}(E'(a))] \cdot b \overset{3}{=} [(F_{h+1} \circ F_h \circ \ldots \circ F_1)(a)] \cdot b = E(a)\,b.$$

En $\overset{1}{=}$ se ha usado la hipótesis de inducción, pues E' es producto de h operaciones fila; en $\overset{2}{=}$ el caso $h = 1$ aplicando $F = F_{h+1}$ al producto $E'(a) \cdot b$ (se sustituye a por $E'(a)$); y en $\overset{3}{=}$ la definición de composición de funciones.

Problema 51 *Probar por inducción que si E es composición de operaciones elementales, $E = F_h \circ \cdots \circ F_1$, se tiene $E(I) = \phi_h \cdot \cdots \cdot \phi_1$, siendo $\phi_j = F_j(I)$.*

Queremos cerrar la sección con un importante teorema que dice que si $a \in \Bbbk^{m\times n}$ y con operaciones fila la transformamos en b escalonada reducida, entonces b es única; o sea, puede haber muchas formas de elegir las operaciones, pero siempre se llegará a la misma matriz. Advertimos que si decimos solo "escalonada" en vez de "escalonada reducida", la unicidad es falsa. La consecuencia práctica más importante de este teorema es que al reducirla a forma tan solo escalonada c, el número de filas no nulas en c es el mismo que el que tiene b y por tanto el número de filas no nulas que aparecen al transformar a en alguna de las posibles formas escalonadas c está bien definido. Ya comentamos que este número es el **rango de la matriz** y la importancia del concepto. Advertimos al lector que hay demostraciones mucho más sencillas que esta de la unicidad de la forma escalonada reducida o forma de Hermite con material básico de los capítulos posteriores, y que la mayor virtud de la que aquí damos es que solo usa lo más básico. *Puede considerarse materia optativa,* aunque siempre es una oportunidad de ponerse a prueba. Además es un buen ejemplo de demostración por inducción en donde el "y así sucesivamente...", "lo vemos para n pequeño y lo general es análogo..." no sirve.

Teorema 21 *Supongamos que $a \in \Bbbk^{m\times n}$ es transformable en $b, c \in \Bbbk^{m\times n}$ escalonadas reducidas, habiendo por tanto $p, q \in \Bbbk^{m\times m}$ invertibles tales que $pa = b$ y $qa = c$. Se tiene entonces que $b = c$.*

Demostración. La demostración será por inducción sobre n, el número de *columnas* de a, b y c. Si $n = 1$, b y c son matrices columna y solo hay dos matrices columna en forma escalonada reducida, que son $(0,\ldots,0)^\top$ y $(1,0,0,\ldots,0)^\top$. Si $a = 0$ tienen que ser $b = c = (0,\ldots,0)^\top$ y si es $a \neq 0$ tiene que ser $b = c = (1,0,\ldots,0)^\top$. El teorema es cierto para $n = 1$.

Supongamos el teorema cierto para $n-1$ y probémoslo para n. Sean $\bar a, \bar b, \bar c$ las matrices a, b, c quitándoles la última columna n. Sigue siendo cierto que $p\bar a = \bar b$ y $q\bar a = \bar c$, luego, por la hipótesis inductiva, $\bar b = \bar c$, y solo debemos probar que b y c deben tener igual la última columna porque las $n-1$ primeras son iguales. Lo haremos por reducción al absurdo, suponiendo $b \neq c$ y mostrando al final que $b = c$.

La afirmación clave a probar es esta: ($\bigstar$) *si $b \neq c$, cualquiera de las condiciones $bx = 0$ y $cx = 0$ para un cierto $x \in \Bbbk^n$ implica la otra y la última componente x^n de x es cero.* Probémoslo. Supongamos por ejemplo que $bx = 0$. Aplicando p^{-1} a $0 = bx = pax$ se tiene $0 = ax$ y luego $cx = qax = q0 = 0$. Análogamente, $cx = 0$ implica $bx = 0$. Sea $x \in \Bbbk^n$ tal que $bx = cx = 0$. Al suponer $b \neq c$, por la hipótesis inductiva, la diferencia entre b y c está en la columna n y hay un índice k tal que $b^k_n \neq c^k_n$. Entonces, $0 = (b-c)\,x \in \Bbbk^m$ y el coeficiente k de $(b-c)\,x$ es $\left(b^k_n - c^k_n\right) x^n$ ya que todos los $b^k_j - c^k_j$ para $j < n$ son nulos por la hipótesis inductiva. De $\left(b^k_n - c^k_n\right) x^n = 0$ sale $x^n = 0$.

Supondremos que $b \neq c$ en lo que sigue y llegaremos a contradicción. Sean $1 \le p_1 < \ldots < p_r \le n$ las columnas de los pivotes de b, luego $b^r_{p_r} = 1$ es el último pivote. Veamos que no puede ser $p_r < n$; o sea, lo que pasa es que *el último pivote está en la última columna.* Si fuera falso, b tendría la forma

$$b = \begin{pmatrix}
0 & \cdots & 0 & b^1_{p_1} & * & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & * & b^1_n \\
0 & \cdots & \cdots & \cdots & 0 & b^2_{p_2} & * & \cdots & \cdots & \cdots & \cdots & \cdots & * & b^2_n \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & b^3_{p_3} & * & \cdots & \cdots & \cdots & * & b^3_n \\
\vdots & & & & & & & \vdots & \vdots & \vdots & & & & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & b^r_{p_r} & * & \cdots & * & b^r_n \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & & & & & & & & \vdots & \vdots & & & & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0
\end{pmatrix}$$

con los $b^j_{p_j} = 1$. Vamos a construir un vector x tal que $bx = 0$ pero $x^n = 1$ y esta contradicción con ($\bigstar$) mostrará que el último pivote de b está en la columna n. Construimos x paso a paso, diciendo de momento que $x^n = 1$ y $x^k = 0$ cuando $k \neq p_1, \ldots, p_r$. Entonces, fijado $i = 1, \ldots, m$,

$$(bx)^i = \sum_{k=1}^{n} b^i_k x^k \overset{1}{=} \sum_{j=1}^{r} b^i_{p_j} x^{p_j} + b^i_n \overset{2}{=} b^i_{p_i} x^{p_i} + b^i_n = x^{p_i} + b^i_n.$$

Se usan en $\overset{1}{=}$ las condiciones impuestas a x y en $\overset{2}{=}$ que en la columna pivotal p_j solo es no nulo (de hecho vale 1) el término en la fila j. Como queremos que sea $bx = 0$, es obvio que debemos completar la definición de x con $x^{p_j} = -b^j_n$.

Los índices de las columnas pivotales de b y c son los mismos, con n como índice final. Podría suceder que el pivote de la columna n estuviese a distinta altura, digamos que sea r para b y $s > r$ para c. La fila c^{s-1} debe tener un término no nulo a la izquierda de la columna n y por tanto $\bar{c}$ no tiene todo ceros en $\bar{c}^{r+1}, \ldots, \bar{c}^m$ (observar que $r \leq s-1$). todas las filas nulas a partir de la fila r. Sin embargo, $\bar{b}$ sí tiene todas las filas a partir de la fila r. Esto es imposible porque $\bar{b} = \bar{c}$. Al estar los últimos pivotes de b y c en la misma posición y ser $\bar{b} = \bar{c}$, hemos llegado a $b = c$ aunque hemos partido de $b \neq c$. ♣

1.8. Más demostraciones por inducción

Damos, *con las demostraciones como material opcional,* una serie de resultados que utilizan en su demostración el principio de inducción. Nos permiten saber cuántas funciones o subconjuntos hay de un cierto tipo. Otro de los resultados es el famoso **binomio de Newton**. Tratamos también el concepto de **definición recursiva.**

El lector tiene una buena intuición de lo que es un conjunto finito o infinito. Diremos que X es **finito** si hay una *biyección* de $\{1, 2, \ldots .n\}$ en X y este $n \in \mathbb{N}$ es el número de elementos[17] que se designa por $\#X$. Si $X = \emptyset$ es $\#X = 0$. Un conjunto es **infinito** si no es finito (*por definición*).

Problema 52 *Sea X un conjunto con n elementos y $1 \leq m \leq n$. El número de funciones inyectivas de $\{1, 2, \ldots .m\}$ en X es $n!/(n-m)!$. En particular, el número de funciones biyectivas de $\{1, 2, \ldots .n\}$ en X o de X en X es $n!$.*♦

Solución. Se hace por inducción sobre n pero hay que dejar muy claro cual es la propiedad $P(n)$, que es "para todo $m \leq n$ el número de funciones inyectivas de $\{1, 2, \ldots .m\}$, $m \geq 1$, en *cualquier conjunto* con n elementos es $n!/(n-m)!$ ". Para $n = 1$ la fórmula es cierta pues $m = n = 1$ y no hay más que una biyección de $\{1\}$ en X. Supongamos cierta la fórmula para n. Supongamos cierta la fórmula si $\#X = n$ y tomemos X con $\#X = n+1$. Primeramente, para tener f de $\{1, \ldots, m\}$ en X inyectiva, elegimos $f(m)$ en X de $n+1$ formas posibles. La restricción de f a $\{1, \ldots, m-1\}$ debe tomar valores en $X - \{f(m)\}$, porque no puede ser $f(x) = f(m)$ si $x < m$, *ya que f es inyectiva.* La restricción por tanto tiene $n!/(n-(m-1))!$ formas posible y el total de funciones f es

$$\frac{n!}{(n-(m-1))!}(n+1) = \frac{(n+1)!}{((n+1)-m)!} \text{ funciones inyectivas de } \{1, \ldots, m\} \text{ en } X$$

y el paso inductivo está completo. ♦

El **número combinatorio** $\binom{m}{n}$ con $m, n \in \mathbb{N}$ y $m \geq n$ se define como

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

extendiendo también la definición a $n = 0$, que da $\binom{m}{0} = 1$. Este número tiene una interpretación muy importante en términos de conjuntos y que se usa mucho. Planteamos el problema con dos soluciones. Una usa inducción y la otra no.

[17] Se prueba que este n es único pues no puede haber biyecciones de $\{1, 2, \ldots .m\}$ en $\{1, 2, \ldots .n\}$ si $m \neq n$, algo que, digámoslo otra vez, es totalmente intuitivo pero que es necesario probar para fundamentar sólidamente las Matemáticas. No hay inconveniente en que lo veamos tan obvio como que $2 + 2 = 4$, pero conviene toparse de vez en cuando con este tipo de comentario "cultural".

Problema 53 *Sea X un conjunto con n elementos y $0 \leq m \leq n$. El número de subconjuntos de X con m elementos es $\binom{n}{m}$. (Se cuentan $\emptyset$ y el propio X.)*

Primera solución parcial. Si un huerto tiene p parcelas y cada parcela tiene q árboles, hay $t = pq$ árboles en el huerto. Si conocemos el total de árboles t y los que hay en cada parcela, podremos saber el número de parcelas. Conocemos el total de funciones inyectivas f de $\{1, \ldots, m\}$ en X y podemos dividirlas en "parcelas", habiendo para cada $Z \subset X$ con m elementos una "parcela" constituida por todas las f cuya imagen es Z. Si hay que contar los $Z \ldots$. ♦

Segunda solución parcial. Se hará por inducción sobre n, pero hay que dejar muy claro cual es la propiedad $P(n)$. Dice que "*cualquiera* que sea X con n elementos y *cualquiera* que sea $m \leq n$, el número de subconjuntos con m elementos de X es $\binom{n}{m}$". Hemos debido de precisar esto pues podíamos pensar que se hacía por inducción sobre m. Si $n = 0$ debe ser $X = \emptyset$ y X solo tiene un subconjunto que es X. Todo funciona porque

$$\binom{0}{0} = \frac{0!}{0!0!} = \frac{1}{1 \cdot 1} = 1.$$

Si X tiene $n = 1$ elementos solo hay dos subconjuntos, que son $\emptyset$ y X. Todo va bien porque

$$\binom{1}{0} = \frac{1!}{1!0!} = \frac{1}{1 \cdot 1} = 1, \qquad \binom{1}{1} = \frac{1!}{1!1!} = \frac{1}{1 \cdot 1} = 1.$$

Sea $n \geq 1$. Fijamos un punto $p \in X$ y dividimos el conjunto $\mathcal{S}$ de todos los subconjuntos de X con $n+1$ elementos en dos partes disjuntas $\mathcal{A}$ y $\mathcal{B}$, definiendo que Y está en $\mathcal{A}$ si $p \notin Y$ e Y está en $\mathcal{B}$ en caso contrario. Contar los elementos de $\mathcal{A}$ y $\mathcal{B}$ y la suma será los elementos que tiene $\mathcal{S}$.♦

Una aplicación muy importante de esto y donde ayuda ver $\binom{n}{m}$ como número de subconjuntos es el **binomio de Newton.** Nos dan números o matrices cuadradas que conmutan a y b y $n \in \mathbb{N}$. Entonces[18]

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

Problema 54 *Probar la fórmula del binomio de Newton.*

Solución parcial. Veamos primero una fórmula más general: la de $(a_1 + b_1)(a_2 + b_2) \cdots (a_n + b_n)$ suponiendo que todos los elementos conmutan entre sí. Para cada subconjunto S, incluso vacío, de $\{1, 2, \ldots .n\}$, representamos por a_S el producto de los a_i con $i \in S$. Si $S = \emptyset$ definimos $a_S = 1$. La definición de b_S es análoga. Pues bien, si S^* es el complementario de S; es decir, $S^* = \{1, 2, \ldots .n\} - S$, tenemos que

$$(a_1 + b_1)(a_2 + b_2) \cdots (a_n + b_n) = \sum_S a_S b_{S^*},$$

extendido el sumatorio a *todos* los subconjuntos de $\{1, 2, \ldots ., n\}$. Por tanto $b_1 \cdots b_n$ aparecerá como sumando correspondiente a $S = \emptyset$ y $a_1 \cdots a_n$ como correspondiente a $S = \{1, 2, \ldots .n\}$. Si $n = 5$ tenemos por ejemplo $a_{\{1,3\}} = a_1 a_3$ y $b_{\{1,3\}^*} = b_2 b_4 b_5$. Si tomamos como caso particular que todos los a_i sean a y todos los b_i sean b, los sumandos $a_S b_{S*}$ toman la forma $a^m b^{n-m}$, siendo m el número de elementos de S. La cuestión es que muchos sumandos están repetidos. ¿Cuántas veces aparece repetido $a^m b^{n-m}$? ♦

Hay una construcción muy relacionada con la inducción que es la llamada **definición recursiva** o **por inducción**. La definición intuitiva es fácil de comprender. Tenemos un conjunto X y vamos a definir para cada $n \in \mathbb{N}$ un elemento $x_n \in X$. Tendremos así lo que se llama una **sucesión** en X que, formalmente, es una función[19] de $\mathbb{N}$ en X. La definición recursiva es un procedimiento para construir sucesiones en los más variopintos X, aunque el lector va a encontrar muchos ejemplos con $X = \mathbb{R}$ en Análisis. La herramienta para las definiciones más sencillas es una función $\phi : X \to X$. Se elige $x_1 \in X$ y se define $x_2 = \phi(x_1)$. A continuación se repite el proceso, $x_3 = \phi(x_2)$, $x_4 = \phi(x_3)$ y, en general, $x_n = \phi(x_{n-1})$. Dicho de otra manera,

$$x_1 \text{ de libre elección}, \; x_2 = \phi(x_1), \; x_3 = \phi^2(x_1), \; x_4 = \phi^3(x_1), \ldots, x_n = \phi^{n-1}(x_1),$$

[18] La fórmula vale si los "objetos" se pueden multiplicar de modo conmutativo; o sea, $ab = ba$. Así por ejemplo, la fórmula vale si a y b son matrices cuadradas que conmutan o, en general, elementos de un anillo que conmutan.

[19] Se suele denotar una función f de $\mathbb{N}$ en X acordando que $f(n)$ represente el valor de f aplicado a n. Tiene sus ventajas y una larga tradición de uso la notación con índices con la que f_n es el valor de f aplicado a n. Es importante en todo caso no olvidar que una sucesión es una *función*. y que si nos hablan de la sucesión (x_n) nos hablan de una función x de $\mathbb{N}$ en X, por mucho que elegir x para denotar una función suene extraño.

siendo $\phi^k = \phi \circ \phi \circ \cdots \circ \phi$ la composición de ϕ consigo misma k veces. Si tenemos por ejemplo $X = \mathbb{R}$ y $\phi(x) = 2x + 3$, tras elegir $x_1 = 0$ tenemos

$$x_1 = 0,\ x_2 = 2 \cdot 0 + 3 = 3,\ x_3 = 2 \cdot 3 + 3 = 9,\ x_4 = 2 \cdot 9 + 3 = 21, \ldots$$

y la sucesión es $(0, 3, 9, 21, \ldots)$. Frecuentemente se piden probar propiedades de la sucesión definida recursivamente. Por ejemplo, si $x_1 > 2$ probamos que $x_n > 2^n$. Se hace por inducción, siendo obvia la certeza si $n = 1$. Suponiendo que $x_n > 2^n$ se tiene que $x_{n+1} = 2x_n + 3 > 2x_n > 2 \cdot 2^n = 2^{n+1}$.

Teorema 22 *Si tenemos una función $\phi : X \to X$ y elegimos x_1 en X, hay una sucesión (x_n) en X correctamente definida por $x_1 = x_1$ y $x_{n+1} = \phi(x_n)$ para cada $n \in \mathbb{N}$.*

Nosotros hemos dicho sin mayor solemnidad que si a es un número, matriz o polinomio y si $n \in \mathbb{N}$ se define a^n, la potencia n de a, como $a \cdot a \cdot \ldots \cdot a$, que es a por sí mismo n veces y para algo tan simple, el teorema anterior es exagerado. Hay cierta razón cuando la sucesión a construir es "sencilla" pero puede haber casos más complicados y es bueno saber que esto funciona siempre. Sin embargo es frecuente que el problema no sea dar rigor a la definición (podemos tener cierta manga ancha) sino conjeturar y probar algo que deben cumplir todas las x_n. Hay muchísimos problemas de ese tipo donde P_1 cierta significa que x_1 cumple la propiedad o tiene la expresión que queramos, y el paso inductivo es probar que si x_n propiedad o tiene la expresión (esto es la veracidad de P_n) entonces $x_{n+1} = \phi(x_n)$ también tiene esa propiedad o expresión, y al hacer esto se habrá probado que P_n implica P_{n+1}.

Hemos dado como algo sin vuelta de hoja que $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$ está perfectamente definido y el lector puede darse por conforme. Sin embargo, si queremos ver cómo se define el factorial con el teorema 22 le pedimos al lector que complete lo que vamos a bosquejar. Tomamos $X = \mathbb{N} \times \mathbb{N}$, definimos $\phi(a, b) = (a(b + 1), b + 1)$ y elegimos $x_1 = (1, 1)$. Construimos con el teorema 22 la sucesión (x_n), que está en $\mathbb{N} \times \mathbb{N}$ (no en $\mathbb{N}$). Cada x_n tiene dos componentes y la primera componente de x_n es $n!$

Capítulo 2

Espacios vectoriales

2.1. El espacio vectorial estándar

Nos interesan particularmente los espacios de las matrices fila $\mathbb{k}^{1\times n}$ y las matrices columna $\mathbb{k}^{m\times 1}$. Por razones tipográficas parece mejor concentrar la atención en $\mathbb{k}^{1\times n}$ cuyos elementos se escriben como $(x_1,,\ldots,x_n)$ y no como columnas que requieren mucho más espacio. Además, para $\mathbb{k}=\mathbb{R}$ estamos acostumbrados a representar los puntos del plano y el espacio $\mathbb{R}^2$ y $\mathbb{R}^3$ como pares o ternas (x,y) o (x,y,z). *No seguiremos esa costumbre* y los puntos del espacio de cualquier dimensión, se representarán por vectores *columna*; es decir, hablaremos del punto $(2,-3)^\top$ pues hay que trasponer una matriz fila para llegar a una matriz columna.[1] Diremos que $\mathbb{k}^{m\times 1}$, escrito casi siempre como $\mathbb{k}^m$, es el **espacio vectorial estándar** y sus elementos se llamarán **vectores**. Es razonable suprimir el subíndice, que siempre sería 1, luego un vector $x\in\mathbb{k}^m$ se escribe como

$$x=\begin{pmatrix} x^1\\ x^2\\ \vdots\\ x^m\end{pmatrix} \text{ o bien } x=\left(x^1,x^2,\ldots,x^n\right)^\top$$

y los superíndices representan el número de la coordenada *y no un exponente*. Los vectores se pueden sumar y multiplicar por escalares $\lambda\in\mathbb{k}$ con las operaciones generales que ya conocemos como matrices que son. Las propiedades que recoge el teorema 8 son básicas para el manejo de vectores. *La definición de espacio vectorial, esencial pero que aparece en la sección siguiente, es una abstracción de* $\mathbb{k}^m$.

Si tenemos una sucesión de vectores $(a_1,\ldots,a_n)$ en $\mathbb{k}^m$ (y puede ser $m\neq n$) definiremos una **combinación lineal** de $(a_1,\ldots,a_n)$ como un vector $v\in\mathbb{k}^m$ de la forma $v=\lambda^1a_1+\ldots+\lambda^na_n$, siendo $\lambda^1,\ldots,\lambda^n$ elementos del cuerpo $\mathbb{k}$. Evidentemente, al variar la sucesión $\left(\lambda^1,\ldots,\lambda^n\right)$, varía x y hay muchas combinaciones de los elementos $(a_1,\ldots,a_n)$. La condición expresada en coordenadas es

$$\begin{pmatrix} v^1\\ \vdots\\ v^i\\ \vdots\\ v^m\end{pmatrix}=\lambda^1\begin{pmatrix} a_1^1\\ \vdots\\ a_1^i\\ \vdots\\ a_1^m\end{pmatrix}+\ldots+\lambda^j\begin{pmatrix} a_j^1\\ \vdots\\ a_j^i\\ \vdots\\ a_j^m\end{pmatrix}+\ldots+\lambda^n\begin{pmatrix} a_n^1\\ \vdots\\ a_n^i\\ \vdots\\ a_n^m\end{pmatrix} \tag{2.1}$$

o con notación matricial

$$\begin{pmatrix} a_1^1 & \cdots & a_j^1 & \cdots & a_n^1\\ \vdots & \ddots & \vdots & & \vdots\\ a_1^i & \cdots & a_j^i & \cdots & a_n^i\\ \vdots & & \vdots & \ddots & \vdots\\ a_1^m & \cdots & a_j^m & \cdots & a_n^m\end{pmatrix}\begin{pmatrix} \lambda^1\\ \vdots\\ \lambda^i\\ \vdots\\ \lambda^n\end{pmatrix}=\begin{pmatrix} v^1\\ \vdots\\ v^i\\ \vdots\\ v^m\end{pmatrix}. \tag{2.2}$$

[1] La razón es que al intervenir funciones lineales, la función L evaluada en x vendrá dada por el producto ax siendo x una matriz y para poder multiplicar a y x se necesita que x sea un vector columna.

Es evidente que para saber si v es combinación lineal de $a_1, \ldots, a_n$, lo que hay que hacer es formar la matriz $a \in \Bbbk^{m\times n}$ cuyas columnas son $a_1, \ldots, a_n$ y plantear el sistema $ax = v$ (con v como dato en lugar de la tradicional y) y mirar si tiene solución. Si no la tiene v no es combinación lineal de $(a_1, \ldots, a_n)$ y si la tiene, *cualquier* solución $\left(\lambda^1, \ldots, \lambda^n\right)$ sirve para expresar esa combinación. Conviene que el lector repase el capítulo anterior y vea que hay muchos problemas de sistemas lineales que se traducen en preguntas sobre si ciertos vectores son combinación lineal de otros y con qué coeficientes.

Otro concepto esencial es el de **vectores linealmente independientes**. Diremos que la *sucesión* $(a_1, \ldots, a_n)$ de vectores de $\Bbbk^m$ es **linealmente independiente** si la combinación lineal $0 = \lambda^1 a_1 + \ldots + \lambda^n a_n$ solo es posible cuando $\lambda^1 = \ldots = \lambda^n = 0$. Es inmediato con el párrafo anterior que esto equivale a decir que el sistema *homogéneo* $ax = 0$ solo tiene la solución $x = 0$. Cuando $(a_1, \ldots, a_n)$ *no es* linealmente independiente, se dice que es **linealmente dependiente** (la doble negación equivale a una afirmación). El lector debe comprobar algo muy fácil pero de uso continuo sin comentarios y es que si en $(a_1, \ldots, a_n)$ hay algún $a_j = 0$ o para $j \neq k$ se tiene $a_j = a_k$, entonces $(a_1, \ldots, a_n)$ es dependiente. Hay una cierta confusión sobre si al hablar de independencia lineal se habla de *sucesiones* de vectores o de *conjuntos* de vectores. Por lo que se refiere a independencia (no así con la dependencia) el riesgo es mínimo. Acabamos de decir que si $(a_1, \ldots, a_n)$ es independiente, no pueden repetirse los a, luego $\{a_1, \ldots, a_n\}$ tiene n elementos. Ya dijimos en la página 11 que $\{a_1, \ldots, a_n\}$ no es un conjunto a secas sino un conjunto *ordenado* o *numerado*. Se puede hablar del *conjunto independiente* $\{a_1, \ldots, a_n\}$ y pasar a la sucesión $(a_1, \ldots, a_n)$ sin ambigüedad. Si $\{a_1, \ldots, a_n\}$ es un conjunto con los a distintos dos a dos, también se pasa a la sucesión $(a_1, \ldots, a_n)$ sin ambigüedad. Mientras se suponga al hablar de dependencia o independencia lineal que no hay repeticiones y que el concepto no depende de cómo se numeren los elementos, se puede tratar con *conjuntos* dependientes e independientes que es lo que hacen la mayoría de los autores.

Los subconjuntos de $\Bbbk^m$ que más nos interesan son los **subespacios vectoriales**. Si nos restringimos a $\mathbb{R}^2$ y $\mathbb{R}^3$, un subespacio vectorial es una recta o un plano que pasa por el origen, con el origen y todo $\mathbb{R}^2$ o $\mathbb{R}^3$ como casos extremos. El lector sabe que estos subespacios vienen determinados por cierto tipo de ecuaciones como $(x, y) = t\,(2, 3) + (1, 0)$ o $2x + 2y - 5z = 0$ y puede pensar que la generalización a $\mathbb{R}^m$ o $\Bbbk^m$ va a ser definir subespacio vectorial como subconjunto expresable por cierto tipo de ecuaciones. No es la vía correcta. Lo que se hace es definir un subespacio vectorial $\mathbb{F}$ como un cierto tipo de subconjunto (de momento, subconjunto de $\Bbbk^m$; luego se generalizará).[2] Un **subespacio vectorial** $\mathbb{F}$ de $\Bbbk^m$ es un subconjunto *no vacío* de $\Bbbk^m$ que cumple que si $x, y \in \mathbb{F}$ y $\lambda \in \Bbbk$ entonces $x + y$ y λx también están en $\mathbb{F}$. Digamos informalmente que un subespacio vectorial es un subconjunto de $\Bbbk^m$ "suficientemente regular" para que al sumar sus elementos o multiplicarlos por escalares $\lambda \in \Bbbk$ no nos salgamos de él. Es fácil visualizar que los subespacios vectoriales de $\mathbb{R}^2$ y $\mathbb{R}^3$ son las rectas y planos *a través de* 0 y, como casos extremos pero admisibles, el subconjunto formado tan solo por el origen 0 (el **subespacio cero**, denotado por 0) y el subconjunto de todos los puntos de $\mathbb{R}^2$ o $\mathbb{R}^3$, el **subespacio total**. Si trabajamos con $\mathbb{R}^m$ y $m > 4$ se pierde la intuición visual y si se cambia de $\mathbb{R}$ a $\Bbbk$, se pierde también aunque sea $m \leq 3$. *Aun así, guiarse de lo que se intuye cierto para* $\mathbb{R}^2$ *y* $\mathbb{R}^3$ *es a veces conveniente.* El origen $0 = (0, \ldots, 0)^\top$ de $\Bbbk^m$ está siempre en todos los subespacios vectoriales, luego el vacío $\emptyset$ *nunca puede ser* un subespacio vectorial (lo dice la definición, pero es que ahora vamos a ser más concretos diciendo que $0 \in \mathbb{F}$). En efecto, como $\mathbb{F} \neq \emptyset$ existe un $x \in \mathbb{F}$ y tomando $\lambda = 0$, se tiene que $\lambda \cdot x = 0 \cdot x = 0$ debe estar también en $\mathbb{F}$.

Si buscamos ejemplos concretos de subespacios en $\Bbbk^m$ vemos dos grandes grupos.

Como queremos que las matrices sean $m \times n$, vamos a dar ejemplos de subespacios de $\Bbbk^n$ (se cambia m en las definiciones generales por n). Se toma $a \in \Bbbk^{m\times n}$ y se define $\mathbb{F} = \{x \in \Bbbk^n \mid ax = 0\}$; o sea, $\mathbb{F}$ es el espacio de soluciones del sistema *homogéneo* $ax = 0$. Es inmediato que $\mathbb{F} \neq \emptyset$ porque $0 \in \mathbb{F}$ y si $x, y \in \mathbb{F}$ y $\lambda \in \Bbbk$ entonces $a\,(x + y) = ax + ay = 0 + 0 = 0$ y $a\,(\lambda x) = \lambda\,(ax) = \lambda \cdot 0 = 0$, luego también $x + y$ y λx están en $\mathbb{F}$. Podemos resumir diciendo que *los espacios de soluciones de un sistema homogéneo*

[2] Letras con tipo *negrita de pizarra* (*blackboard bold* en inglés) como $\mathbb{E}, \mathbb{F}, \ldots, \mathbb{S}$ se reservarán para espacios y subespacios vectoriales.

$ax = 0$ *son un ejemplo de espacio vectorial.* Los $x \in \mathbb{F}$ son las soluciones del sistema

$$\left\{\begin{array}{l} a_1^1 x^1 + \cdots + a_j^1 x^j + \cdots + a_n^1 x^n = \sum_{j=1}^n a_j^1 x^j = 0 \\ \qquad\qquad\qquad\qquad \vdots \\ a_1^i x^1 + \cdots + a_j^i x^j + \cdots + a_n^i x^n = \sum_{j=1}^n a_j^i x^j = 0 \\ \qquad\qquad\qquad\qquad \vdots \\ a_1^m x^1 + \cdots + a_j^m x^j + \cdots + a_n^m x^n = \sum_{j=1}^n a_j^m x^j = 0 \end{array}\right. \quad \text{que se condensa en } ax = 0.$$

y se dice que esta matriz o sistema de ecuaciones dan $\mathbb{F}$ **en forma implícita.**

Otro tipo de ejemplos, ahora de subespacios de $\Bbbk^m$, es elegir una sucesión de vectores $(a_1, \ldots, a_n)$ en $\Bbbk^m$ y definir $\mathbb{F}$ como el conjunto de todas las combinaciones lineales de $(a_1, \ldots, a_n)$. Cada sucesión $\left(\lambda^1, \ldots, \lambda^m\right)$ en $\Bbbk$ da lugar a una combinación lineal y por tanto a un elemento de $\mathbb{F}$. Es fácil ver que $\mathbb{F}$ es un subespacio. Si todas las λ son 0 obtenemos que $0 \in \mathbb{F}$. Suponiendo $v, w \in \mathbb{F}$ y $\lambda \in \Bbbk$ escribimos

$$\left\{\begin{array}{l} v = \lambda^1 a_1 + \ldots + \lambda^n a_n \\ w = \mu^1 a_1 + \ldots + \mu^n a_n \end{array}\right. \quad \text{y entonces } v + w = \left(\lambda^1 + \mu^1\right) a_1 + \ldots + \left(\lambda^n + \mu^n\right) a_n$$

y la sucesión $\left(\left(\lambda^1 + \mu^1\right), \ldots, \left(\lambda^n + \mu^n\right)\right)$ nos prueba que $v + w \in \mathbb{F}$. De modo análogo, $\left(\lambda\lambda^1, \ldots, \lambda\lambda^n\right)$ demuestra que $\lambda v \in \mathbb{F}$. Detallando con coordenadas obtenemos para $\mathbb{F}$ las ecuaciones (2.1) y (2.2). Al mover los n parámetros λ en $\Bbbk$ se obtienen los elementos de $\mathbb{F}$ y por esto estas ecuaciones se llaman **ecuaciones paramétricas**. Introducimos la notación $\lg(a_1, \ldots, a_n)$ para denotar el subespacio de todas las combinaciones lineales de $(a_1, \ldots, a_n)$, que llamamos también el **subespacio generado por** $(a_1, \ldots, a_n)$ y decimos que $(a_1, \ldots, a_n)$ es **sucesión generadora** o que **genera** $\lg(a_1, \ldots, a_n)$.[3] Cada sucesión $\left(\lambda^1, \ldots, \lambda^n\right)$ da un $v \in \mathbb{F}$ pero puede suceder que $\left(\mu^1, \ldots, \mu^n\right) \neq \left(\lambda^1, \ldots, \lambda^n\right)$ dé el mismo v. Si todos los λ_j son cero excepto $\lambda_k = 1$ se obtiene a_k, luego $\{a_1, \ldots, a_n\} \subset \lg(a_1, \ldots, a_n)$. Nos gustaría que hubiese correspondencia biunívoca entre sucesiones $\left(\lambda^1, \ldots, \lambda^n\right)$ de parámetros y vectores v de $\mathbb{F}$. Puede conseguirse.

Teorema 23 *Para que la correspondencia entre sucesiones de parámetros* $\left(\lambda^1, \ldots, \lambda^n\right)$ *y elementos de* $\mathbb{F} = \lg(a_1, \ldots, a_n)$ *sea biunívoca (o biyectiva) es necesario y suficiente que la sucesión* $(a_1, \ldots, a_n)$ *de vectores sea independiente.*

Demostración. Si $(a_1, \ldots, a_n)$ es independiente y $v = \lambda^1 a_1 + \ldots + \lambda^n a_n = \mu^1 a_1 + \ldots + \mu^n a_n$, restamos y

$$0 = v - v = v = \left(\lambda^1 - \mu^1\right) a_1 + \ldots + \left(\lambda^n - \mu^n\right) a_n.$$

La independencia nos da $\lambda^1 - \mu^1 = \ldots = \lambda^n - \mu^n = 0$ y $\left(\lambda^1, \ldots, \lambda^n\right) = \left(\mu^1, \ldots, \mu^n\right)$. Recíprocamente si hay correspondencia biunívoca y $0 = \lambda^1 a_1 + \ldots + \lambda^n a_n$, observamos que también $0 = 0 \cdot a_1 + \ldots + 0 \cdot a_n$ luego $\lambda^1 = 0, \ldots, \lambda^n = 0$, que es la condición de independencia lineal. ♣

Es fácil imaginar que al considerar un subespacio $\mathbb{F} = \lg(a_1, \ldots, a_n)$ nos interesa que $(a_1, \ldots, a_n)$ sea independiente. Esto nos lleva a una definición esencial, que en principio la hacemos para $\Bbbk^m$ y luego se extenderá. Diremos que $(a_1, \ldots, a_n)$ es **base** de $\Bbbk^m$ si $(a_1, \ldots, a_n)$ es independiente y genera $\Bbbk^m$. Al generar, cada $v \in \mathbb{F}$ se escribe como $v = \lambda^1 a_1 + \ldots + \lambda^n a_n$ *como mínimo* de una manera, y al haber independencia, $v = \lambda^1 a_1 + \ldots + \lambda^n a_n$ se escribe *como máximo* de una manera. El resumen es que cada $v \in \Bbbk^m$ se escribe de modo único. Advertimos que en la definición se admite la posibilidad (que luego veremos que no se puede dar) de que sea $m \neq n$. La ventaja de tener una base $(a_1, \ldots, a_n)$ es poder asignar a cada $x \in \Bbbk^m$ unas **coordenadas** $\left(\xi^1, \ldots, \xi^n\right)$ en esa base cuando $x = \xi^1 a_1 + \ldots + \xi^n a_n = \sum_{i=1}^n \xi^i a_i$. Remitirse a las coordenadas a veces simplifica y a veces complica; ya veremos por qué.

Cuesta poco obtener bases de $\Bbbk^m$. Basta tomar una matriz *invertible* $a \in \Bbbk^{m \times m}$ (¡ojo!, $m = n$) y considerar las columnas de a como la sucesión $(a_1, \ldots, a_m)$. Para cualquier $v \in \Bbbk^m$, el sistema $ax = v$ tiene solución única $x = a^{-1}v$, luego con las columnas $(a_1, \ldots, a_m)$ de a podemos escribir de modo único (equivalente a la independencia lineal por el teorema 23) cada $v \in \Bbbk^m$. El caso más obvio, y sin embargo esencial, es aquel en que se toma como a la matriz unidad I. Se *reserva para siempre* e_j en vez de I_j

[3] La (rara) notación $\lg(\bullet)$ se elige por las iniciales de *linealmente generado.*

para indicar la columna j de I, y tenemos una base

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots \quad , e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

que se llama la **base estándar de** $\Bbbk^m$.[4] Obsérvese la ecuación simple pero muy importante

$$x = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ \vdots \\ x^n \end{pmatrix} = x^1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x^3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + x^n \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{m} x^i e_i.$$

Es muy fácil realizar cálculos numéricos si se saben manejar los sistemas lineales. Sea $\Bbbk = \mathbb{R}$. Definamos en $\mathbb{R}^3$

$$a_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}, \quad v = \begin{pmatrix} -3 \\ 0 \\ 3 \end{pmatrix}, \quad w = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

y sea a la matriz con columnas a_j. Si $x = (3, 2, -2)^\top$, vemos que $ax = v$, que es como decir que

$$3a_1 + 2a_2 - 2a_3 = 3 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 2 \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - 2 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 3 \end{pmatrix} = v$$

y v es combinación lineal de (a_1, a_2, a_3) con coeficientes $(3, 2, -2)$. De modo más general se puede plantear el sistema $ax = v$ cuya solución general es $(\lambda + 5, -2\lambda - 2, \lambda)^\top$ con $\lambda \in \mathbb{R}$. Para cualquier elección de λ tenemos una terna de coeficientes $\left(\lambda^1, \lambda^2, \lambda^3\right)$ de modo que $\lambda^1 a_1 + \lambda^2 a_2 + \lambda^3 a_3 = v$ mostrando que v se puede expresar de muchos modos como combinación lineal de (a_1, a_2, a_3). Por el teorema 23 ni (a_1, a_2, a_3) es independiente ni es base. El hecho de que el sistema $ax = w$ no sea compatible equivale a que w no sea combinación lineal de (a_1, a_2, a_3) y una nueva confirmación de que no es base. Sin embargo las columnas de una *nueva* matriz

$$a = (a_1, a_2, a_3) = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

forman base puesto, que a es triangular sin ceros en la diagonal (¡invertible!). Dado que

$$a^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}, \quad a^{-1}w = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$$

el vector w que con la *antigua* a no se podía expresar como combinación lineal de las columnas, sí se puede con la *nueva* a siendo $w = a_1 - 3a_2 + 2a_3$. Las coordenadas de $w = (1, 1, 2)^\top$ en la base (a_1, a_2, a_3) son $(1, -3, 2)$ aunque no hay duda que en la base estándar son $(1, 1, 2)^\top$. Como se ve, determinar si hay combinación lineal, independencia, base, etc. tiene *mientras estemos en* $\Bbbk^m$, una verificación sencilla con sistema lineales, aunque quizás laboriosa en los cálculos.

Problema 55 *Para* $\Bbbk = \mathbb{C}$ *y* $\mathbb{Z}_2$ *nos dan las matrices*

$$a = \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Estudiar si (a_1, a_2) *es base de* $\mathbb{C}^2$ *y* (b_1, b_2, b_3) *es base de* $(\mathbb{Z}_2)^3$. *Cuando lo sean, expresar* $(1, 1)$ *o* $(1, 1, 1)$, *según proceda o no, como combinación lineal de elementos de la base. Cuando no sean base, dar un ejemplo concreto de un* w *que no se pueda expresar como combinación lineal de* (a_1, a_2) *o* (b_1, b_2, b_3) *según el caso.*

[4] Otros autores hablan de *base canónica* por ser el canon, modelo o patrón de las demás bases. Se usa también *base natural* o *base usual*. La RAE dice que *estándar* significa *que sirve como tipo, modelo, norma, patrón o referencia.*

2.1.1. Los teoremas principales

Vamos a estudiar metódicamente el concepto de base y otros relacionados. Dada una sucesión $(z_1,\ldots,z_k)$ en cualquier conjunto Z, diremos que k es la **longitud de la sucesión**. Obsérvese que $(z_1,z_2,z_3)=(5,5,5)$ tiene longitud 3 pero $\{z_1,z_2,z_3\}=\{5,5,5\}=\{5\}$ tiene solo un elemento. Veremos que en las demostraciones teóricas que siguen las sucesiones de máxima o mínima longitud de cierto tipo juegan un papel esencial. Las demostraciones son ingeniosas aunque no las mejores para un cálculo meramente computacional.

Teorema 24 *Si un subespacio $\mathbb{F}$ está generado por n vectores, $\mathbb{F}=\lg(a_1,\ldots,a_n)$, y $(b_1,\ldots,b_k)$ es una sucesión independiente de vectores de $\mathbb{F}$, entonces $k\leq n$.*

Demostración. Es equivalente probar que si $k>n$, entonces $(b_1,\ldots,b_k)$ es dependiente. Como los b están en $\mathbb{F}$ son por ello combinación lineal de los a y hay coeficientes h tales que

$$\begin{cases} b_1=h_1^1a_1+\ldots+h_1^na_n \\ \vdots \\ b_k=h_k^1a_1+\ldots+h_k^na_n \end{cases}\text{, que con sumatorios es }\begin{cases} b_1=\sum_{i=1}^n h_1^ia_i \\ \vdots \\ b_k=\sum_{i=1}^n h_k^ia_i \end{cases}$$

Afirmamos que pueden elegirse $\lambda^1,\ldots,\lambda^k$ *no todos nulos* de modo que $0=\lambda^1b_1+\ldots+\lambda^kb_k$, probando así la dependencia. Sustituimos

$$\begin{aligned}\lambda^1b_1+\ldots+\lambda^kb_k&=\lambda^1\left(h_1^1a_1+\ldots+h_1^na_n\right)+\ldots+\lambda^k\left(h_k^1a_1+\ldots+h_k^na_n\right)\\&=\left(\lambda^1h_1^1+\ldots++\lambda^kh_k^1\right)a_1+\ldots+\left(\lambda^1h_1^n+\ldots+\lambda^kh_k^n\right)a_n.\end{aligned}$$

(Para entrenamiento del lector, el cálculo con sumatorios es

$$\sum_{j=1}^k\lambda^jb_j=\sum_{j=1}^k\lambda^j\left(\sum_{i=1}^n h_j^ia_i\right)=\sum_{i=1}^n\left(\sum_{j=1}^k h_j^i\lambda^j\right)a_i$$

y comprobará que se dice lo mismo con distinta notación.) Si los λ, aún no determinados, cumplen

$$\begin{cases} 0=\lambda^1h_1^1+\ldots+\lambda^kh_k^1 \\ \vdots \\ 0=\lambda^1h_1^n+\ldots+\lambda^kh_k^n \end{cases}\text{ y en forma matricial }\begin{pmatrix} h_1^1 & \cdots & h_k^1 \\ \vdots & \ddots & \vdots \\ h_1^n & \cdots & h_k^n \end{pmatrix}\begin{pmatrix}\lambda^1\\ \vdots\\ \lambda^k\end{pmatrix}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

se tiene $0=\sum_{j=1}^k\lambda^jb_j$ y si alguno no es nulo, la dependencia lineal de $(b_1,\ldots,b_k)$. Pero el sistema precedente es homogéneo con más incógnitas k que ecuaciones n, luego hay soluciones $\left(\lambda^1,\ldots,\lambda^k\right)$ no nulas como queríamos. ♣

Una **base del subespacio** $\mathbb{F}$ es una sucesión generadora e independiente. Aún no sabemos si hay bases en un subespacio arbitrario, pero, si las hay (y anticipamos que las hay) tienen el mismo número de elementos.

Teorema 25 *Si un subespacio $\mathbb{F}$ tiene bases $(a_1,\ldots,a_n)$ y $(b_1,\ldots,b_p)$, entonces $n=p$.*[5]

Demostración. Se aplica el teorema 24 permutando los papeles de $(a_1,\ldots,a_n)$ y $(b_1,\ldots,b_p)$, obteniéndose con la primera aplicación que $n\leq p$ y con la segunda que $p\leq n$. ♣

La **dimensión** de $\Bbbk^n$ o de uno de sus subespacios $\mathbb{F}$ es el número de elementos una de sus bases, denotado por $\dim(\mathbb{F})$. Como $\Bbbk^n$ tiene la base estándar con n elementos, deducimos que $\dim(\Bbbk^n)=n$. *¡Obvio pero de uso continuo!* Enseguida mostraremos que todos los subespacios de $\Bbbk^n$ tienen bases.

Diremos que $(a_1,\ldots,a_p)$ es **sucesión generadora mínima** de $\mathbb{F}$, un espacio o subespacio, si tiene longitud mínima entre todas las sucesiones que lo generan; o sea, no se puede generar con una sucesión más corta. Diremos que $(b_1,\ldots,b_q)$ es **sucesión independiente máxima** si tiene longitud máxima entre todas las sucesiones que de vectores de $\mathbb{F}$ que son independientes; o sea, no las hay más largas. Podría suceder que no existieran, bien porque por muy larga que fuera la sucesión siempre quedara algún $x\in\mathbb{F}$ no expresable como combinación lineal suya, o bien porque se pudieran elegir sucesiones tan largas como se quisiera, con lo cual la longitud no tendría un máximo. Advertimos que ninguna de las dos cosas sucede con subespacios de $\Bbbk^n$, pero puede pasar en espacios más generales.

[5] Ya dijimos que si $(a_1,\ldots,a_n)$ es independiente, los a_j son distintos entre sí, luego $\{a_1,\ldots,a_n\}$ tiene n elementos. Por eso se dice que dos bases tienen el mismo número de elementos (y no "la misma longitud") pues no hay equívoco.

Teorema 26 *Sean* $(a_1,\ldots,a_p)$ *y* $(b_1,\ldots,b_q)$ *respectivamente una sucesión generadora mínima y una sucesión independiente máxima del subespacio* $\mathbb{F}$ *de* $\Bbbk^n$. *Entonces ambas son bases de* $\mathbb{F}$.

Demostración. Falta ver que $(a_1,\ldots,a_p)$ es independiente y que $(b_1,\ldots,b_q)$ es generadora. Si $(a_1,\ldots,a_p)$ no lo fuera, habría una combinación $\lambda^1 a_1+\ldots+\lambda^p a_p=0$ con algún $\lambda_j\neq 0$. Entonces

$$a_j=\frac{1}{\lambda_j}\left(-\lambda^1 a_1-\ldots-\lambda^{j-1}a_{j-1}-\lambda^{j+1}a_{j+1}-\ldots-\lambda^p a_p\right)$$

y $(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_p)$, *que es más corta que* $(a_1,\ldots,a_p)$, generaría $\mathbb{F}$. En efecto,

$$\begin{aligned}x&=\xi^1 a_1+\ldots+\xi^{j-1}a_{j-1}+\xi^j a_j+\xi^{j+1}a_{j+1}+\ldots+\xi^p a_p\\&=\sum_{i=1}^{j-1}\xi^i a_i+\frac{\xi^j}{\lambda_j}\left(-\lambda^1 a_1-\ldots-\lambda^{j-1}a_{j-1}-\lambda^{j+1}a_{j+1}-\ldots-\lambda^p a_p\right)+\sum_{i=j+1}^{p}\xi^i a_i,\end{aligned}$$

mostrándose x como combinación de $(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_p)$. Esta contradicción prueba la independencia de $(a_1,\ldots,a_p)$.

Si $(b_1,\ldots,b_q)$ no fuese generadora habría un $x\in\mathbb{F}$ no expresable como combinación de $(b_1,\ldots,b_q)$. Afirmamos que se tendría que $(b_1,\ldots,b_q,x)$, *que es más larga que* $(b_1,\ldots,b_q)$, es independiente. Sea $0=\mu^1 b_1+\ldots\mu^q b_q+\theta x$. No puede ser $\theta\neq 0$ porque llegaríamos a $x=-\left(\mu^1/\theta\right)b_1-\ldots-\left(\mu^q/\theta\right)b_q$ lo que contradice que x no sea combinación de $(b_1,\ldots,b_q)$. Visto que $\theta=0$, queda $0=\mu^1 b_1+\ldots\mu^q b_q$ y, por la independencia de $(b_1,\ldots,b_q)$ es $\mu^1=\ldots=\mu^q=0$. Tal como anunciábamos, $(b_1,\ldots,b_q,x)$ es independiente, contradiciendo la hipótesis sobre $(b_1,\ldots,b_q)$. ♣

Teorema 27 *Todo subespacio* $\mathbb{F}\neq 0$ *de* $\Bbbk^n$ *tiene una base y* $\dim(\mathbb{F})\leq n$. *Más aún, si* $(b_1,\ldots,b_k)$ *es una sucesión independiente de vectores de* $\mathbb{F}$ *podemos añadir* $b_{k+1},\ldots,b_m$ *en* $\mathbb{F}$ *y que* $(b_1,\ldots,b_k,\ldots,b_m)$ *sea base de* $\mathbb{F}$.

Demostración. Consideramos sucesiones *independientes* de vectores de $\mathbb{F}$. Si ya nos han dado $(b_1,\ldots,b_k)$ serán solo de la forma $(b_1,\ldots,b_k,\ldots,b_h)$; si no nos han dado $(b_1,\ldots,b_k)$ empezamos con cualquier (b_1) de longitud 1, siendo $b_1\neq 0$. Afirmamos que $h\leq n$ porque si fuera $h>n$ tendríamos en $\Bbbk^n$, generado por $(e_1,\ldots,e_n)$, una sucesión independiente $(b_1,\ldots,b_k,\ldots,b_h)$ y $h>n$. Elegimos de entre todas $(b_1,\ldots,b_k,\ldots,b_m)$ de máxima longitud.[6] El teorema 26 implica que $(b_1,\ldots,b_k,\ldots,b_m)$ es una base. Se ha obtenido por el camino que $m=\dim(\mathbb{F})\leq n=\dim(\Bbbk^n)$. ♣

El teorema 27 dice que se puede ampliar un sucesión independiente hasta conseguir una base. Se puede rizar el rizo y mejorar el teorema 27 todavía más porque se puede ampliar un sucesión independiente, *añadiendo elementos de una sucesión generadora previa,* hasta conseguir una base. Es el llamado **teorema de Steinitz.**

Problema 56 *Sean* $(a_1,\ldots,a_p)$ *y* $(b_1,\ldots,b_q)$ *respectivamente una sucesión generadora y una sucesión independiente del subespacio* $\mathbb{F}$ *de* $\Bbbk^n$. *Renumerando* $(a_1,\ldots,a_p)$ *se pueden elegir* $a_1,\ldots,a_h$ *de modo que* $(b_1,\ldots,b_q,a_1,\ldots,a_h)$ *sea base de* $\mathbb{F}$. *Indicación: Considerar sucesiones* $(b_1,\ldots,b_q,a_\bullet,\ldots,a_\bullet)\ldots$ *¿cómo?.*

Con un lenguaje llano, buscando sobre todo recordar las proposiciones, sintetizamos lo obtenido, *importantísimo y de uso continuo.* Algunas afirmaciones están en las demostraciones, no en los enunciados.

1. Una sucesión de vectores independientes no puede ser estrictamente más larga que una generadora.
2. Una sucesión generadora mínima o una sucesión independiente máxima es una base.
3. Todas las bases de $\mathbb{F}$ subespacio de $\Bbbk^n$ tienen el mismo número de elementos $m\leq n$.
4. Todos los subespacios $\mathbb{F}$ de $\Bbbk^n$ tienen una base, obtenible alargando cualquier sucesión suya de vectores independientes. Esto se puede conseguir incluso eligiendo los nuevos elementos en una sucesión previa generadora de $\mathbb{F}$.

[6] Si no hubiésemos probado que las sucesiones que nos interesan tienen longitud $\leq n$, podría ser imposible elegir una de longitud máxima, ya que $\max(S)$ no está definido si $S\subset\mathbb{N}$ es infinito.

2.1.2. Espacios dados en implícitas

Ya hemos dicho que los teoremas precedentes no dan el método más efectivo para calcular bases y dimensiones. Vamos a estudiar esto ahora, advirtiendo que se va a hacer siempre suponiendo que $\mathbb{F}$ viene dado en forma implícita o paramétrica. Podría pensarse que quizás haya subespacios $\mathbb{F}$ que no se puedan expresar así, pero luego veremos que es un temor infundado. Suponemos fija en adelante una matriz $a \in \Bbbk^{m\times n}$ y que $\mathbb{F}$ viene dado, bien en implícitas por $ax = 0$, por tanto (¡ojo!) $\mathbb{F} \subset \Bbbk^n$, bien en paramétricas, como $\mathbb{F} = \lg(a_1, \ldots, a_n)$ (¡ojo!, como los $a_j \in \Bbbk^m$ ahora se tiene $\mathbb{F} \subset \Bbbk^m$). Varía por tanto la dimensión del espacio estándar donde $\mathbb{F}$ es un subconjunto.

Teorema 28 *Sea $\mathbb{F}$ el espacio de soluciones del sistema homogéneo $ax = 0$ con $a \in \Bbbk^{m\times n}$. Pasemos a a formas escalonadas b y c con b reducida y c solo escalonada. El número r de filas no nulas de b y c es el mismo y $\dim(\mathbb{F}) = n - r$. Al expresar la solución general en la forma $x = x^{q_1}u_1 + \ldots + x^{q_{n-r}}u_{n-r}$ con las variables libres $x^{q_1}, \ldots, x^{q_{n-r}}$, se tiene que $(u_1, \ldots, u_{n-r})$ es base de $\mathbb{F}$.*

Demostración. Sabemos que $bx = 0$ o $cx = 0$ sirven también como ecuaciones implícitas de $\mathbb{F}$. Si b tiene r filas no nulas, hay $n - r$ variables libres $x^{q_1}, \ldots, x^{q_{n-r}}$ y vectores $u_1, \ldots, u_{n-r}$ de modo que cada $x \in \mathbb{F}$, por ser solución de $bx = 0$, se puede escribir como $x = x^{q_1}u_1 + \ldots + x^{q_{n-r}}u_{n-r}$ con las x^{q_j} unívocamente fijadas por x. Esto implica que $(u_1, \ldots, u_{n-r})$ es base de $\mathbb{F}$. Al trabajar con c solo escalonada, podemos considerar las variables pivotales $(x^{t_1}, \ldots, x^{t_{n-s}})$ y para ciertos vectores v expresar cada solución de $cx = 0$ como $x = x^{t_1}v_1 + \ldots + x^{t_{n-s}}v_{n-s}$, aunque no decimos que sea $(q_1, \ldots, q_{n-r}) = (t_1, \ldots, t_{n-s})$. Sin embargo, es cierto que s es es el número de filas no nulas de c y este es el mismo número que el que tiene b; o sea, $r = s$. Como tanto $(u_1, \ldots, u_{n-r})$ como $(v_1, \ldots, v_{n-r})$ generan $\mathbb{F}$ y $(u_1, \ldots, u_{n-r})$ es base, $(v_1, \ldots, v_{n-r})$ es también base de $\mathbb{F}$. ♣

2.1.3. Espacios dados en paramétricas

Vamos a estudiar en general sucesiones $(a_1, \ldots, a_n)$ de vectores columna en $\Bbbk^m$. Definimos tres operaciones en paralelo con las operaciones elementales sobre matrices solo que ahora se harán sobre el espacio $\mathcal{S}$ de estas sucesiones. Las **operaciones elementales sobre sucesiones** serán Φ_{uv}, $\Phi_{uv}[\lambda]$ y $\Phi_w[\mu]$ con u, v, w índices en $\{1, \ldots, n\}$ y $\lambda, \mu \in \Bbbk$ con $\mu \neq 0$.

1. La sucesión $(b_1, \ldots, b_n) = \Phi_{uv}(a_1, \ldots, a_n)$ se obtiene permutando a_u y a_v y dejando invariables las demás a_w. Con símbolos,
$$b_u = a_v, \qquad b_v = a_u, \qquad b_w = a_w \quad \text{si } w \neq u, v.$$
2. La sucesión $(b_1, \ldots, b_n) = \Phi_{uv}[\lambda](a_1, \ldots, a_n)$ se obtiene sustituyendo a_u por la fila $a_u + \lambda a_v$ dejando invariables las demás a_w, incluida a_w. Con símbolos,
$$b_u = a_u + \lambda a_v, \qquad b_w = a_w \quad \text{si } w \neq u.$$
3. La sucesión $(b_1, \ldots, b_n) = \Phi_w[\mu](a_1, \ldots, a_n)$ se obtiene sustituyendo a_w por μa_w dejando invariables las demás a_u. Con símbolos,
$$b_w = \mu a_w, \qquad b_u = a_u \quad \text{si } w \neq u.$$

Diremos que dos sucesiones $(a_1, \ldots, a_n)$ y $(b_1, \ldots, b_n)$ son **sucesiones equivalentes (de vectores en $\Bbbk^m$),** y escribiremos $(a_1, \ldots, a_n) \sim (b_1, \ldots, b_n)$, si se puede transformar a en b con un número finito de operaciones elementales. Es como decir que si juntamos las sucesiones para que formen las columnas de matrices a y b, entonces se puede pasar de a a b con un número finito de operaciones columna. Es pura rutina probar que esta relación es de equivalencia y que estas operaciones tienen operaciones inversas, que son
$$(\Phi_{uv})^{-1} = \Phi_{uv}, \quad (\Phi_{uv}[\lambda])^{-1} = \Phi_{uv}[-\lambda], \quad (\Phi_w[\mu])^{-1} = \Phi_w[\mu].$$
Hemos empezado con sucesiones en $\Bbbk^m$ de vectores columna, pero también se pueden definir operaciones elementales sobre sucesiones de matrices fila $\left(a^1, \ldots, a^m\right)$ en $\Bbbk^{1\times n}$ y la equivalencia $\left(a^1, \ldots, a^m\right) \sim \left(b^1, \ldots, b^m\right)$ y quedan los detalles a cargo del lector.

Teorema 29 *Si $(a_1,\ldots,a_n)$ y $(b_1,\ldots,b_n)$ son equivalentes generan el mismo subespacio de $\Bbbk^m$. Hay un resultado análogo para sucesiones de matrices fila.*

Demostración. Sean $\mathbb{F}_a = \lg(a_1,\ldots,a_n)$ y $\mathbb{F}_b = \lg(b_1,\ldots,b_n)$. Basta que probemos que $\mathbb{F}_a = \mathbb{F}_b$ cuando se pasa de $(a_1,\ldots,a_n)$ a $(b_1,\ldots,b_n)$ con una sola operación elemental; por ejemplo, $(b_1,\ldots,b_n) = \Phi_{uv}[\lambda](a_1,\ldots,a_n)$. Un elemento arbitrario x de $\mathbb{F}_b$ se puede escribir (suponiendo $u<v$)

$$\begin{aligned} x &= \xi^1 b_1 + \ldots + \xi^u b_u + \ldots + \xi^v b_v + \ldots + \xi^n b_n \\ &= \xi^1 a_1 + \ldots + \xi^u (a_u + \lambda a_v) + \ldots + \xi^v a_v + \ldots + \xi^n a_n \\ &= \xi^1 a_1 + \ldots + \xi^u a_u + \ldots + (\xi^v + \xi^u\lambda) a_v + \ldots + \xi^n a_n \end{aligned}$$

y por tanto $x \in \mathbb{F}_a$ porque lo hemos expresado como combinación lineal de $(a_1,\ldots,a_n)$. Se tiene pues $\mathbb{F}_b \subset \mathbb{F}_a$. De modo análogo, como $a_u = b_u - \lambda a_v = b_u - \lambda b_v$,

$$\begin{aligned} x &= \eta^1 a_1 + \ldots + \eta^u a_u + \ldots + \eta^v a_v + \ldots + \eta^n a_n \\ &= \eta^1 b_1 + \ldots + \eta^u (b_u - \lambda b_v) + \ldots + \eta^v b_v + \ldots + \eta^n b_n \\ &= \eta^1 b_1 + \ldots + \eta^u b_u + \ldots + (\eta^v - \eta^u\lambda) a_v + \ldots + \eta^n a_n \end{aligned}$$

probando que $\mathbb{F}_a \subset \mathbb{F}_b$. Para la operación $\Phi_{uv}[\lambda]$ se ha probado que $\mathbb{F}_a = \mathbb{F}_b$ y los otros casos son análogos. Queda también para el lector el caso de las matrices fila. ♣

Para obtener una base para $\mathbb{F} = \lg(a_1,\ldots,a_n)$ se hacen operaciones columna en $a \in \Bbbk^{m\times n}$ de modo que se pase a $b = (b_1,\ldots,b_n) \in \Bbbk^{m\times n}$, b en forma escalonada reducida. Sea

$$b = \begin{pmatrix} 0 & 0 & \vdots & & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \vdots & & 0 & \cdots & 0 \\ b_1^{q_1} & \vdots & \vdots & & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & b_2^{q_2} & \cdots & & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & b_r^{q_r} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & & \vdots \\ * & * & \cdots & * & 0 & \cdots & 0 \end{pmatrix}$$

Los pivotes de columna $b_i^{q_i}$ son todos 1 y en su fila, que es la fila q_i, todos los demás términos son 0. Por el teorema 29, $\mathbb{F}$ es también $\lg(b_1,\ldots,b_n)$. Si quitamos $b_{r+1},\ldots,b_n$, que son cero, $\mathbb{F} = \lg(b_1,\ldots,b_r)$. Los generadores $b_1,\ldots,b_r$ son independientes. La razón es que si $0 = \lambda^1 b_1 + \ldots + \lambda^r q_r$, los coeficientes en los lugares $q_1 < q_2 < \ldots < q_r$ valen respectivamente $\lambda^1,\ldots,\lambda^r$ por cómo son la filas $q_1 < q_2 < \ldots < q_r$. Por consiguiente $\lambda^1 = \ldots = \lambda^r = 0$ y $(b_1,\ldots,b_r)$ es una base de $\mathbb{F}$. Claramente, $\dim(\mathbb{F})$ es el número de columnas no nulas de la forma escalonada reducida. Si solo nos interesa la dimensión de $\mathbb{F}$ basta con transformar a con operaciones columna elementales a una forma c escalonada aunque no esté reducida. Si c tiene r columnas no nulas, también b transformada de c hasta la forma reducida, tendrá r columnas no nulas y será la dimensión de $\mathbb{F}$. Incluso si necesitamos una base, las columnas de c forman base. La justificación es esta: por el teorema 29 generan $\mathbb{F}$ y sabemos de antes que $\dim(\mathbb{F}) = r$, que es justamente la longitud de $(c_1,\ldots,c_r)$. El teorema 29 nos dice que $(c_1,\ldots,c_r)$ es base de $\mathbb{F}$.

Problema 57 *Sea a la matriz*

$$a = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & h & 0 \end{pmatrix}$$

Consideramos el espacio $\mathbb{F}$ de soluciones del sistema $ax = 0$ y $\mathbb{G}$ generado por las columnas de a. Calcular bases y dimensiones de $\mathbb{F}$ y $\mathbb{G}$. ♦

Solución. El paso a forma escalonada reducida es

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & h & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & h & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

y el sistema y la solución general son

$$\begin{cases} x^1 = -x^3 \\ x^2 = 0 \end{cases}, \quad \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = x^3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Por tanto, solo el vector $(-1,0,1)^\top$ da una base de $\mathbb{F}$ y $\dim(\mathbb{F}) = 3-2 = 1$ pues en forma escalonada hay dos filas no nulas.

Para $\mathbb{G}$ las operaciones son

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & h & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ h & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h & 0 & 0 \end{pmatrix}$$

.La dimensión es 2 y una base es $\left((1,0,h)^\top, (0,1,0)^\top\right)$. ♦

Problema 58 *Hacer lo mismo que en el problema anterior para la matriz*

$$a = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 1 & 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix},$$

El lector puede convertir problemas del capítulo anterior en problemas sobre bases y dimensiones, que son muy pesados pero fáciles de resolver con lo visto sobre sistemas y cálculo matricial. Siguiendo con problemas pesados, pero que tarde o temprano son necesarios, nos preguntamos ¿cómo se pasa de ecuaciones implícitas a paramétricas y viceversa? La utilidad de saber pasar de unas a otras ecuaciones es porque ciertos cálculos sobre subespacios se hacen mejor según estén sus ecuaciones de una u otra forma. En realidad ya sabemos pasar de implícitas a paramétricas porque en la subsección anterior vimos que al resolver el sistema dando la solución general, aparece una base y tener una base permite enseguida escribir la ecuación en paramétricas.

Supongamos ahora que tenemos la ecuación en paramétricas y $\mathbb{F} = \lg(a_1, \ldots, a_n) \subset \Bbbk^m$. Sabemos que $y \in \mathbb{F}$ si y solo si existen $x^1, \ldots, x^n \in \Bbbk$ tales que $y = x^1 a_1 + \ldots + x^n a_n$. Si unimos los vectores columna $a_1, \ldots, a_n$ para obtener la matriz $a \in \Bbbk^{m\times n}$, queda claro que $y \in \mathbb{F}$ equivale a que $ax = y$, con y "dato variable", tenga al menos una solución x, o, dicho de otra manera, que $ax = y$ sea compatible. Si seguimos el procedimiento para resolver $ax = y$ ampliando a a $(a \mid y)$, realizando operaciones fila sobre $(a \mid y)$ para convertirla en $(b \mid z)$ con b escalonada reducida. Como estas operaciones se realizan en toda $(a \mid y)$, la última columna z tiene unos coeficientes z^i dependientes de y, digamos que $z^i = f^i(y)$, $1 \le i \le m$. Si las filas nulas de b son $b^{r+1}, \ldots, b^m$, sabemos que la compatibilidad del sistema equivale a que y verifique las últimas $m-r$ ecuaciones $z^i = f^i(y) = 0$, $r+1 \le i \le m$. Estas son las ecuaciones implícitas de $\mathbb{F}$, pues expresan los $y \in \mathbb{F}$ como soluciones de un sistema de ecuaciones.

Problema 59 *Para la matriz real*

$$a = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix},$$

escribir en forma implícita el espacio de soluciones generado por sus columnas. ♦

Solución. Transformamos $(a \mid y)$ a $(b \mid z)$ con operaciones fila buscando que b esté escalonada

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & y^1 \\ 2 & 5 & 8 & y^2 \\ 3 & 6 & 9 & y^3 \end{array}\right) \xrightarrow{1} \left(\begin{array}{ccc|c} 1 & 4 & 7 & y^1 \\ 0 & -3 & -6 & y^2 - 2y^1 \\ 0 & -6 & -12 & y^3 - 3y^1 \end{array}\right) \xrightarrow{2} \left(\begin{array}{ccc|c} 1 & 4 & 7 & y^1 \\ 0 & -3 & -6 & y^2 - 2y^1 \\ 0 & 0 & 0 & \left(y^3 - 3y^1\right) - 2\left(y^2 - 2y^1\right) \end{array}\right).$$

En $\xrightarrow{1}$ se hacen las operaciones $F_{21}[-2]$ y $F_{31}[-3]$ y en $\xrightarrow{2}$ es $F_{32}[-2]$. Como sistema queda

$$\begin{cases} x^1 + 4x^2 + 7x^3 = y^1 \\ -3x^2 - 6x^3 = y^2 - 2y^1 \\ 0 = \left(y^3 - 3y^1\right) - 2\left(y^2 - 2y^1\right) \end{cases}$$

La última ecuación $y^1 - 2y^2 + y^3 = 0$ da el espacio generado por las columnas en implícitas. ♦

Problema 60 *Poner en implícitas el subespacio de* $\mathbb{k}^3$ *generado por las columnas de*

$$a = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Podríamos añadir aquí el espacio de filas y columnas de una matriz, que está en una sección más adelante en el capítulo y donde todo se sigue haciendo "con coordenadas y ecuaciones" pero creemos que ya debemos de entrar en los espacios vectoriales en general.

2.2. Espacios vectoriales

Un **espacio vectorial** es un objeto matemático formado por un conjunto $\mathbb{E}$, un cuerpo $\mathbb{k}$, y dos funciones de $\mathbb{E} \times \mathbb{E}$ en $\mathbb{E}$ y de $\mathbb{k} \times \mathbb{E}$ en $\mathbb{E}$. Para la primera, el elemento que se asigna a $(x, y) \in \mathbb{E} \times \mathbb{E}$ se denota por $x + y$, y para la segunda, el elemento que se asigna a (λ, x) se denota por $\lambda \cdot x$. Estas funciones, comúnmente llamadas operaciones, cumplen los axiomas

1. **Asociatividad.** Para todo $x, y, z \in \mathbb{E}$ se tiene que $(x + y) + z = x + (y + z)$.
2. **Conmutatividad.** Para todo $x, y \in \mathbb{E}$ se tiene $x + y = y + x$.
3. **Existencia de cero**. Existe un elemento en $\mathbb{E}$, que denotaremos por 0, tal que para todo $x \in \mathbb{E}$ es $x + 0 = 0 + x = x$.
4. **Existencia de opuesto.** Para cada $x \in \mathbb{E}$ existe un elemento que se denota por $-x$ tal que $x + (-x) = 0$.
5. **Distributividad.** Para todo $x, y \in \mathbb{E}$ y todo $\lambda, \mu \in \mathbb{k}$ se tiene que $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$ y $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$.
6. Para todo $x \in \mathbb{E}$ y todo $\lambda, \mu \in \mathbb{k}$ se tiene que $(\lambda\mu) \cdot x = \lambda \cdot (\mu \cdot x)$ y $1 \cdot x = x$.

Los elementos de $\mathbb{E}$ se llaman **vectores** y los de $\mathbb{k}$ **escalares**, aunque nos referiremos a ellos muchas veces como "números". La operación $+$ es la **suma** y $\cdot$ el **producto por escalares**. A 0 en el axioma **3** se le llama el **cero** (del espacio vectorial) y a $-x$ el **opuesto** o **inverso aditivo** de x. El uso del artículo "el" puede hacer pensar que 0 y $-x$ son únicos, y así es, pero no hay que postularlo porque es consecuencia de los axiomas. Si por ejemplo, hubiera 0 y $0'$ cumpliendo **3** se tendría $0 = 0 + 0' = 0'$ (¿qué se usa en las igualdades?). Damos por hecho que el lector entiende el significado de los paréntesis como signo de prioridad, de modo que mucho de lo que dicen los axiomas es que ciertas operaciones pueden combinarse sin que importe el orden en que se hagan. Por ejemplo, $(\lambda\mu) \cdot x = \lambda \cdot (\mu \cdot x)$ nos dice que si multiplicamos *primero* los números λ y μ resultando $\theta \in \mathbb{k}$ y *en segundo lugar* multiplicamos $\theta = \lambda\mu$ por x, obtenemos lo mismo que si multiplicamos *primero* μ por x, resultando $y \in \mathbb{E}$, *después* multiplicamos λ por $y = \mu \cdot x$. Si $\mathbb{k}$ es $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$, el espacio vectorial se llama **racional**, **real** o **complejo.**

El ejemplo principal de espacio vectorial es $\mathbb{E} = \mathbb{k}^{m\times n}$ con la suma y producto por escalares (y no el producto entre matrices). Vemos que se cumplen los axiomas pues esto es lo que dice el teorema 8. El espacio $\mathbb{k}^m = \mathbb{k}^{m\times 1}$ de las matrices columna, el que hemos llamado espacio estándar en la sección anterior, se incluye dentro de $\mathbb{E} = \mathbb{k}^{m\times n}$ que es más general.

Si somos estrictos un espacio vectorial es una *terna* $(\mathbb{E}, +, \cdot)$ cuyas operaciones $+$ y $\cdot$ verifican ciertos axiomas, y no un simple *conjunto* $\mathbb{E}$, por mucho que se hable (y se hablará) del "espacio vectorial $\mathbb{E}$". Tanta precisión es necesaria pues con otro par de operaciones $(\oplus, \odot)$ diferentes, $(\mathbb{E}, +, \cdot)$ y $(\mathbb{E}, \oplus, \odot)$ serían diferentes espacios vectoriales. Sin embargo, en la práctica, las operaciones en $\mathbb{E}$ están fijadas desde el principio, y se habla del espacio vectorial $\mathbb{E}$ sin más. Los dos problemas siguientes muestran espacios raros pero sirven para que el lector maneje y recuerde los axiomas de espacio vectorial.

Problema 61 *En el conjunto* $\mathbb{R}^2$ *se dan dos operaciones*

$$(x_1, x_2) \oplus (y_1, y_2) = (x_1 + y_1, x_2 + y_2), \qquad \lambda \odot (x_1, x_2) = (\lambda x_1, 0).$$

¿Qué axiomas de espacio vectorial se verifican? ¿Es un nuevo espacio vectorial?

Problema 62 *Se toma una biyección* $\phi : \mathbb{R}^n \to \mathbb{R}^n$ *y se definen en* $\mathbb{R}^n$ *operaciones*

$$x \oplus y = \phi^{-1}(\phi(x) + \phi(y)), \qquad \lambda \odot x = \phi^{-1}(\lambda \phi(x))$$

pasando con ϕ *al segundo ejemplar de* $\mathbb{R}^n$*, operando allí del modo usual y volviendo con lo obtenido a* $\mathbb{R}^n$ *con* ϕ^{-1}*. ¿Es* $(\mathbb{R}^n, \oplus, \odot)$ *un espacio vectorial?*

Hay una serie de propiedades de uso continuo para manejar vectores. La suma $x + (-y)$ se abrevia a $x - y$. Suele prescindirse también del punto en $\lambda \cdot x$ y solo se pone por claridad o por énfasis. Las identidades que siguen son obvias en $\mathbb{R}^n$. La razón de citar el teorema es que estas propiedades son consecuencia de los axiomas de la definición. No se trata de aprender la demostración de memoria sino de convencerse que estas propiedades "evidentes" se cumplen en cualquier espacio vectorial. De manera análoga a como pasa con un cuerpo se tienen las propiedades del teorema.

Teorema 30 *Se verifican las identidades para todo* $x \in \mathbb{E}$ *y* $\lambda \in \Bbbk$*,*

$$\lambda \cdot 0_{\mathbb{E}} = 0_{\mathbb{E}} = 0_{\Bbbk} \cdot x_{\mathbb{E}}, \quad \lambda \cdot (-x) = -(\lambda \cdot x) = (-\lambda) \cdot x, \quad \textit{si } \lambda \cdot x = 0_{\mathbb{E}}, \textit{ entonces } \lambda = 0_{\Bbbk} \textit{ o } x = 0_{\mathbb{E}}.$$

Demostración. Por ejemplo, $\lambda \cdot 0_{\mathbb{E}} = \lambda \cdot (0_{\mathbb{E}} + 0_{\mathbb{E}}) = \lambda \cdot 0_{\mathbb{E}} + \lambda \cdot 0_{\mathbb{E}}$, luego $\lambda \cdot 0_{\mathbb{E}} = 0_{\mathbb{E}}$. Otro caso, supongamos que $\lambda \cdot x = 0_{\mathbb{E}}$. Si $\lambda = 0_{\Bbbk}$ hemos acabado y si no, se multiplica por λ^{-1} y

$$0_{\mathbb{E}} = \lambda^{-1} \cdot 0 = \lambda^{-1} \cdot (\lambda \cdot x) = \left(\lambda^{-1} \cdot \lambda\right) \cdot x = 1 \cdot x = x.$$

♣

En lo sucesivo quitaremos casi siempre los puntos y la distinción entre $0_{\mathbb{E}}$ y $0_{\Bbbk}$.

Problema 63 *Problema rápido. Acabamos de ver en el teorema que para todo* $x \in \mathbb{E}$ *es* $0 \cdot x = 0$*. Supongamos que tenemos* $\mathbb{E}$ *con* $+$ *y* $\cdot$ *que cumple, en principio, todos los axiomas excepto la existencia de opuesto, pero sí cumple "para todo* $x \in \mathbb{E}$ *es* $0 \cdot x = 0$*". Probar que se cumple entonces el axioma de la existencia de opuesto. ¿Cuál es* $-x$*?*

Hasta ahora, aparte de los espacios de matrices y unos ejemplos "raros" no hemos visto más ejemplos de espacios vectoriales. Muchos ejemplos de espacios vectoriales tienen como vectores matrices, de tipo más o menos restringido, que se suman y multiplican por escalares con las operaciones usuales. No son operaciones "raras" como en los ejemplos, sino que $\mathbb{E}$ no es todo $\Bbbk^{m \times n}$. Cuando veamos la definición de subespacio estos ejemplos serán subespacios del espacio de todas las matrices. Hay sin embargo un tipo de espacios vectoriales que a veces cuesta intuir porque sus elementos, sus vectores, no son matrices sino son *funciones*. En efecto, muchos espacios vectoriales tienen como vectores cierto tipo de funciones de un cierto conjunto I en $\Bbbk$ (en los ejemplos más interesantes, $\Bbbk$ es $\mathbb{R}$ o $\mathbb{C}$).

Dadas funciones $f, g : I \to \Bbbk$ y $\lambda \in \Bbbk$ se definen nuevas funciones de I en $\Bbbk$ por

$$(f + g)(x) = f(x) + g(x), \qquad (\lambda \cdot f)(x) = \lambda f(x).$$

Se entiende mejor con palabras: la nueva función $f + g$ manda x al número resultante de sumar $f(x)$ y $g(x)$ y λf manda x al producto de los números λ y $f(x)$.[7] Remarcamos que $(f + g)(x) = f(x) + g(x)$ no es algo "evidente", sino una *definición que se elige*, aunque es muy razonable y resulta ser interesante. Denotamos por $\mathcal{F}(I, \Bbbk)$ al **espacio vectorial de las funciones** de I en $\Bbbk$.

Teorema 31 *Con estas operaciones, las funciones de* I *en* $\Bbbk$ *forman un espacio vectorial con cuerpo* $\Bbbk$*.*

[7] Casi siempre se prescindirá del punto $\cdot$, luego $\lambda \cdot f = \lambda f$ y $(\lambda f)(x) = \lambda f(x)$.

Demostración. Empezamos diciendo que el cero de $\mathcal{F}(I,\Bbbk)$ va a ser la **función cero**, definida como la que manda todo x a $0\in\Bbbk$. Esta función se denota por 0, luego $0(x)=0$, si bien 0 significa dos cosas distintas en $0(x)=0$, siendo la *función* 0 a la izquierda y el *escalar* 0 a la derecha. El opuesto de f se va a denotar por $-f$ y se define por $(-f)(x)=-f(x)$. Sucede algo similar, pues $-$ tiene dos significados distintos: es una notación para el opuesto de f a la izquierda y la notación del opuesto en $\Bbbk$ a la derecha. Con palabras, $-f$ manda cada x al opuesto de $f(x)$. Verificamos

$$(f+0)(x)=f(x)+0(x)=f(x)+0=f(x),\qquad (f+(-f))(x)=f(x)+(-f(x))=0=0(x)$$

y, análogamente, $0+f=f$ y $(-f)+f=0$.

Verificamos alguna otra propiedad, por ejemplo $(\lambda\mu)\cdot f=\lambda\cdot(\mu\cdot f)$. Hay que percatarse que estamos verificando una igualdad de funciones y que $f=g$ por definición si y solo si para cualquier x es $f(x)=g(x)$. Calculamos

$$[(\lambda\mu)\cdot f](x)=(\lambda\mu)f(x),\qquad [\lambda\cdot(\mu\cdot f)](x)=\lambda[(\mu\cdot f)(x)]=\lambda[\mu f(x)]$$

y hemos acabado porque la asociatividad en $\Bbbk$ dice que $(\lambda\mu)f(x)=\lambda[\mu f(x)]$ para los tres *números* $\lambda,\mu,f(x)$. ♣

Se puede particularizar esta definición tomando un I concreto; por ejemplo,

1. Si $I=\{1,2,\ldots,n\}$, un elemento $x\in\mathcal{F}(\{1,2,\ldots,n\},\Bbbk)$ es, por definición de producto cartesiano, un elemento de $\Bbbk^n=\Bbbk\times\ldots\times\Bbbk$, que se suele representar por $(x_1,\ldots,x_n)$. Por tanto, $\mathcal{F}(\{1,2,\ldots,n\},\Bbbk)$ es el **espacio estándar** $\Bbbk^n$, independientemente de que nuestra costumbre sea ver $\Bbbk^n$ como "listas de números" y no como funciones. (Pero ¿qué es una "lista ordenada" sino un cierto tipo de función?)

2. Si I es los números naturales $\mathbb{N}$. Como en el caso anterior, $x\in\mathcal{F}(\mathbb{N},\Bbbk)$ es una función que se suele representar por (x_n) siendo x_n la imagen de n por x, eludiendo la notación $x(n)$. Así pues $\mathcal{F}(\mathbb{N},\Bbbk)$ está formado por las **sucesiones en** $\Bbbk$, que son fundamentales en Análisis con $\Bbbk=\mathbb{R},\mathbb{C}$. Se suman sucesiones y multiplican por números "término a término"; es decir $(x_n)+(y_n)=((x+y)_n)$ y $\lambda(x_n)=(\lambda x_n)$.

3. Si I es un intervalo de $\mathbb{R}$ y $\Bbbk=\mathbb{R},\mathbb{C}$ es $\mathcal{F}(I,\Bbbk)$ el conjunto de todas las funciones del intervalo I en $\mathbb{R}$ o $\mathbb{C}$; o sea, las funciones reales o complejas definidas sobre I. Este espacio es demasiado grande e interesan más bien subespacios[8] suyos como las funciones continuas, diferenciables, acotadas, etc. definidas en I con valores en $\Bbbk=\mathbb{R},\mathbb{C}$. Para probar, por ejemplo, que $\mathcal{C}(I,\mathbb{R})$, las funciones continuas del intervalo I en $\mathbb{R}$, forman un espacio vectorial, necesitamos saber *un teorema de Análisis que no de Álgebra Lineal* que dice que $f+g$ y λf son continuas si f y g lo son. Con esto ya casi no hay que hacer nada más pues una propiedad como $f+g=g+f$ es cierta para $\mathcal{C}(I,\mathbb{R})$ pues se cumple en el espacio más grande $\mathcal{F}(I,\mathbb{R})$. Si el lector no sabe probar que el espacio $\mathcal{B}(I,\mathbb{R})$ de las funciones acotadas sobre I es un espacio vectorial es porque no sabe probar $f+g$ y λf son acotadas si f y g lo son, que es *un teorema de Análisis y no de Álgebra Lineal*. El lenguaje algebraico de espacios vectoriales, que parece tener como justificación primordial el estudio de sistemas lineales y los problemas geométricos del plano y del espacio, es esencial para la formulación del Análisis (como el lector verá en el curso paralelo a este).

   Los espacios recién descritos de funciones continuas, diferenciables, acotadas, etc, son de dimensión infinita, concepto aún no definido, pero que para nosotros significa de momento "inmanejables a través de coordenadas". Este curso se va a centrar en espacios de dimensión finita pero no debe pensarse que lo que veamos sea irrelevante para el Análisis. El Análisis se relaciona, y no es más que un ejemplo, con Álgebra Lineal en dimensión finita al tratar las Ecuaciones Diferenciales Lineales, que normalmente no es materia de Primer Curso. Se debe tanto a que el espacio de soluciones de estas ecuaciones es de dimensión finita como a que su cálculo requiere la exponencial de matrices y ahí sí interviene el Álgebra Lineal.

4. ¿Qué es un **polinomio**? Hay dos candidatos a la definición. La primera es una "expresión"

$$f(X)=a_0+a_1X+a_2X^2+\ldots+a_{n-1}X^{n-1}+a_nX^n,\qquad a_i\in\Bbbk,\qquad a_n\neq 0,\tag{2.3}$$

[8] No se ha definido aún formalmente "subespacio".

y los polinomios se suman y multiplican entre ellos y se multiplican por elementos de $\Bbbk$ con las reglas conocidas. Se puede decir que $f(X)$ es esencialmente una sucesión $(a_0, a_1, a_2, \ldots, a_{n-1}, a_n, 0, 0, \ldots)$ donde a partir de un cierto lugar n todos los elementos a_j, $j > n$ son nulos. En el sentido laxo de la definición, tendríamos un "subespacio" de $\mathcal{F}(\mathbb{N}, \Bbbk)$. Si se usa la primera notación es por su utilidad para manejar productos de polinomios, aunque, repetimos, el producto de polinomios no es una operación para dar al **espacio de polinomios** estructura de espacio vectorial. Este espacio se denotará por $\Bbbk[X]$.

La segunda definición de polinomio es un cierto tipo de *función* de $\Bbbk$ en $\Bbbk$; precisamente la dada por la fórmula de más arriba. Resulta intuitivamente cierto que la *expresión* $f(X)$ determina la *función* $f : \Bbbk \to \Bbbk$ de modo unívoco (a cada polinomio una función y solo una) *pero es falso.* Si $\Bbbk$ tiene infinitos elementos, caso de $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$, esto es cierto y, como es el caso más frecuente, no hay peligro al, ver según se quiera, un polinomio como una función o una "expresión" (léase, la sucesión de sus coeficientes, con $a_j = 0$ a partir de un cierto n). En un cuerpo como $\mathbb{Z}_2, Z_3$ o $\mathbb{Z}_p$,

$$f_2(X) = X(X-1), \quad f_3(X) = X(X-1)(X-2), \ldots, f_p(X) = X(X-1)\ldots(X-(p-1))$$

cumplen respectivamente que la función asociada f_p es cero pero $f_p \neq 0$ (hay coeficientes $\neq 0$). Así pues $f_p \neq 0$ como polinomio con la primera definición, pero como función da la función cero como el polinomio 0.

Para no complicar demasiado la situación, consideraremos por defecto que "polinomio" es sinónimo de **función polinomial (o polinómica)**, que es la que tiene una fórmula $f(X) = a_0 + a_1 X + a_2 X^2 + \ldots + a_n X^n$ y nos limitaremos casi siempre a que $\Bbbk$ sea $\mathbb{Q}, \mathbb{R}$ o $\mathbb{C}$ para evitar problemas. El espacio $\Bbbk[X]$ es de dimensión infinita, pero, una vez fijado $n \in \mathbb{N}$, se puede considerar $\Bbbk_n[X]$, los **polinomios de grado menor o igual a** n, que es un espacio vectorial y, cuando se defina "dimensión" veremos que tiene dimensión finita $n+1$.

5. Si $\mathbb{E}_1, \ldots, \mathbb{E}_k$ son espacios vectoriales el **producto cartesiano de espacios vectoriales** es un espacio vectorial si definimos $(x_1, \ldots, x_k) + (y_1, \ldots, y_k) = (x_1 + y_1, \ldots, x_k + y_k)$ y $\lambda(x_1, \ldots, x_k) = (\lambda x_1, \ldots, \lambda x_k)$, como el lector comprobará.

Es muy importante que el lector se acostumbre a manejar espacios vectoriales generales donde el obstáculo es no poder ver sus elementos como matrices de uno u otro tipo, y hay que fundamentar cuanto se haga solo con los axiomas de espacio vectorial. La manera de salvar ese obstáculo es ir examinando minuciosamente todas las demostraciones que vienen ahora, donde las coordenadas no aparecen. Es un grave error pensar que porque muchos problemas se hacen en espacios "con coordenadas", y donde el objetivo es la aplicación de un algoritmo, se pueden tranquilamente ignorar las demostraciones y los casos más generales. Hemos empezado el capítulo con el espacio vectorial estándar para que el lector no se pierda en un mundo abstracto, pero ahora tiene que familiarizarse con lo general.

En adelante, letras como $\mathbb{E}, \mathbb{F}, \mathbb{G}\ldots$ denotarán espacios vectoriales con un mismo cuerpo $\Bbbk$.

2.3. Subespacios vectoriales

En toda la sección $\mathbb{E}$ será un espacio vectorial sobre el cuerpo $\Bbbk$. En los ejemplos, si no se precisa, $\Bbbk = \mathbb{R}$. Si tenemos una sucesión de vectores $(a_1, \ldots, a_n)$ en $\mathbb{E}$, definiremos una **combinación lineal** de $(a_1, \ldots, a_n)$ como un vector $v \in \mathbb{E}$ de la forma $v = \lambda^1 a_1 + \ldots + \lambda^n a_n$, siendo $\lambda^1, \ldots, \lambda^k$ elementos del cuerpo $\Bbbk$. Evidentemente, al variar la sucesión $(\lambda^1, \ldots, \lambda^k)$ varía x y hay muchas combinaciones de $(a_1, \ldots, a_n)$. El lector puede constatar que hemos copiado palabra por palabra la definición correspondiente con $\mathbb{E} = \Bbbk^m$. Lo que resulta nuevo es que, de momento, no tenemos coordenadas. Si tomamos por ejemplo como $\mathbb{E}$ el espacio de las funciones continuas de $\mathbb{R}$ en $\mathbb{R}$ y $a_1 = \cos t$, $a_2 = \sin t$, el coseno y el seno, podemos decir que $f(t) = 3\cos t - 5\sin t$ es un elemento de $\mathbb{E}$ combinación lineal de a_1 y a_2, pero no se puede simplificar del modo que en $\mathbb{E} = \mathbb{R}^2$ se puede decir que $3(2,2)^\top - 5(-1,7)^\top = (11,-29)^\top$. Esta sensación extraña de que "no hay ejemplos" en cuanto abandonamos $\mathbb{E} = \Bbbk^{m\times n}$ es inevitable.

Una sucesión de vectores $(a_1, \ldots, a_n)$ en $\mathbb{E}$ es **(linealmente) independiente** si al poner $0 \in \mathbb{E}$ como combinación lineal $0 = \lambda^1 a_1 + \ldots + \lambda^n a_n$ se tiene necesariamente que $\lambda^1 = \ldots = \lambda^n = 0$. Si la sucesión no es (linealmente) independiente se dice que es **(linealmente) dependiente**. Al preguntarnos en el ejemplo de antes si $(\cos t, \sin t)$ es linealmente independiente, la respuesta es sí, pero no se puede echar

mano de los sistemas lineales sino que ha de trabajarse directamente con la definición. En efecto, si tomamos λ^1 y λ^2 tales que $\lambda^1 \cos t + \lambda^2 \sin t = 0$, y aquí 0 es la función 0, evaluamos $\lambda^1 \cos t + \lambda^2 \sin t$ en $t = 0$ y $t = \pi/2$ y

$$0 = \lambda^1 \cos(0) + \lambda^2 \sin(0) = \lambda^1 \cdot 1 + \lambda^2 \cdot 0 = \lambda^1, \quad 0 = \lambda^1 \cos(\pi/2) + \lambda^2 \sin(\pi/2) = \lambda^1 \cdot 0 + \lambda^2 \cdot 1 = \lambda^2.$$

El lector puede pensar que si le dan la sucesión de tres funciones $\left(\cos t, 1 - t^3, \log\left(1 + t^2\right)\right)$ quizás podría probar la independencia con las ideas anteriores, pero no tiene disponible todo lo que sabe sobre ecuaciones lineales. En el Álgebra Lineal, cuando $\mathbb{E}$ está formado por funciones, se pueden necesitar procedimientos diferentes a cuando $\mathbb{E}$ está formado por matrices si bien en este curso nos centraremos en espacios que permiten el cálculo con coordenadas.

Un **subespacio vectorial** $\mathbb{F}$ de $\mathbb{E}$ es un subconjunto *no vacío* de $\mathbb{E}$ que cumple que si $x, y \in \mathbb{F}$ y $\lambda \in \Bbbk$, entonces $x + y$ y λx también están en $\mathbb{F}$. Sin duda $\mathbb{F} = \mathbb{E}$ y $\mathbb{F} = 0$ son subespacios vectoriales, el **subespacio total** y el **subespacio cero**, que se suelen llamar los **subespacios triviales** (por lo fácil que es ver que lo son). Los restantes subespacios de $\mathbb{E}$ se llaman **subespacios propios**. Las definiciones son idénticas al caso $\mathbb{E} = \Bbbk^m$ y tal como allí se hizo repetimos que 0, el cero de $\mathbb{E}$, siempre está en cualquier subespacio. El lector comprobará, una vez que le han advertido que $0 = 0_{\mathbb{E}} \in \mathbb{F}$, que los vectores de $\mathbb{F}$ con las operaciones de $\mathbb{E}$ restringidos a ellos, es un espacio vectorial. Con esta observación, todo lo que se pruebe o defina para un espacio vectorial general vale para sus subespacios mientras solo se usen los axiomas de espacio vectorial. Damos una larga lista de ejemplos de subespacios vectoriales y, por lo que acabamos de decir, de espacios vectoriales.

1. El espacio estándar $\Bbbk^m$ tiene muchos subespacios vectoriales $\mathbb{F}$. Pueden expresarse como $\mathbb{F} = \lg(a_1, \ldots, a_n)$ o como espacio de soluciones de un sistema homogéneo.

2. Tomamos $\mathbb{E} = \Bbbk^{m \times m}$ (matrices *cuadradas*), que tiene muchos subespacios de fácil descripción. Por ejemplo las matrices triangulares (tanto superiores como inferiores), y las matrices simétricas y antisimétricas. También se puede usar la traza (recordemos que $\operatorname{tr}(a) = a_1^1 + \ldots + a_m^m$) y considerar como $\mathbb{F}$ el subespacio de matrices de traza nula.

3. Generalizando a $\mathbb{E} = \Bbbk^{m \times n}$ podemos construir subespacios fijando una matriz b de dimensiones adecuadas y definir $\mathbb{F} = \{a \in \Bbbk^{m \times n} \mid ab = 0\}$ o bien $\mathbb{F} = \{a \in \Bbbk^{m \times n} \mid ab = ba\}$.

4. En cualquier $\mathbb{E}$ elegimos una sucesión de sus elementos $(a_1, \ldots, a_n)$ y definimos $\mathbb{F} = \lg(a_1, \ldots, a_n)$, el conjunto de las combinaciones lineales de $(a_1, \ldots, a_n)$. Tenemos entonces un subespacio vectorial, como enseguida se ve. Se llamará el **subespacio generado por** $(a_1, \ldots, a_n)$. Si por ejemplo, $\mathbb{E}$ es el espacio de las funciones continuas de $\mathbb{R}$ en $\mathbb{R}$, podemos considerar $\mathbb{F} = \lg(\cos t, \sin t)$ y decir que $f(t) = 3\cos t - 5\sin t \in \mathbb{F}$.

5. El espacio $\Bbbk[X]$ de los polinomios tiene muchos subespacios de fácil descripción, poniendo alguna restricción sobre los coeficientes o pidiendo, si los vemos como funciones, que tomen determinado valor. Por ejemplo, los polinomios cuyos coeficientes impares son nulos (todo $a_{2k-1} = 0$ para $k \in \mathbb{N}$) forman un subespacio. También los que cumplen $f(1) = 0$. Es muy importante el subespacio $\Bbbk_n[X]$ de los polinomios de grado $\leq n$.[9]

6. Si I tiene cierta "estructura", al considerar $\mathcal{F}(I, \Bbbk)$ aparecen funciones "distinguidas" que forman subespacios de $\mathcal{F}(I, \Bbbk)$. Para I un intervalo de $\mathbb{R}$ y $\Bbbk = \mathbb{R}$ o $\mathbb{C}$, tenemos los subespacios de funciones continuas, derivables, acotadas, etc. o que son soluciones de ciertas ecuaciones diferenciales. (No hay que saber las definiciones con precisión.) Si $I = \Bbbk$ tenemos el subespacio de las funciones polinomiales y si I es un conjunto infinito de cualquier tipo de objetos, tenemos el subespacio de las funciones f que se anulan en todos los $x \in I$ excepto en un número finito de elementos (si se prefiere, se anulan fuera de un subconjunto finito de I). Pueden obtenerse muchas veces ejemplos de subespacios poniendo a las funciones la condición de anularse en ciertos $x_0 \in I$.

7. En $\mathbb{E}_1 \times \mathbb{E}_2$, el subconjunto $\mathbb{E}_1 \times \{0\}$ es un subespacio vectorial, que se suele identificar con $\mathbb{E}_1$.

[9] Una digresión necesaria: El grado de $f(X) = a_0 + a_1 X + \ldots + a_n X^n$ es n, el lugar que ocupa el último coeficiente $\neq 0$. No obstante, debe observarse que con esta definición no hemos cubierto el caso $f(X) = 0$ pues no hay coeficientes $\neq 0$. Ciertos autores dicen que el polinomio cero no tiene grado definido y otros (y es lo que haremos) le atribuyen el grado $-\infty$. No afinamos la definición de $-\infty$ pero este símbolo cumple sin duda $-\infty < n$ para $n = 0, 1, 2, \ldots$. De esta manera, $\Bbbk_n[X]$ contiene al polinomio 0, que es el cero de $\Bbbk[X]$.

Problema 64 *Verificar las afirmaciones que se consideren no evidentes, en especial que para I conjunto infinito, el subespacio de las funciones f que se anulan en todos los $x \in I$ excepto en un número finito de elementos es de verdad un subespacio de $\mathcal{F}(I, \Bbbk)$.*

Problema 65 *Consideramos $\mathbb{E} = \mathbb{R}[X]$ vistos los polinomios como funciones. Estudiar si los subconjuntos $\mathbb{F}_i$ que damos son o no son subespacios vectoriales*

$$\mathbb{F}_1 : f(1) = 0, \qquad \mathbb{F}_2 : f(1) = 1, \qquad \mathbb{F}_3 : f(1) = f(2) = 0, \qquad \mathbb{F}_4 : f'(1) = 0, \qquad \mathbb{F}_5 : f'(1) = 1.$$

Problema 66 *Seguimos con $\mathbb{E} = \mathbb{R}[X]$ y preguntamos si las condiciones (a)"f está acotado" (las cotas pueden ser distintas); (b) "f toma infinitas veces el valor 0"; (c) "f toma infinitas veces el valor 1 y $f(0) = 0$"; (d) "f es divisible por $X^2 + X + 1$", dan o no dan subespacios de $\mathbb{E} = \mathbb{R}[X]$.*

El último resultado es muy rutinario, pero tiene utilidad. Pedimos al lector que vaya a la sección *Espacios dados en paramétricas* y defina para sucesiones $(a_1, \ldots, a_n)$ de vectores de $\mathbb{E}$ las tres operaciones elementales Φ y el concepto de sucesiones equivalentes. Ahora no se pide que $\mathbb{E}$ tenga dimensión finita ni que los a_j sean vectores fila o columna. Se tiene exactamente el mismo teorema que el teorema 29 pero muy generalizado.

Teorema 32 *Si $(a_1, \ldots, a_n)$ y $(b_1, \ldots, b_n)$ son sucesiones de vectores equivalentes en $\mathbb{E}$, entonces generan el mismo subespacio de $\mathbb{E}$.*

2.4. Bases

En toda la sección $\mathbb{E}$ será un espacio vectorial sobre el cuerpo $\Bbbk$. En los ejemplos, si no se precisa, $\Bbbk = \mathbb{R}$. Aquí vamos a tratar aquellos espacios vectoriales en que se pueden asignar de modo unívoco coordenadas a los vectores. Son los llamados espacios de **dimensión finita**. Antes de la definición recordamos otras realizadas para el espacio estándar y que son formalmente idénticas en $\mathbb{E}$ arbitrario. Para una sucesión $(a_1, \ldots, a_n)$ de vectores de $\mathbb{E}$ definimos $\lg(a_1, \ldots, a_n)$ como el conjunto (de hecho es un subespacio vectorial) de todas las combinaciones de $(a_1, \ldots, a_n)$. Si $\lg(a_1, \ldots, a_n) = \mathbb{F}$ se dice que $(a_1, \ldots, a_n)$ **genera** $\mathbb{F}$ y que $(a_1, \ldots, a_n)$ es **sucesión generadora d**e $\mathbb{F}$. Una sucesión generadora que es además independiente se llama **base**. Puede suceder que un espacio vectorial arbitrario no admita bases, pero cuando sí sucede se dice que $\mathbb{E}$ es de **dimensión finita.** Al haber independencia, en $(a_1, \ldots, a_n)$ no hay repeticiones, luego $\{a_1, \ldots, a_n\}$ tiene n elementos y por eso es admisible decir que un espacio de dimensión finita tiene una base con un número finito de elementos. Pronto veremos que, tal como sucede con los subespacios de $\Bbbk^n$, todas las bases tienen el mismo número de elementos y este número se llama la **dimensión** del espacio.

Antes de entrar en los espacios de dimensión finita hay que comentar qué pasa con los demás. No hemos definido ni vamos a definir "bases infinitas" (posible, pero técnicamente complicado). Por eso resulta extraño que se diga que, por definición, un **espacio de dimensión infinita** es el que no es de dimensión finita; o sea, el que no tiene bases (finitas). Pronto se verá que en estos espacios para cualquier sucesión $(a_1, \ldots, a_n)$ hay algún elemento que no puede ponerse como combinación lineal de ella y que para cada m, por grande que sea, se puede conseguir una sucesión $(a_1, \ldots, a_m)$ independiente. Y esto es lo que hace que estos espacios sean diferentes. Tienen sentido en ellos conceptos como combinación, dependencia e independencia lineales, así como subespacio (que puede ser de dimensión finita o no). El obstáculo está en que no se pueden identificar con $\Bbbk^n$ para algún $n \in \mathbb{N}$. Sin embargo un espacio $\mathbb{E}$ de dimensión infinita puede tener subespacios "interesantes" de dimensión finita. El espacio $\Bbbk[X]$ de los polinomios es el ejemplo más claro de espacio de dimensión infinita porque si $(P_1(X), \ldots, P_n(X))$ fuese base de $\Bbbk[X]$ y m el máximo grado de los $P_j(X)$, un polinomio de grado $m+1$ no se podría poner como combinación lineal de $(P_1(X), \ldots, P_n(X))$. En general, los espacios de funciones son de dimensión infinita, constituidos por funciones continuas, diferenciables (del grado de diferenciabilidad que sea), acotadas, periódicas, etc.. Los tres teoremas que siguen tienen idéntica demostración y generalizan los teoremas 23, 24 y 25 probados para subespacios del espacio estándar.

Teorema 33 *Para que la correspondencia entre sucesiones de parámetros $\left(\lambda^1, \ldots, \lambda^n\right)$ y elementos de $\mathbb{F} = \lg(a_1, \ldots, a_n)$ sea biunívoca (o biyectiva) es necesario y suficiente que la sucesión $(a_1, \ldots, a_n)$ de vectores sea independiente.*

Teorema 34 *Si* $\mathbb{E}$ *está generado por* n *vectores,* $\mathbb{E} = \lg(a_1, \ldots, a_n)$, *y* $(b_1, \ldots, b_k)$ *es una sucesión independiente de vectores de* $\mathbb{E}$, *entonces* $k \leq n$.

Teorema 35 *Si* $\mathbb{E}$ *tiene bases* $(a_1, \ldots, a_n)$ *y* $(b_1, \ldots, p_n)$, *entonces* $n = p$.

En efecto, en las demostraciones no se usa en absoluto que los vectores tengan forma $\left(x^1, \ldots, x^n\right)^\top$ típica del espacio estándar. Aparentemente las coordenadas surgen en la demostración del teorema 24, pero es a través de un sistema auxiliar que se construye con vectores que pueden ser de tipo arbitrario. El número de elementos común a todas las bases de $\mathbb{E}$ es la **dimensión de** $\mathbb{E}$, escrita $\dim(\mathbb{E})$. Si no hay bases se dice que $\mathbb{E}$ tiene **dimensión infinita**. Para una sucesión $(a_1, \ldots, a_n)$ de $\mathbb{E}$, los conceptos de **longitud**, **sucesión generadora mínima** y **sucesión independiente máxima** son idénticos a los definidos en el espacio estándar. También vale la demostración de este teorema.

Teorema 36 *Sean* $(a_1, \ldots, a_p)$ *y* $(b_1, \ldots, b_q)$ *respectivamente una sucesión generadora mínima o una sucesión independiente máxima del espacio* $\mathbb{E}$. *Entonces son bases de* $\mathbb{E}$.

En $\mathbb{E}$ arbitrario puede no haber sucesiones generadoras. En el momento en que las hubiera se elegiría entre todas una de longitud mínima y se tendría una base. En $\mathbb{E}$ arbitrario siempre hay sucesiones independientes (por ejemplo (v) para $v \neq 0$) pero puede suceder que las haya ilimitadamente largas así que no existen de longitud máxima. Teniendo presente el teorema 36 podemos deducir que un espacio $\mathbb{E}$ con sucesiones generadoras o para el que exista un $k \in \mathbb{N}$ tal que toda sucesión independiente tenga longitud $\leq k$, ha de ser de dimensión finita. De modo equivalente, si $\mathbb{E}$ tiene dimensión infinita, no hay sucesiones generadoras y las independientes son tan largas como se quiera.

El teorema 27 no es válido pero sí una versión modificada de enunciado y demostración.

Teorema 37 *Todo subespacio* $\mathbb{F} \neq 0$ *de otro espacio* $\mathbb{E}$ *de dimensión finita* n *tiene una base y* $\dim(\mathbb{F}) \leq n$. *Más concretamente, si* $(b_1, \ldots, b_k)$ *es una sucesión independiente de vectores de* $\mathbb{F}$ *se pueden añadir* $b_{k+1}, \ldots, b_m \in \mathbb{F}$ *de modo que* $(b_1, \ldots, b_k, \ldots, b_m)$ *sea base de* $\mathbb{F}$.

Una consecuencia de este teorema es que si $\mathbb{E}$ tiene un subespacio $\mathbb{F}$ de dimensión infinita se sigue que $\mathbb{E}$ es también de dimensión infinita. Finalmente el **teorema de Steinitz** es casi igual.

Problema 67 *Sean* $(a_1, \ldots, a_p)$ *y* $(b_1, \ldots, b_q)$ *respectivamente una sucesión generadora y una sucesión independiente de* $\mathbb{E}$ *(por tanto de dimensión finita). Renumerando* $(a_1, \ldots, a_p)$ *se pueden elegir* $a_1, \ldots, a_h$ *de modo que* $(b_1, \ldots, b_q, a_1, \ldots, a_h)$ *sea base de* $\mathbb{F}$. *Indicación: Considerar sucesiones* $(b_1, \ldots, b_q, a_\bullet, \ldots, a_\bullet) \ldots$ *¿cómo?.*

Se puede hacer una síntesis de las propiedades principales, algunas aparecen en la demostración.

1. Una sucesión de vectores independientes no puede ser estrictamente más larga que una de generadores.
2. Una sucesión generadora mínima o una sucesión independiente máxima es una base (¡si existen!)
3. Todas las bases de un espacio de dimensión finita tienen el mismo número de elementos.
4. Si $\mathbb{E}$ es de dimensión finita tienen una base obtenible alargando cualquier sucesión suya de vectores independientes. Esto se puede conseguir incluso eligiendo los nuevos elementos en una sucesión generadora de $\mathbb{F}$.

Problema 68 *Sea* $\mathbb{E}$ *el espacio de las sucesiones* $x = (x_1, \ldots, x_n, \ldots)$ *que es como decir las funciones de* $\mathbb{N}$ *en* $\Bbbk$. *Probar que es un espacio de dimensión infinita.*

Problema 69 *Dar, aparte de* $\mathbb{R}[X]$, *visto como espacio de funciones, tres ejemplos más de espacios de dimensión infinita.*

Veamos ejemplos de espacios de dimensión finita, que son con los que vamos principalmente a trabajar. Acabamos de ver que si $\mathbb{E}$ es uno de estos ejemplos, cualquier subespacio suyo $\mathbb{F}$ tiene también dimensión finita. Vamos por la vía más directa, que es dar ejemplos de sucesiones $(a_1, \ldots, a_m)$ de forma que $\mathbb{E} = \lg(a_1, \ldots, a_m)$.

1. Ya sabemos desde la primera sección que $\mathbb{E} = \Bbbk^m$ es de dimensión finita y que tiene una base $\mathcal{E} = (e_1, \ldots, e_m)$ que es la **base estándar**, siendo $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ con solo un 1 en el lugar j. Esta base es muy cómoda porque si $x = \left(x^1, \ldots, x^m\right)$, las coordenadas de x en $\mathcal{E}$ son justamente $x^1, \ldots, x^m$. pero deja de ser cierto si tomamos otra base.

2. Supongamos $\mathbb{E} = \Bbbk^{m\times n}$. Hay una generalización de lo anterior. Definamos $\mathbf{e}_q^p$ como la matriz de $\Bbbk^{m\times n}$ con todo ceros excepto 1 en fila p columna q. Por ejemplo, en $\Bbbk^{3\times 2}$ hay seis matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

   Queremos definir con las matrices $\mathbf{e}_q^p$ la **base estándar**, pero queda por decir en qué orden van las $\mathbf{e}_q^p$. No hay un convenio universal. Nosotros adoptaremos el más natural que supone ir desplazando los 1 de izquierda a derecha y de arriba abajo, como al leer un papel. La **base estándar** de $\Bbbk^{m\times n}$ se define

$$\mathcal{E} = \left(\mathbf{e}_1^1, \ldots, \mathbf{e}_n^1, \mathbf{e}_1^2, \ldots, \mathbf{e}_n^2, \ldots, \mathbf{e}_1^{m-1}, \ldots, \mathbf{e}_n^{m-1}, \mathbf{e}_1^m, \ldots, \mathbf{e}_n^m\right),$$

   y es base pues

$$a = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} = a_1^1\mathbf{e}_1^1 + \ldots + a_n^1\mathbf{e}_n^1 + \ldots\ldots + a_1^m\mathbf{e}_1^m + \ldots + a_n^m\mathbf{e}_n^m = \sum_{\substack{1\le i\le m \\ 1\le j\le n}} a_j^i\mathbf{e}_j^i. \quad (2.4)$$

   Un poco más arriba está la base estándar de $\Bbbk^{3\times 2}$. Evidentemente $\Bbbk^m$ tiene dimensión m y $\Bbbk^{m\times n}$ tiene dimensión $m \cdot n$.

3. Un último ejemplo por el momento de espacio de dimensión finita es $\Bbbk_n[X]$, los polinomios con coeficientes en $\Bbbk$ de grado $\le n$. La **base estándar** de este espacio es $\mathcal{E} = \left(1, X, X^2, \ldots, X^{n-1}, X^n\right)$ y es base pues

$$P(X) = a_0 + a_1X + a_2X^2 + \ldots + a_{n-1}X^{n-1} + a_nX^n$$

   luego los coeficientes del polinomio son (numerados desde cero) las coordenadas en $\mathcal{E}$. Claramente $\dim(\Bbbk_n[X]) = n + 1$ (y no n como tenemos tendencia a decir).

4. Si en $\mathbb{E}$ de dimensión finita o infinita nos dan una sucesión $(a_1, \ldots, a_n)$, sabemos, *por la propia definición,* que $\mathbb{F} = \lg(a_1, \ldots, a_n)$ tiene dimensión finita, aunque no sabemos si $\dim(\mathbb{F}) = n$ y solo se puede decir que $\dim(\mathbb{F}) \le n$.

5. La dimensión del producto cartesiano de espacios es la suma (¡y no el producto!) de las dimensiones de los factores. Si, para simplificar, consideramos tan solo $\mathbb{E} \times \mathbb{F}$ con bases $\mathcal{U}$ y $\mathcal{V}$, la base

$$((u_1, 0), \ldots, (u_n, 0), (0, v_1), \ldots, (0, v_p))$$

   es base de $\mathbb{E} \times \mathbb{F}$.

Problema 70 *Sea $\mathbb{F} = \lg(u_1, \ldots, u_n)$ un subespacio de dimensión n de un subespacio $\mathbb{E}$ (no importa la dimensión finita o infinita de $\mathbb{E}$). Probar que si $v \notin \mathbb{F}$ entonces* $\dim \lg(u_1, \ldots, u_n, v) = n + 1$.

Los espacios $\mathbb{E}$ de dimensión finita n "son esencialmente" $\Bbbk^n$. Lo explicamos. Sea $\mathcal{U} = (u_1, \ldots, u_n)$ una de sus bases y asignemos a cada $x = \sum_{i=1}^n x^iu_i$ la matriz $\mathrm{mat}^{\mathcal{U}}(x) = \left(x^1, \ldots, x^n\right)^\top \in \Bbbk^n$ formada por sus coordenadas en esa base. Tememos una función

$$\Phi : \mathbb{E} \to \Bbbk^n, \quad \Phi(x) = \mathrm{mat}^{\mathcal{U}}(x) = \left(x^1, \ldots, x^n\right)^\top.$$

La función Φ es lineal pues, por ejemplo, $\mathrm{mat}^u(\lambda x) = \lambda\,\mathrm{mat}^{\mathcal{U}}(x)$ y es biyectiva. Es un caso particular de **isomorfismo**. En general, $\Phi : \mathbb{E} \to \mathbb{F}$ es un **isomorfismo** (de espacios vectoriales) si es lineal y biyectiva. Decir que $\mathbb{E}$ y $\mathbb{F}$ son isomorfos es la manera rigurosa de decir que, salvo la designación de sus elementos, son en esencia iguales. En efecto, cualquier propiedad que tengan los vectores de $\mathbb{E}$ como por ejemplo, tener una sucesión generadora $(a_1, \ldots, a_n)$, se traslada inmediatamente a $\mathbb{F}$ porque $(\Phi(a_1), \ldots, \phi(a_n))$ será sucesión generadora de $\mathbb{F}$. Es obvio porque si tenemos $y \in \mathbb{F}$ y queremos ponerlo

como $y = \lambda^1\Phi(a_1) + \ldots + \lambda_n\Phi(a_n)$ para ciertos λ, tomamos el único x tal que $\Phi(x) = y$, escribimos $x = \lambda^1 a_1 + \ldots + \lambda^n a_n$ y, tras aplicar Φ, se obtiene $y = \Phi(x) = \lambda^1\Phi(a_1) + \ldots + \lambda_n\Phi(a_n)$ por la linealidad de Φ. Volviendo a nuestro caso particular donde $\Phi(x) = \mathrm{mat}^{\mathcal{U}}(x)$, vemos que $\mathbb{E}$ y $\Bbbk^n$ son "identificables" (eso sí, después de elegir una base $\mathcal{U}$). Por tanto la regla para calcular es identificar (más precisamente, usar el isomorfismo que da $\mathcal{U}$) cada ente de $\mathbb{E}$ con su homólogo de $\Bbbk^n$, trabajar en $\Bbbk^n$, y volver a $\mathbb{E}$ con la conclusión obtenida.

Problema 71 *En $\mathbb{E} = \mathbb{R}_2[X]$ se consideran los polinomios $1 + X + X^2$, $1 - X + X^2$ y $1 + X - X^2$. ¿Forman base?* ♦

Solución. Identificamos $\mathbb{E}$ con $\mathbb{R}^3$ por medio de la base estándar $\mathcal{E} = \left(1, X, X^2\right)$. Los tres polinomios dados se identifican con las tres columnas de la matriz

$$a = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Sabemos que los n vectores $a_1, \ldots, a_n$ de $a \in \Bbbk^{n\times n}$ forman base de $\Bbbk^n$ si a es invertible. En nuestro caso, a es invertible, (a_1, a_2, a_3) es base de $\mathbb{R}^3$, y los tres polinomios de donde provienen los a_i son base de $\mathbb{E} = \mathbb{R}_2[X]$. ♦

Problema 72 *Sea $\mathcal{U} = (u_1, \ldots, u_n)$ una base de $\mathbb{E}$. Con una matriz $a \in \Bbbk^{n\times n}$ construyen nuevos vectores $v_j = \sum_{i=1}^m a_j^i u_i$, $1 \le j \le n$. Probar que $(v_1, \ldots, , v_n)$ es base si y solo si a es invertible.*

Problema 73 *Tomamos $\mathbb{E} = \mathbb{C}$, los números complejos y $\Bbbk = \mathbb{R}$. Es un espacio vectorial con suma la suma de números complejos y con producto $\lambda \cdot (a + bi) = \lambda a + \lambda bi$, siendo $a, b, \lambda \in \mathbb{R}$. Definimos la función $\Phi : \mathbb{C} \to \mathbb{R}^2$, $\Phi(a + bi) = (a, b)$. Probar que Φ es un isomorfismo en el que la base*[10] *$(1, i)$ de $\mathbb{C}$ se corresponde con la base estándar. Dado $\sigma = a + bi \neq 0$, estudiar cuando $(\sigma, i\sigma)$, sucesión constituida por σ y su producto por i, es una base.*

Aunque pretendemos que el lector haga el problema por el procedimiento de pasar por $\mathbb{R}^2$, le advertimos que hay procedimientos más eficaces. Por ejemplo, le vamos a desvelar la segunda parte diciéndole que $(\sigma, i\sigma)$ es *siempre* base (supuesto $\sigma \neq 0$). En efecto, si no lo fuera, $i\sigma$ sería combinación de σ, que en este caso significa que existe $\lambda \in \mathbb{R}$ tal que $i\sigma = \lambda\sigma$. Como $\sigma \neq 0$, se puede dividir por σ e $i = \lambda$, que es imposible si $\lambda \in \mathbb{R}$. Advertimos que al trabajar en $\mathbb{C}$ se tiene $\mathbb{R}^2$ como espacio vectorial (lo que se describe en el problema) y mucho más porque hay una multiplicación, conjugación, etc. que simplifican a veces mucho el trabajo, permitiendo además formular muchos aspectos geométricos del plano.

Si en el espacio $\mathbb{E}$ de dimensión finita en donde estamos no nos dan los datos es función de una base concreta, hay que decidir si se intenta resolver el problema eligiendo una, escribiendo todo con ella y pasando a $\Bbbk^m$ resolver ahí el problema; o resolver el problema por la vía directa sin esta transcripción. No hay regla fija y a veces la primera vía, no es la mejor. Supongamos que en $\mathbb{E} = \mathbb{R}_2[X]$, el espacio de los polinomios de grado ≤ 2, nos preguntan si $\mathbb{F} = \{P(X) \mid P(4) = 0\}$ es un subespacio. Aplicando la definición directa, si $P(X)$ y $Q(X)$ están en $\mathbb{F}$, se evalúa $P(X)+Q(X)$ en $X = 4$ y queda $P(4)+Q(4) = 0 + 0 = 0$, luego $P(X) + Q(X) \in \mathbb{F}$ y se ve de modo similar que $\lambda P(X) \in \mathbb{F}$ cuando $P(X) \in \mathbb{F}$. Si se quiere hacer con coordenadas, se toma la base estándar $\mathcal{E} = \left\{1, X, X^2\right\}$ de $\mathbb{R}_2[X]$. El polinomio $P(X) = p_0 + p_1X + p_2X^2$ tiene coordenadas (p_0, p_1, p_2) y la condición $P(X) \in \mathbb{F}$ equivale a $0 = p_0 + 4p_1 + 16p_2 = 0$ que es un sistema homogéneo en $\mathbb{R}^3$ con una sola ecuación. Su espacio de soluciones $\mathbb{S}$ se corresponde con $\mathbb{F}$ y $\mathbb{F}$ es un subespacio vectorial de $\mathbb{E} = \mathbb{R}_2[X]$. Esta segunda parte es más complicada. Tiene por lo menos la ventaja de que si nos hubiesen preguntado la dimensión de $\mathbb{F}$ responderíamos enseguida que como $\dim(\mathbb{S}) = 2$ ha de ser $\dim(\mathbb{F}) = 2$.

Problema 74 *Recordamos que la traza de $\Bbbk^{m\times m}$ en $\Bbbk$ es $\mathrm{tr}(a) = \sum_{i=1}^m a_i^i$, la suma de los términos de la diagonal. En $\mathbb{E} = \Bbbk^{m\times m}$, probar que $\mathbb{F} = \{a \in \Bbbk^{m\times m} \mid \mathrm{tr}(a) = 0\}$ es un subespacio vectorial. ¿Cuál es su dimensión?*

Problema 75 *En $\mathbb{E} = \Bbbk_2[X]$, el espacio de los polinomios de grado ≤ 2 nos dan la sucesión de tres polinomios $\mathcal{P} = \left(1, \alpha + X, \alpha X + X^2\right)$. Dígase si, dependiendo del valor de α, forman una base y, en caso afirmativo, las coordenadas de $P(X) = \alpha^2 + X^2$ en ella.*

[10] Cuidado con la notación. Una base de $\mathbb{C}$ será una sucesión de (dos) elementos de $\mathbb{C}$ y eso es lo que representa $(1, i)$.

Problema 76 *En los espacios* $\mathbb{C}^{m\times n}$ *y* $\Bbbk_n[X]$ *vamos a cambiar las bases estándar por otras sucesiones de vectores*

$$\mathcal{G}=\left(ie_1^1,\ldots,ie_n^1,ie_1^2,\ldots,ie_n^2,\ldots,ie_1^m,\ldots,ie_n^m\right),\qquad \mathcal{G}=\left(1,-X,X^2,\ldots,(-1)^k X^k,\ldots,(-1)^n X^n\right).$$

¿Forman base de los respectivos espacios? Si es así, ¿cómo se modifican las coordenadas?

Hasta aquí hemos visto procedimientos para determinar si una sucesión de vectores en un espacio de dimensión finita es o no independiente. Nos preguntamos qué se puede hacer si el espacio $\mathbb{E}$ es de dimensión infinita, aparte de aplicar directamente la definición. Hay un teorema útil para el espacio $C^\infty(I,\mathbb{R})$ de las funciones infinitamente diferenciables (i.e, se pueden derivar cuantas veces se quiera) de un intervalo I en $\mathbb{R}$. El teorema vale también para cualquier subespacio; por ejemplo, el formado por las funciones polinómicas. Para $f:I\to\mathbb{R}$ representaremos por $f^{(i)}(t)$ a la derivada de orden i en t. Se entiende que $f'(t)=f^{(1)}(t)$ y $f^{(0)}(t)=f(t)$.

Teorema 38 *Sean* $f_1,\ldots,f_n$ *funciones de* $C^\infty(I,\mathbb{R})$. *Si existe un* $t\in I$ *tal que la matriz*

$$a(t)=\begin{pmatrix} f_1^{(0)}(t) & f_2^{(0)}(t) & \cdots & f_n^{(0)}(t)\\ f_1^{(1)}(t) & f_2^{(1)}(t) & \cdots & f_n^{(1)}(t)\\ \vdots & \vdots & \ddots & \vdots\\ f_1^{(n-1)}(t) & f_2^{(n-1)}(t) & \cdots & f_n^{(n-1)}(t)\end{pmatrix}$$

es invertible, entonces la sucesión $(f_1,\ldots,f_n)$ *es independiente.*

Demostración. Sean $\lambda^1,\ldots,\lambda_n\in\mathbb{R}$ tales que $\lambda^1 f_1+\ldots+\lambda^n f_n=f$ sea la función 0. Derivándola tantas veces como queramos se tendrá la función 0 y, al evaluar en cualquier $t\in I$ se tendrá también $\lambda^1 f_1^{(i)}(t)+\ldots+\lambda^n f_n^{(i)}(t)=f^{(i)}(t)=0$. Resultan n ecuaciones que involucran a las λ^j y que en forma matricial se escriben como $a(t)\left(\lambda^1,\ldots,\lambda^n\right)^\top=0$. Basta que en algún t sea invertible $a(t)$ para obtener $\left(\lambda^1,\ldots,\lambda^n\right)^\top=0$ y la independencia de $(f_1,\ldots,f_n)$.

El determinante de $a(t)$ es el **Wronskiano**. *La recíproca del teorema es falsa* (diría que si la sucesión de funciones es independiente, $a(t)$ debe ser invertible para al menos un t). Si $f(t)=t^3$ y $g(t)=\left|t^3\right|$ se tiene

$$a(t)=\begin{pmatrix} t^3 & \left|t^3\right|\\ 3t^2 & 3|t|\,t\end{pmatrix}$$

y claramente t por la segunda fila da la primera al ser $|t|^2=t^2$. (Quizás el lector vea difícil probar que $g'(t)=3|t|\,t$.) Sin embargo hay independencia, porque si se tuviera $0=\lambda t^3+\mu|t|^3$, para $t=\pm1$ queda $\lambda+\mu=0$ y $-\lambda+\mu=0$, de donde $\lambda=\mu=0$. ♣

Problema 77 *Probar que las funciones* $f,g:(0,\infty)$ *dadas por* $f(t)=t^k$, $k\in\mathbb{N}$ *fijo, y* $g(t)=1/t$, *son independientes.* ♦

Solución. Se tiene con la notación anterior

$$a(t)=\begin{pmatrix} t^k & \frac{1}{t}\\ kt^{k-1} & -\frac{1}{t^2}\end{pmatrix},\quad a(1)=\begin{pmatrix} 1 & 1\\ k & -1\end{pmatrix}.$$

Como $a(1)$ es invertible, el teorema 38 nos da la independencia. ♦

Problema 78 *Probar que las funciones* $\cos t$, $\sin t$, $\cos^2 t$ *y* $\sin^2 t$ *son independientes.*

2.5. El rango de una matriz

Vamos a ver cómo se determinan bases y dimensiones con ciertos algoritmos que permiten ahorrar trabajo. Fijemos $a\in\Bbbk^{m\times n}$ y definamos el **espacio de las columnas** $\mathbb{K}(a)$ y el **espacio de las filas** $\mathbb{F}(a)$ **de la matriz** a. El espacio $\mathbb{K}(a)$ es el subespacio de $\Bbbk^m$ generado por las columnas $(a_1,\ldots,a_n)$ de

a y el espacio $\mathbb{F}(a)$ es el subespacio de $\Bbbk^n$ generado por las filas $\left(a^1,\ldots,a^m\right)$ de a. Dicho con símbolos, son los subespacios $\lg(a_1,\ldots,a_n)\subset\Bbbk^m$ y $\lg\left(a^1,\ldots,a^m\right)\subset\Bbbk^n$. Por ejemplo,

$$\mathbb{K}(a)=\lg\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}2\\-1\end{pmatrix}\right)\ \text{ y }\ \mathbb{F}(a)=\lg\left((1,1,2),(0,1,-1)\right)\ \text{ si }\ a=\begin{pmatrix}1&1&2\\0&1&-1\end{pmatrix}.$$

Estos ejemplos son sumamente manejables en lo referente a dimensión y bases, y permiten conocer dimensiones y verificar dependencia e independencia lineal tal como se ha definido en los párrafos anteriores. Es obvio, especialmente si $m\neq n$, que el espacio de columnas $\mathbb{K}$ y el espacio de filas $\mathbb{F}$ de la matriz a son diferentes, porque sus vectores dependen de m y n parámetros, pero incluso si $m=n$, estos espacios son diferentes. (Pensar en una matriz 2×2 con columnas $(1,1)^\top$ y $(0,0)^\top$.). La dimensión de $\mathbb{K}$ se llamará el **rango por columnas** de a y la de $\mathbb{F}$ será el **rango por filas** de a y se denotarán por $\operatorname{rg}_C(a)$ y $\operatorname{rg}_F(a)$. Por lo que hemos advertido se puede pensar que $\operatorname{rg}_C(a)\neq\operatorname{rg}_F(a)$ pero...

Teorema 39 *El rango por filas es igual al rango por columnas*

Daremos un poco más adelante la demostración manteniendo de momento la distinción entre $\operatorname{rg}_C(a)$ y $\operatorname{rg}_F(a)$, pero cuando probemos el teorema 39 escribiremos $\operatorname{rg}_C(a)=\operatorname{rg}_F(a)=\operatorname{rg}(a)$ y le llamaremos el **rango de la matriz** a.

La demostración del teorema 39 se basa en una acotación de $\operatorname{rg}_C(a)$ y $\operatorname{rg}_F(a)$ (admitimos todavía que pueden ser distintos) *cuando a es producto de dos matrices.* Sean pues $a=uv$ con $a\in\Bbbk^{m\times n}$, $u\in\Bbbk^{m\times p}$ y $v\in\Bbbk^{p\times n}$. Tenemos tres fórmulas con índices que expresan lo mismo: que $a=uv$. Son

$$\begin{cases}\Bbbk\ni a^i_j=\sum_{k=1}^p u^i_k v^k_j, & 1\le i\le m,\quad 1\le j\le n\\ \Bbbk^m\ni a_j=\sum_{k=1}^p v^k_j u_k, & 1\le j\le n\\ \Bbbk^n\ni a^i=\sum_{k=1}^p u^i_k v^k, & 1\le i\le m\end{cases}\tag{2.5}$$

Hemos puesto $\Bbbk,\Bbbk^m,\Bbbk^n\ni\bullet$ para indicar que la ecuación $\bullet$ es una ecuación en $\Bbbk$ (ecuación numérica) o es una ecuación en $\Bbbk^m$ o $\Bbbk^n$ (ecuación vectorial, de filas o columnas). La primera ecuación de las tres es sin duda cierta por la fórmula del producto de matrices. Para comprobar la segunda se observa que lo que está a ambos lados de $=$ es un vector de $\Bbbk^m$ porque lo es la columna j de a y las columnas u_k de $u\in\Bbbk^{m\times p}$. Si tenemos en $\Bbbk^m$ una combinación $x=\lambda^1y_1+\ldots+\lambda^ry_r$, la componente i de x es $x^i=\lambda^1y^i_1+\ldots+\lambda^ry^i_r$ siendo y^i_j la coordenada i del vector y_j. Por consiguiente,

$$a^i_j=\sum_{k=1}^p v^k_j u^i_k=\sum_{k=1}^p u^i_k v^k_j$$

precisamente la primera ecuación, que es cierta luego la segunda ecuación es cierta. La demostración de que la tercera ecuación es cierta es análoga. Lo importante es que las ecuaciones segunda y tercera nos dicen que

$$a_j\in\lg(u_1,\ldots,u_p),\ 1\le j\le n,\quad\text{y}\quad a^i\in\lg\left(v^1,\ldots,v^p\right),\ 1\le i\le m,$$

y como consecuencia,

$$\mathbb{K}\subset\lg(u_1,\ldots,u_p),\quad\dim(\mathbb{K})\le p,\quad\mathbb{F}\subset\lg\left(v^1,\ldots,v^p\right),\quad\dim(\mathbb{F})\le p.$$

Hemos probado este teorema que será esencial para el teorema 39.

Teorema 40 *Sea $a=uv$ con $a\in\Bbbk^{m\times n}$, $u\in\Bbbk^{m\times p}$ y $v\in\Bbbk^{p\times n}$. El espacio $\mathbb{K}$ de las columnas de a está contenido en el de las columnas de u y el espacio $\mathbb{F}$ de las filas de a está contenido en el de las filas de v. Como estos espacios de u y v tienen dimensión $\le p$ se deduce que*

$$\operatorname{rg}_C(a)=\dim(\mathbb{K})\le p,\quad\operatorname{rg}_F(a)=\dim(\mathbb{F})\le p.$$

Demostración. (del teorema 39) Todo es cosa de factorizar $a=uv$ de forma adecuada. Probaremos que si $p=\operatorname{rg}_C(a)$ y $q=\operatorname{rg}_F(a)$ debe ser $p\le q$ y $q\le p$. Mostremos que $q\le p$ pues $p\le q$ tiene demostración análoga. Por hipótesis $\dim(\mathbb{K})=p$, y hay una base de $\mathbb{K}$ de longitud p, digamos que es

$(b_1,\dots,b_p)$. Las columnas a_j serán combinaciones de $(b_1,\dots,b_p)$ luego existen coeficientes $h^\bullet_\bullet \in \Bbbk$ de forma que

$$\Bbbk^m \ni a_j = \sum_{k=1}^{p} h^k_j b_k, \quad 1 \le j \le n.$$

Si cada una de estas n ecuaciones se desdobla en ecuaciones escalares obtenemos

$$\Bbbk \ni a^i_j = \sum_{k=1}^{p} h^k_j b^i_k = \sum_{k=1}^{p} b^i_k h^k_j, \quad 1 \le i \le m, \quad 1 \le j \le n,$$

que es decir en coordenadas que $a = bh$ con $b \in \Bbbk^{m\times p}$ y $h \in \Bbbk^{p\times n}$. El teorema 40 nos dice que la dimensión q del espacio de filas de a es $\le p$, que es la primera mitad que íbamos a probar. Que $p < q$ tiene demostración análoga. ♣

Problema 79 *Escribir la demostración de que $p \le q$.*

El rango de una matriz a es por tanto una *dimensión* de un subespacio de $\Bbbk^m$ o $\Bbbk^n$ según nos convenga para el cálculo. Vamos a describir el sistema efectivo de cálculo del rango. Tomemos $a \in \Bbbk^{m\times n}$ y sean $b, c \in \Bbbk^{m\times n}$ las formas escalonada y escalonada reducida de a, digamos que por filas. Aunque b y c intervienen en el razonamiento, será suficiente con conocer b. El teorema 29 o el teorema 32 nos dicen que $\lg(a_1,\dots,a_n) = \lg(b_1,\dots,b_n) = \lg(c_1,\dots,c_n) = \operatorname{rg}(a)$. Sin embargo el rango de c es muy fácil de obtener; de hecho, $\operatorname{rg}(c) = r$ siendo r el número de filas no nulas de c. Esto es casi inmediato porque $\lg(c_1,\dots,c_n) = \lg(c_1,\dots,c_r)$ y, además, $(c_1,\dots,c_r)$ es independiente. En efecto, si $c^1_{p_1} = \dots = c^r_{p_r} = 1$ son los pivotes de c, y $\lambda_1 c^1 + \dots + \lambda_r c^r = s = 0$, las componentes $p_1 < p_2 < \dots < p_r$ de s, que son nulas porque $s = 0$, son precisamente $\lambda_1,\dots,\lambda_r$, lo que prueba la idependencia de $(c_1,\dots,c_r)$ y que $\operatorname{rg}(c) = r$. Hemos llegado a que $\operatorname{rg}(a) = \operatorname{rg}(b) = \operatorname{rg}(c)$ es el número de filas no nulas de c, que es también el de filas no nulas de b. Advertimos que es a veces más rápido transformar a en d en forma reducida. Si se permutan sus filas adecuadamente, d pasa a forma escalonada, luego por lo anterior, el número de filas no nulas de d es también el rango de a.

Teorema 41 *Sea $a \in \Bbbk^{m\times n}$ y b, c y d alguna de sus formas escalonada, escalonada reducida y reducida por filas (respectivamente, columnas). Entonces, el número de filas (respectivamente, columnas) no nulas es el mismo en b, c y d, y es el rango de a.*

Corolario 2 *El espacio de soluciones $\mathbb{S}$ del sistema homogéneo $ax = 0$ tiene dimensión $n - \operatorname{rg}(a)$.*

Demostración. Hemos demostrado en *Espacios dados en implícitas* que $\dim(\mathbb{S})$ es el número de variables libres que aparecen al resolver $ax = 0$ y este número es precisamente $n - r$, siendo r el número de filas no nulas, que es justamente el rango de a. ♣

El siguiente problema da resultados muy fáciles pero de uso frecuentísimo.

Problema 80 *El rango de una matriz (posiblemente no cuadrada) y su traspuesta es el mismo. Para una matriz $a \in \Bbbk^{m\times m}$ cuadrada, tener el rango máximo m equivale a ser invertible.*

El que sea $\operatorname{rg}(a) = \operatorname{rg}\left(a^\top\right)$ implica que si nos gusta, por ejemplo, calcular rangos trabajando por filas podemos elegir entre a o $a^\top$ la que mejor se adapte al cálculo por filas.

Teorema 42 (Rouché-Frobenius) *Dado el sistema $ax = y$ con $a \in \Bbbk^{m\times n}$, es necesario y suficiente para que tenga solución que a y la matriz ampliada $(a \mid y)$ tengan el mismo rango.*

Demostración. El que $ax = y$ tenga solución $x = \left(x^1,\dots,x^n\right)$ equivale a que sea cierta la combinación $y = x^1 a_1 + \dots + x^n a_n$, que es como decir que $y \in \lg(a_1,\dots,a_n)$. Entonces,

$$\lg(a_1,\dots,a_n,y) = \lg(a_1,\dots,a_n), \quad \dim\lg(a_1,\dots,a_n,y) = \dim\lg(a_1,\dots,a_n), \quad \operatorname{rg}(a) = \operatorname{rg}(a \mid y).$$

Si, recíprocamente, y no fuera solución, se tendría $y \notin \lg(a_1,\dots,a_n) = \mathbb{K}$, el espacio de las columnas de a. Vimos en el problema 70 que si v no está en un subespacio $\mathbb{F}$, el subespacio $\mathbb{F}'$ generado por $\mathbb{F}$ y v tiene dimensión $\dim(\mathbb{F}') = \dim(\mathbb{F}) + 1$. Aplicando esto a $\mathbb{F} = \mathbb{K}$ y $v = y$ vemos que $\operatorname{rg}(a \mid y) = \operatorname{rg}(a) + 1 \ne \operatorname{rg}(a)$. ♣

Problema 81 *Estudiese en función de h si los sistemas lineales en $\mathbb{R}$ (se supone $n \geq 2$)*

$$\left\{ \begin{array}{r} x^1 + 2x^2 + 3x^3 + \ldots + nx^n = 1 \\ 2x^1 + 3x^2 + 4x^3 + \ldots + (n+1)\, x^n = h \end{array} \right. , \qquad \left\{ \begin{array}{r} x^1 + x^2 + x^3 + \ldots + x^n = h \\ x^1 + x^2 + x^3 + \ldots + x^{n-1} + hx^n = h \end{array} \right.$$

tienen soluciones.

Le vamos a explicar el lector como puede ponerse muchísimos problemas numéricos. Consideremos la matriz

$$a = \begin{pmatrix} -4 & 1 & 5 & 2 \\ 3 & 0 & -2 & -1 \\ -1 & 1 & 3 & 1 \end{pmatrix}.$$

Se han tomado dos vectores a_2 y a_4 independientes y se ha puesto $a_1 = 2a_2 - 3a_4$ y $a_3 = -a_2 + 2a_4$. Por construcción, $\lg(a_1, a_2, a_3, a_4) = \lg(a_2, a_4)$ y el espacio de las columnas tiene dimensión 2. Nos planteamos el sistema $ax = 0$. Como hemos preparado el problema para que sea $\operatorname{rg}(a) = 2$, al pasar a a forma escalonada b se sabe que tendrá rango 2, luego b tendrá una fila nula y habrá $4 - 2 = 2$ variables libres, por tanto, el espacio de soluciones tendrá dimensión 2. Esto no nos da de por sí una base de este espacio pero nos marca ciertas condiciones que deben verificarse en el proceso de solución. Otra posibilidad es tomar

$$a = \begin{pmatrix} h^2 - h & h & h^2 + h & h \\ -h & 0 & h^2 & h \\ h^2 & h & h & 0 \end{pmatrix}$$

con $h \in \Bbbk$. Las columnas a_2 y a_4 son independientes y $a_1 = ha_2 - a_4$, $a_3 = a_2 - ha_4$. Nuevamente a tiene rango 2 y el sistema $ax = 0$ tiene espacio de soluciones con dimensión 2. Por supuesto los cálculos son laboriosos y el problema puede considerarse poco recomendable. Para simplificarlo se puede suponer $\Bbbk = \mathbb{C}$ y que sea $h = i$ con lo que $h^2 = -1$. Simplemente cuesta calcular el rango de a, pero si además hay un parámetro h y el resultado depende del valor de h, el trabajo es aún mayor. Un ejemplo

$$u = \begin{pmatrix} 1 & 1-h \\ 1 & 1 \end{pmatrix}, \quad v = \begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix}, \quad a = uv = \begin{pmatrix} 1 & -h(h-1) \\ 1 & h \end{pmatrix}$$

Como por el teorema 40 el espacio de las filas de a está contenido en el de las filas de v, se intuye que si $h = 0$ el espacio de las filas de a tendrá dimensión 1 y no puede ser $\operatorname{rg}(a) = 2$. Un problema de poco cálculo es calcular el rango de a en función del valor de h.

Problema 82 *Proponer a un compañero lector o proponerse a uno mismo algún problema numérico de cálculo de rango de una matriz o de un conjunto de vectores como los que acabamos de indicar.*

El siguiente problema exige más destreza.

Problema 83 *Sean $\left(\lambda^1, \ldots, \lambda^m\right)$ y $(\mu_1, \ldots, \mu_m)$ dos sucesiones de números, ambas con algún valor $\neq 0$. Con las sucesiones se construye la matriz $a \in \mathbb{R}^{m \times m}$ dada por $a^i_j = \lambda^i \mu_j$. ¿Cuál es su rango?*

El siguiente teorema es muy útil y generaliza el problema 72.

Teorema 43 *Sea $\mathcal{U} = (u_1, \ldots, u_m)$ una base de $\mathbb{E}$ de dimensión m.*[11] *Con una matriz $a \in \Bbbk^{m \times n}$ (quizás no cuadrada) se construyen nuevos vectores*

$$v_j = \sum_{i=1}^{m} a^i_j u_i, \qquad 1 \leq j \leq n.$$

Entonces la dimensión del espacio $\lg(v_1, \ldots, v_n)$ es el rango de a. En el caso particular $m = n$, (entonces a es cuadrada) se obtiene que $(v_1, \ldots, v_m)$ es base si a es invertible.

Demostración. Consideramos el isomorfismo $\Phi : \mathbb{E} \to \Bbbk^m$ que lleva la base $\mathcal{U}$ en la estándar $\mathcal{E}$ de $\Bbbk^m$. Es obvio que $\Phi(v_j) = \left(a^1_j, \ldots, a^m_j\right)$, $1 \leq j \leq n$. Como Φ es un isomorfismo,

$$\dim \lg(v_1, \ldots, v_n) = \dim \lg(\Phi(v_1), \ldots, \Phi(v_n)) = \dim \lg(a_1, \ldots, a_n)$$

[11] El que sea $m = \dim(\mathbb{E})$ y no n es para que a sea $m \times n$.

y el teorema es ya inmediato. ♣

Puede ser útil observar que las columnas de a están formadas por las coordenadas de los v en $\mathcal{U}$. En todo caso, la manera práctica de calcular con este problema es transformar las ecuaciones con sumatorios a una "forma matricial"

$$(v_1,\ldots,v_n)=(u_1,\ldots,u_m)\begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix}.$$

Hemos puesto comillas en "forma matricial" porque $(v_1,\ldots,v_n)$ y $(u_1,\ldots,u_m)$ no son sucesiones de números sino de vectores, pero se puede hacer la multiplicación formal sin problema. Escritas así las cosas, el rango de a es $\dim \lg (v_1,\ldots,v_n)$. Si nos dicen por ejemplo que

$$\begin{cases} v_1=u_1+2u_2+3u_3 \\ v_2=4u_1+5u_2+6u_3 \\ v_1=7u_1+8u_2+9u_3 \end{cases}, \quad (v_1,v_2,v_3)=(u_1,u_2,u_3)\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$$

y como la matriz tiene rango 2, $\dim \lg (v_1,v_2,v_3)=2$ y (v_1,v_2,v_3) no es una base. La matriz a hay que escribirla como se ha hecho para que sea cierto $(v_1,v_2,v_3)=(u_1,u_2,u_3)\,a$ pero si se hubiera puesto la traspuesta (borrando mentalmente u y v en las fórmulas con sumatorios) no habría problema porque solo se trata de calcular un rango.

Problema 84 *En $\mathbb{E}$ de dimensión m nos dan una base $\mathcal{U}$. Se pide calcular la dimensión de $\mathbb{F}$ el subespacio generado por*

$$v_1=u_1,\quad v_2=u_1+u_2,\quad \ldots\quad v_k=u_1+u_2+\ldots+u_k,\quad \ldots\quad v_m=u_1+u_2+\ldots+u_m.$$

Ídem con $m=4$ para los vectores

$$v_1=u_2+u_3+u_4,\quad v_2=u_2,\quad v_3=u_3,\quad v_4=u_1+u_2+u_3.$$

¿Depende la respuesta de que $\Bbbk$ sea $\mathbb{R}$ o que sea un $\mathbb{Z}_p$?

Problema 85 *En $\mathbb{E}$ tridimensional con $\Bbbk=\mathbb{Z}_p$ nos dan una base $\mathcal{U}$ con la que construimos $v_1=u_2+u_3$, $v_2=u_1+u_3$, $v_3=u_1+u_2$. ¿Cuál es la dimensión de $\mathbb{F}=\lg(v_1,v_2,v_3)$. En general, si en $\mathbb{E}$ de dimensión $p+1$ y $\Bbbk=\mathbb{Z}_p$ se toma una base $\mathcal{U}$ y $p+1$ vectores*

$$v_0=u_1+\ldots+u_p,\ v_1=u_0+u_2+\ldots+u_p,\ v_2=u_0+u_1+u_3+\ldots+u_p,\ \ldots\ ,\ v_p=u_0+\ldots+u_{p-1},$$

¿Cuál es la dimensión de $\mathbb{F}=\lg(v_0,v_1,\ldots,v_p)$?

2.6. Intersección y suma de subespacios vectoriales

Aparte de construir espacios vectoriales con ecuaciones implícitas o paramétricas, hay otros procedimientos. Hay cierta dificultad en el sentido de que muchos resultados son del tipo $\mathbb{G}_1\subset\mathbb{G}_2$ o $\mathbb{G}_1=\mathbb{G}_2$, siendo $\mathbb{G}_1$ y $\mathbb{G}_2$ subespacios obtenidos a partir de $\mathbb{F}_1,\ldots,\mathbb{F}_p$ por determinados procedimientos. Si se trata de probar que $\mathbb{G}_1\subset\mathbb{G}_2$ se está hablando de un contenido de conjuntos y por tanto hay que ver que si $x\in\mathbb{G}_1$, entonces $x\in\mathbb{G}_2$. Este tipo de argumentos resultan nuevos a muchos lectores.

2.6.1. Intersección de subespacios vectoriales

Si $\mathbb{F}_1,\ldots,\mathbb{F}_p$ son subespacios de $\mathbb{E}$, la **intersección de subespacios** $\mathbb{F}_1,\ldots,\mathbb{F}_p$ es su intersección conjuntista; o sea,

$$\mathbb{F}=\mathbb{F}_1\cap\ldots\cap\mathbb{F}_p=\bigcap_{i=1}^{p}\mathbb{F}_i=\{x\in\mathbb{E}\mid\forall i=1,\ldots,p,\ x\in\mathbb{F}_i\}.$$

Dicho con palabras, $x\in\mathbb{F}$ si y solo si está en todos los subespacios $\mathbb{F}_1,\ldots,\mathbb{F}_p$. Le pedimos al lector que haga la sencillísima comprobación de que $\mathbb{F}_1\cap\ldots\cap\mathbb{F}_p$ es otro subespacio de $\mathbb{E}$. El concepto se

generaliza sin problemas si tenemos un conjunto finito o infinito $\mathcal{F}$ de subespacios de $\mathbb{E}$ y se define $\bigcap \mathcal{F}$ como el conjunto de los $x \in \mathbb{E}$ que están en todos los $\mathbb{F} \in \mathcal{F}$ (recordar que los elementos de $\mathcal{F}$ son subespacios). Si $\mathcal{F} = \{\mathbb{F}_1, \ldots, \mathbb{F}_p\}$ se cumple que $\bigcap \mathcal{F} = \mathbb{F}_1 \cap \ldots \cap \mathbb{F}_p$. El querer generalizar el concepto a un conjunto general de subespacios $\mathcal{F}$ (se supone que $\mathcal{F}$ tiene al menos un subespacio) se hace para que sea válida la siguiente construcción. Tomamos un subconjunto X de $\mathbb{E}$ y $\mathcal{F}$ tendrá como elementos todos los subespacios que contienen a X. Desde luego, $\mathbb{E} \in \mathcal{F}$. Se define entonces $\bigcap \mathcal{F}$, que es, con palabras, la intersección de todos los subespacios que contienen a X. Vamos a denotar este subespacio por $\lg(X)$ y se llamará el **subespacio generado por** X. Claramente, $\lg(X)$ es el subespacio más pequeño de todos los que contienen a X, refiriéndose "pequeño" a la relación de contenido. En efecto, si $\mathbb{F} \supset X$ se tiene que $\mathbb{F} \in \mathcal{F}$ y $\mathbb{F} \supset \bigcap \mathcal{F} = \lg(X)$.

Si $X = \{a_1, \ldots, a_n\}$ se pone $\lg(X) = \lg\{a_1, \ldots, a_n\}$, que es una notación próxima a $\lg(a_1, \ldots, a_n)$, el conjunto de las combinaciones lineales de $(a_1, \ldots, a_n)$. En principio, son subespacios distintos, pero en realidad no es así.

Problema 86 *Para un conjunto $X = \{a_1, \ldots, a_n\}$ se tiene que $\lg(\{a_1, \ldots, a_n\}) = \lg(a_1, \ldots, a_n)$.* ♦

Solución. Queda más claro si se prescinde de símbolos y se pide probar que el subespacio $\mathbb{S}_1$ más pequeño de los que contienen a $\{a_1, \ldots, a_n\}$ es el subespacio $\mathbb{S}_2$ de las combinaciones lineales de $(a_1, \ldots, a_n)$. Como S_2 es un subespacio vectorial conteniendo a $\{a_1, \ldots, a_n\}$, al ser $\mathbb{S}_1$ intersección de todos estos subespacios, debe suceder que $\mathbb{S}_1 \subset \mathbb{S}_2$. Sea por otra parte un subespacio $\mathbb{F} \supset \{a_1, \ldots, a_n\}$. Si $x = \lambda^1 a_1 + \ldots + \lambda^n a \in \mathbb{S}_2$ se tiene $x \in \mathbb{F}$ y, al estar x en todos los $\mathbb{F} \supset \{a_1, \ldots, a_n\}$ resulta $x \in \mathbb{S}_1$. ♦

Si $\mathbb{E} = \Bbbk^n$ y los subespacios vienen dados en implícitas $\Bbbk^n$, el sistema que da $\mathbb{F} = \mathbb{F}_1 \cap \ldots \cap \mathbb{F}_p$ tiene una sencilla descripción en implícitas, pues es el sistema que resulta al unir todas las ecuaciones en un sistema mayor. Se entiende muy bien con ejemplos. Si en $\mathbb{R}^3$ tenemos rectas $\mathbb{F}$ y $\mathbb{G}$,

$$\mathbb{F}: \left\{ \begin{array}{c} 2x+3y=0 \\ x-y=0 \end{array} \right. , \qquad \mathbb{G}: \left\{ \begin{array}{c} x+4y=0 \\ 2x-y=0 \end{array} \right. , \qquad \mathbb{F} \cap \mathbb{G}: \left\{ \begin{array}{c} 2x+3y=0 \\ x-y=0 \\ x+4y=0 \\ 2x-y=0 \end{array} \right. .$$

El lector sospechará con acierto que 4 ecuaciones para $\mathbb{F} \cap \mathbb{G}$ son demasiadas y que alguna sobra. Tiene razón pero para abordar este problema con generalidad hay que conocer la fórmula de Grassmann que pronto explicaremos.

Problema 87 *Consideramos los subconjuntos de matrices (a) la intersección de las matrices triangulares superiores y triangulares inferiores; (b) la intersección de las matrices simétricas y antisimétricas; (c) La intersección de las matrices en $\mathbb{R}^{2\times 2}$ que envían $(1,0)^\top$ a $(0,0)^\top$ y las que que envían $(0,1)^\top$ a $(0,0)^\top$; (d) Ídem para las que envían respectivamente $(1,0)^\top$ a $(0,1)^\top$ y $(0,1)^\top$ a $(1,0)^\top$. Estudiar si son o no son subespacios vectoriales.*

2.6.2. Suma de subespacios vectoriales. Sumas directas

Tras considerar la intersección parece natural considerar la *unión* de subespacios. Esto puede hacerse pues los subespacios son conjuntos, pero el concepto no tiene interés, como indica el siguiente problema. (La solución ocupa muy poco, pero posiblemente le llevará algún tiempo al lector resolverlo) Basta de momento que tengamos claro que la unión de subespacios no tiene que ser un subespacio y, a la vista del problema "casi nunca" lo es.

Problema 88 *Si $\mathbb{F}$ y $\mathbb{G}$ son subespacios de $\mathbb{E}$ se tiene que a $\mathbb{F} \cup \mathbb{G}$ es un subespacio solo si a $\mathbb{F} \subset \mathbb{G}$ o $\mathbb{G} \subset \mathbb{F}$.*

El concepto que interesa no es la unión, sino la **suma de subespacios**. Si $\mathbb{F}$ y $\mathbb{G}$ son subespacios de $\mathbb{E}$, su suma es

$$\mathbb{F} + \mathbb{G} = \{x = y + z \mid y \in \mathbb{F},\ z \in \mathbb{G}\};$$

o sea, el conjunto de vectores que resulta al sumar de todas las formas posibles uno de $\mathbb{F}$ y otro de $\mathbb{G}$. En general, para subespacios $\mathbb{F}_1, \mathbb{F}_2, \ldots, \mathbb{F}_k$ se define su suma por

$$\mathbb{F}_1 + \mathbb{F}_2 + \cdots + \mathbb{F}_k = \sum_{i=1}^{k} \mathbb{F}_i = \{x = y_1 + \cdots + y_k \mid y_i \in \mathbb{F}_i,\ \forall i = 1, \ldots, k\}.$$

Con palabras: los vectores x de la suma son los expresables como combinación lineal de los y_i con cada y_i en $\mathbb{F}_i$. Pedimos al lector que haga la sencilla comprobación de que $\mathbb{F}_1+\mathbb{F}_2+\cdots+\mathbb{F}_k$ es un subespacio vectorial. Lo que sucede es que tiene una descripción alternativa como el subespacio generado por el *conjunto* $X=\mathbb{F}_1\cup\cdots\cup\mathbb{F}_k$.

Problema 89 *Sean* $\mathbb{F}_1,\mathbb{F}_2,\ldots,\mathbb{F}_k$ *subespacios de* $\mathbb{E}$ *y* $X=\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k$. *Probar que* $\lg(\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k)=\sum_{i=1}^k\mathbb{F}_i$. ♦

Solución. Sirven las mismas ideas que para el problema 86. Hay que probar que $\mathbb{S}_1=\lg(\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k)$ el subespacio más pequeño que contiene a $\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k$ y $\mathbb{S}_2$ el de las combinaciones lineales $x=\lambda^1a_1+\ldots+\lambda^ka_k$ con $a_i\in\mathbb{F}_i$, se tiene que $\mathbb{S}_1=\mathbb{S}_2$. Dado $a_i\in\mathbb{F}_i$ se tiene $a_i=0\cdot a_1+0\cdot a_2+\ldots+1\cdot a_i+\ldots+0\cdot a_k$, luego $\mathbb{F}_i\subset\mathbb{S}_2$ para todo i y $\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k\subset\mathbb{S}_2$. Al ser $\mathbb{S}_1$ el subespacio más pequeño que contiene a $\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k$, resulta $\mathbb{S}_1\subset\mathbb{S}_2$.

Recíprocamente, dada $x=\lambda^1a_1+\ldots+\lambda^ka_k\in\mathbb{S}_2$ y $\mathbb{F}\supset\mathbb{F}_1\cup\ldots\cup\mathbb{F}_k$ vemos que $x\in\mathbb{F}$, y al ser $\mathbb{S}_1$ la intersección de todos esto $\mathbb{F}$, es $x\in\mathbb{S}_1$ y $\mathbb{S}_2\subset\mathbb{S}_1$. ♦

Con este problema podemos visualizar las sumas de subespacios, al menos en $\mathbb{R}^2$ y $\mathbb{R}^3$. Los subespacios son el origen, el total ($\mathbb{R}^2$ o $\mathbb{R}^3$), las rectas a través del origen (sean del plano o del espacio) y los planos a través del origen (solo en $\mathbb{R}^3$). No hay duda para visualizar intersecciones; por ejemplo, dos rectas se cortan en el origen o son iguales. La suma es el mínimo subespacio que contiene a los sumandos. En el plano casi no hay nada que ver porque dos rectas distintas dan como suma todo el plano. En $\mathbb{R}^3$ dos rectas distintas dan el plano que las contiene y tres rectas no coplanarias (no en un mismo plano) dan todo $\mathbb{R}^3$. Una recta que corte transversalmente a un plano da el total $\mathbb{R}^3$ como suma de esa recta y plano. Hay que tomar nota de esto pues ayuda a la intuición.

Cuando $\mathbb{F}$ y $\mathbb{G}$ vienen dados por ecuaciones podemos preguntarnos como serán las ecuaciones de $\mathbb{F}+\mathbb{G}$. Si se trata de hacer esto para $\mathbb{F}\cap\mathbb{G}$, lo que va bien es tener $\mathbb{F}$ y $\mathbb{G}$ en implícitas porque $\mathbb{F}\cap\mathbb{G}$ viene dado por el sistema resultante al unir las ecuaciones. Pero ahora nos ocupa $\mathbb{F}+\mathbb{G}$. Lo que va bien es que $\mathbb{F}$ y $\mathbb{G}$ vengan dados por ecuaciones *paramétricas* o, de modo más general, que vengan dados como $\mathbb{F}=\lg(a_1,\ldots,a_p)$ y $\mathbb{G}=\lg(b_1,\ldots,b_q)$.

Problema 90 *Si vienen dados así* $\mathbb{F}$ *y* $\mathbb{G}$, *entonces,* $\mathbb{F}+\mathbb{G}=\lg(a_1,\ldots,a_p,b_1,\ldots,b_q)$.

Problema 91 *Tomamos* $\Bbbk$ *de característica* cualquiera *y en* $\Bbbk^3=\mathbb{E}$ *cuatro vectores*

$$a_1=\begin{pmatrix}1\\1\\0\end{pmatrix},\quad a_2=\begin{pmatrix}0\\1\\1\end{pmatrix},\quad a_3=\begin{pmatrix}1\\-1\\0\end{pmatrix},\quad a_4=\begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

Determinar si alguno de los subespacios $\mathbb{F}=\lg(a_1,a_2)$, $\mathbb{G}=\lg(a_3,a_4)$ *y* $\mathbb{F}+\mathbb{G}$ *son* $\Bbbk^3=\mathbb{E}$.

Problema 92 *Sean* $\mathbb{F},\mathbb{G},\mathbb{H}$ *subespacios de* $\mathbb{E}$. *Estudiar si son ciertas o falsas las relaciones*

$$\mathbb{F}\cap(\mathbb{G}+\mathbb{H})=(\mathbb{F}\cap\mathbb{G})+(\mathbb{F}\cap\mathbb{H}),\qquad(\mathbb{F}\cap\mathbb{G})+\mathbb{H}=(\mathbb{F}+\mathbb{H})\cap(\mathbb{G}+\mathbb{H}).$$

(Al intersecar con la suma sale la suma de intersecciones y al sumar a la intersección sale la intersección de las sumas. Con números f,g,h, *la ecuación* $f\cdot(g+h)=f\cdot g+f\cdot h$ *es cierta y* $(f\cdot g)+h=(f+h)\cdot(g+h)$ *es falsa. Aquí nos planteamos cuestiones análogas. Si son falsas hay que dar un contraejemplo y si son ciertas probarlas.)*

Interesa ahora el concepto de **suma directa (de subespacios)**. Dados subespacios $\mathbb{F}_1,\ldots,\mathbb{F}_k$ de $\mathbb{E}$ diremos que la suma de estos subespacios es directa si cumple la siguiente propiedad *adicional*: la expresión de cada $x\in\sum_{i=1}^k\mathbb{F}_i$ en la forma $x=y_1+\ldots+y_k$ con cada y_j en $\mathbb{F}_j$ es única. Resaltamos que si se tienen $\mathbb{F}_1,\ldots,\mathbb{F}_k$, la suma $\mathbb{F}_1+\ldots+\mathbb{F}_k$ siempre está definida, pero el que la suma sea directa es una propiedad que puede ser cierta o no. Vamos a insistir un poco. Si $x\in\sum_{i=1}^k\mathbb{F}_i$, se tiene que poder escribir por definición de suma, *como mínimo* de una forma $x=y_1+\ldots+y_k$ con cada y_j en $\mathbb{F}_j$. Pero si la suma es directa, lo que tenemos es que se escribe *como máximo* de una forma $x=y_1+\ldots+y_k$ con cada y_j en $\mathbb{F}_j$; luego, si la suma es directa, la suma es única. El subespacio $\sum_{i=1}^k\mathbb{F}_i$ se escribirá como $\bigoplus_{i=1}^k\mathbb{F}_i$ para indicar que la suma es directa. Poder expresar $\mathbb{E}$ como suma directa de $\mathbb{F}_1,\ldots,\mathbb{F}_k$ es parecido (de hecho una generalización) de que $(a_1,\ldots,a_n)$ sea una base de $\mathbb{E}$. En ambos casos se puede expresar cada $x\in\mathbb{E}$ de modo único como suma de otros vectores y_j. En el caso de suma directa los $y_j\in\mathbb{F}_j$ y en el caso de base los y_j son de la forma $y_j=\lambda^ja_j$.

Problema 93 *Sea $\mathcal{U} = (u_1, \ldots, u_m)$ una base de $\mathbb{E}$. Probar que $\mathbb{E} = \lg(u_1) \oplus \ldots \oplus \lg(u_m)$.*

Teorema 44 *Sea $\mathbb{F} = \sum_{i=1}^{k} \mathbb{F}_i$ una suma de subespacios de $\mathbb{E}$. Son equivalentes las siguientes propiedades*

1. *Para cada $j = 1, 2, \ldots, k$, se tiene que*
$$[\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] \cap \mathbb{F}_j = 0;$$
o sea, al intersecar cada $\mathbb{F}_j$ con la suma de los otros $k-1$ subespacios resulta el subespacio cero.[12]
2. *Si se tiene $0 = z_1 + z_2 + \ldots + z_k$ con $z_j \in \mathbb{F}_j$ para cada j, se sigue que $z_1 = z_2 = \ldots = z_k = 0$.*
3. *La suma es directa; es decir, la expresión de cada $x \in \sum_{i=1}^{k} \mathbb{F}_i$ en la forma $x = y_1 + \ldots + y_k$ con cada y_j en $\mathbb{F}_j$ es única.*

Demostración. La demostración es más sencilla de seguir en el caso $k = 2$. La hacemos primero en ese caso, aunque, advertimos, no es una demostración por inducción sobre k, y el lector audaz puede ir directamente a la demostración general. La demostración será circular, la más elegante y económica muchas veces, con $1 \implies 2 \implies 3 \implies 1$.

Siempre $\mathbb{F} = \mathbb{F}_1 + \mathbb{F}_2$. La condición **1** equivale a $\mathbb{F}_1 \cap \mathbb{F}_2 = 0$. Supongamos que **1** es cierta y que $0 = z_1 + z_2$ con $z_i \in \mathbb{F}_i$. Se tiene $z_1 = -z_2$ luego $z_1 \in \mathbb{F}_1 \cap \mathbb{F}_2 = 0$ y, de $0 = z_1 + z_2$ sale $z_2 = 0$. Hemos probado **2.** Supongamos que **2** es cierta y que $x = y_1 + y_2 = y_1' + y_2'$. Restando, $0 = (y_1 - y_1') + (y_2 - y_2')$. Cada $z_i = y_i - y_i'$ está en $\mathbb{F}_i$, $i = 1, 2$, de modo que **2** nos da $y_1 - y_1' = y_2 - y_2' = 0$, equivalente a $y_1 = y_1'$ e $y_2 = y_2'$ y ya hemos probado **3**. Finalmente, supongamos **3** cierta y sea $x \in \mathbb{F}_1 \cap \mathbb{F}_2$. Se escribirá

$$x = x_1 + 0 = 0 + x_2, \quad x_1 \in \mathbb{F}_1,\ x_2 \in \mathbb{F}_2 \ \text{ para } \ x = x_1 = x_2,$$

y como la expresión es única, $x_1 = 0$ y $0 = x_2$, así que $x = 0$.

Para k arbitrario se hace así. Si se cumple **1** y $0 = z_1 + z_2 + \ldots + z_k$ con $z_j \in \mathbb{F}_j$ para cada j tenemos

$$z_j = -z_1 - z_2 - \ldots - z_{j-1} - z_{j+1} - \ldots - z_k,$$

luego $z_j \in [\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] \cap \mathbb{F}_j = 0$ y cada z_j es cero. El caso **2** $\implies$ **3** es análogo a lo visto con $k = 2$. Finalmente, si se cumple **3** y $x \in [\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] \cap \mathbb{F}_j$ se podrá escribir

$$x = 0 + \ldots + 0 + x_j + 0 + \ldots + 0, \qquad x = x_1 + \ldots + x_{j-1} + 0 + x_{j+1} + \ldots + x_k.$$

con cada $x_j \in \mathbb{F}_j$. La unicidad de la descomposición da que todos los x_j son nulos y $x = x_j = 0$. ♣

Es muy fácil visualizar en $\mathbb{R}^2$ y $\mathbb{R}^3$ una suma directa de dos factores. Por ejemplo, una suma de recta y plano en $\mathbb{R}^3$ se tiene en cuanto la recta es transversal al plano, pero como la recta corta al plano (al haber transversalidad) solo en el origen, vemos que la suma es directa. Dos planos de $\mathbb{R}^3$ se cortan en una recta o un plano (si los dos planos son iguales, caso límite). Hay suma, si los planos son distintos, pero nunca es directa. Si $n \geq 4$ la situación es difícil de visualizar aunque no de intuir. En general, si $\mathbb{F}_1$ y $\mathbb{F}_2$ dependen de demasiados parámetros, se facilita que sea $\mathbb{F}_1 \cap \mathbb{F}_2 \neq 0$ y no habrá suma directa. Esto se puede expresar con rigor una vez visto el concepto de dimensión y es consecuencia de la **fórmula de Grassmann** que pronto veremos.

Problema 94 *Sea $\Bbbk$ con característica que no sea 2 y sea $\mathbb{E}$ el espacio $\Bbbk^{n\times n}$ de las matrices cuadradas. Definimos $\mathbb{F}_1$ y $\mathbb{F}_2$ respectivamente como los subespacios de las matrices simétricas y antisimétricas. Probar que $\mathbb{E} = \mathbb{F}_1 \oplus \mathbb{F}_2$.* ♦

Solución. Recordamos que las condiciones $a^\top = a$ y $a^\top = -a$ dan respectivamente las matrices simétricas y antisimétricas. Para cualquier matriz $a \in \Bbbk^{n\times n}$ se tiene que

$$\left(a + a^\top\right)^\top = a^\top + \left(a^\top\right)^\top = a^\top + a = a + a^\top, \qquad \left(a - a^\top\right)^\top = a^\top + \left(-a^\top\right)^\top = a^\top - a = -\left(a - a^\top\right),$$

luego las matrices $a + a^\top$ y $a - a^\top$ son respectivamente simétricas y antisimétricas. Cualquier $a \in \Bbbk^{n\times n}$ se escribe como

$$a = \frac{1}{2}\left(a + a^\top\right) + \frac{1}{2}\left(a - a^\top\right),$$

[12] Obsérvese que para $k = 2$, el caso más frecuente, la condición es $\mathbb{F}_1 \cap \mathbb{F}_2 = 0$.

lo que nos muestra que $\mathbb{E} = \mathbb{F}_1 + \mathbb{F}_2$. Para tener $\mathbb{E} = \mathbb{F}_1 \oplus \mathbb{F}_2$ necesitamos ver que $\mathbb{F}_1 \cap \mathbb{F}_2 = 0$. Pero si a es a la vez simétrica y antisimétrica, se cumple $a^\top = a$ y $a^\top = -a$, luego $a = -a$ y $a = 0$. ♦

Hay un problema análogo con **funciones pares** e **impares**. Se dice que $f : \mathbb{R} \to \mathbb{R}$ es par o impar según se tenga respectivamente que para todo $x \in \mathbb{R}$ es $f(-x) = f(x)$ o $f(-x) = -f(x)$. Las funciones $f(x) = x^n$ son ejemplos de funciones pares e impares según n sea par o impar. La función coseno es par y la función seno es impar porque $\cos(-x) = \cos x$ y $\sin(-x) = -\sin x$.

Problema 95 *Sea $\mathbb{E}$ el espacio $\mathbb{k}^{n\times n}$ de las funciones de $\mathbb{R}$ en $\mathbb{R}$. Definimos $\mathbb{F}_1$ y $\mathbb{F}_2$ respectivamente como los subespacios de las funciones pares e impares. Probar que $\mathbb{E} = \mathbb{F}_1 \oplus \mathbb{F}_2$. Indicación: considerar funciones que combinan $f(x)$ y $f(-x)$.*

Problema 96 *Sea $\mathbb{E}$ el espacio $\mathbb{k}^{n\times n}$ de las matrices cuadradas. Definimos $\mathbb{F}_1$ y $\mathbb{F}_2$ respectivamente como los subespacios de las matrices triangulares superiores y triangulares inferiores[13]. Estudiar si $\mathbb{E} = \mathbb{F}_1 \oplus \mathbb{F}_2$. ¿Varía la situación si cambiamos $\mathbb{F}_2$ por $\mathbb{F}_2'$ de las matrices triangulares inferiores estrictas que son las que tienen nulos todos los coeficientes de la diagonal principal y los que están sobre ella?*

Si tenemos una suma directa y en cada sumando una base, alineando una tras otra todas las bases se obtiene una base del subespacio suma directa. Lo presentamos con más detalle.

Problema 97 *Sean $\mathbb{F}$ y $\mathbb{G}$ subespacios de $\mathbb{E}$ con bases respectivas $(u_1, \dots, u_n)$ y $(v_1, \dots, v_p)$ y $\mathbb{F} \cap \mathbb{G} = 0$. Probar que $(u_1, \dots, u_n, v_1, \dots, v_p)$ es base de $\mathbb{F} + \mathbb{G} = \mathbb{F} \oplus \mathbb{G}$. Generalizar al caso de una suma $\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_k$ con bases $\mathcal{U}_j = \left(u_{j1}, u_{j2}, \dots, u_{jn_j}\right)$ de $\mathbb{F}_j$ probando que la sucesión*

$$\mathcal{U} = (u_{11}, u_{12}, \dots, u_{1n_1}, u_{21}, u_{22}, \dots, u_{2n_2}, \dots, u_{k1}, u_{k2}, \dots, u_{kn_k})$$

obtenida alineando las bases es una base de $\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_k$.

Dados dos subespacios $\mathbb{F}$ y $\mathbb{G}$ de $\mathbb{E}$, las dimensiones $\dim(\mathbb{F} + \mathbb{G})$, $\dim(\mathbb{F} \cap \mathbb{G})$, $\dim(\mathbb{F})$ y $\dim(\mathbb{G})$ están relacionadas.[14]

Teorema 45 (fórmula de Grassmann) *Sean $\mathbb{F}$ y $\mathbb{G}$ subespacios de $\mathbb{E}$ de dimensiones n y p. Entonces*

$$\dim(\mathbb{F} + \mathbb{G}) + \dim(\mathbb{F} \cap \mathbb{G}) = \dim(\mathbb{F}) + \dim(\mathbb{G}).$$

Demostración. Sea $\mathbb{H} = \mathbb{F} \cap \mathbb{G}$ de dimensión q. Si $q = 0$, la suma es directa y el problema 97 nos da la fórmula de Grassmann. Sea $q > 0$. La estrategia de demostración es ampliar adecuadamente una base $\mathcal{W} = (h_1, \dots, h_q)$ de $\mathbb{H}$. Usaremos el teorema 37 que dice que una base de un subespacio $\mathbb{S}$ contenido en otro subespacio $\mathbb{T}$ puede alargarse hasta una base de $\mathbb{T}$. Ampliamos $\mathcal{W}$ a bases $\mathcal{U} = (h_1, \dots, h_q, f_{q+1}, \dots, f_n)$ de $\mathbb{F}$ y $\mathcal{V} = (h_1, \dots, h_q, g_{q+1}, \dots, g_p)$ de $\mathbb{G}$. Afirmamos que

$$\mathcal{B} = (h_1, \dots, h_q, f_{q+1}, \dots, f_n, g_{q+1}, \dots, g_p)$$

es base de $\mathbb{F} + \mathbb{G}$. Vemos que $\mathcal{B}$ genera $\mathbb{F} + \mathbb{G}$ porque $x \in \mathbb{F} + \mathbb{G}$ se puede escribir como $x = y + z$ con $y \in \mathbb{F}$ y $z \in \mathbb{G}$ y, al ser y y z combinaciones de vectores h y f para y y h y g para z, resulta x combinación de vectores de $\mathcal{B}$. Supongamos ahora que

$$0 = \sum_{i=1}^{q} \gamma^i h_i + \sum_{k=q+1}^{n} \alpha^k f_k + \sum_{j=q+1}^{p} \beta^j g_j.$$

Debemos probar que todas las α, β y γ son cero. Tenemos que

$$\sum_{k=q+1}^{n} \alpha^k f_k = -\sum_{i=1}^{q} \gamma^i h_i - \sum_{j=q+1}^{p} \beta^j g_j, \quad \text{con} \quad \sum_{k=q+1}^{n} \alpha^k f_k \in \mathbb{F} \text{ y } -\sum_{i=1}^{q} \gamma^i h_i - \sum_{j=q+1}^{p} \beta^j g_j \in \mathbb{G}.$$

Obtenemos entonces que

$$\sum_{k=q+1}^{n} \alpha^k f_k \in \mathbb{F} \cap \mathbb{G} = \mathbb{H} \text{ y se puede escribir } \sum_{k=q+1}^{n} \alpha^k f_k = \sum_{r=1}^{q} \theta^r h_r$$

[13] Los términos respectivamente bajo o sobre la diagonal principal son nulos.

[14] La demostración no requiere $\dim(\mathbb{E}) < \infty$, lo que tiene su interés cuando $\mathbb{E}$ es un espacio de funciones (casi siempre de dimensión infinita).

para ciertos coeficientes θ. pero entonces

$$0 = \sum_{r=1}^{q} \theta^r h_r - \sum_{k=q+1}^{n} \alpha^k f_k$$

y como $(h_1, \ldots, h_q, f_{q+1}, \ldots, f_n)$ es base de $\mathbb{F}$ todos los coeficientes son 0. La igualdad de partida se simplifica, puesto que las α son cero, convirtiéndose en

$$0 = \sum_{i=1}^{q} \gamma^i h_i + \sum_{j=q+1}^{p} \beta^j g_j \in \mathbb{G},$$

y, como $(h_1, \ldots, h_q, g_{q+1}, \ldots, g_p)$ es base de $\mathbb{G}$, todas las β y γ son nulas. Visto que todas las α, β, γ son nulas, $\mathcal{B}$ es independiente. Ya dijimos al principio que $\mathcal{B}$ generaba $\mathbb{F} + \mathbb{G}$, luego es base de $\mathbb{F} + \mathbb{G}$. Es ya fácil concluir porque $\mathcal{B}$ tiene longitud $q + (n - q) + (p - q)$ y

$$\dim(\mathbb{F} + \mathbb{G}) = q + (n - q) + (p - q) = p + n - q = \dim(\mathbb{F}) + \dim(\mathbb{G}) - \dim(\mathbb{F} \cap \mathbb{G}),$$

que es lo que había que probar. ♣

Problema 98 *Probar que en $\mathbb{E}$ de dimensión 3 dos planos $\mathbb{F}$ y $\mathbb{G}$ distintos se cortan en una recta $\mathbb{H}$ (muy intuitivo si $\mathbb{E} = \mathbb{R}^3$).*

Problema 99 *En $\mathbb{E}$ se toma una base $\mathcal{U} = (u_1, \ldots, u_m)$ y se define $\mathbb{F} = \lg(u_1, \ldots, u_{m-1})$ y $\mathbb{G} = \lg(v)$ siendo $v = u_1 + \ldots + u_m$. Estudiar si la suma $\mathbb{F} + \mathbb{G}$ es directa.*

Problema 100 *En $\mathbb{R}^4$ nos dan $\mathbb{F}$ y $\mathbb{G}$ con las ecuaciones*

$$\mathbb{F} : \begin{cases} x^1 + x^2 = 0 \\ x^1 - x^2 = 0 \end{cases}, \quad \mathbb{G} : \begin{cases} x^3 + x^4 = 0 \\ x^3 - hx^4 = 0 \end{cases}$$

Determinar en función de h cuándo puede haber suma directa $\mathbb{F} \oplus \mathbb{G}$. Si no la hubiera, ¿cuánto valen $\dim(\mathbb{F} \cap \mathbb{G})$ y $\dim(\mathbb{F} + \mathbb{G})$?

Problema 101 *Sea $\mathbb{E}$ de dimensión m y $(a_1, \ldots, a_n)$ y $(a_{n+1}, \ldots, a_{n+p})$ dos sucesiones independientes de vectores que generan $\mathbb{F}$ y $\mathbb{G}$. Supongamos que ningún a_{n+j} depende linealmente de $(a_1, \ldots, a_n)$ y ningún a_k con $k \leq n$ de $(a_{n+1}, \ldots, a_{n+p})$*

Dados $\mathbb{E}$ y un subespacio $\mathbb{F}$ de $\mathbb{E}$, se dice que otro subespacio $\mathbb{G}$ de $\mathbb{E}$ es un **suplementario de** $\mathbb{F}$ si $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$. Esto se visualiza muy bien en $\mathbb{E} = \mathbb{R}^2$ donde si $\mathbb{F}$ es una recta, $\mathbb{G}$ es otra recta transversal a ella (todas a través del origen). En $\mathbb{E} = \mathbb{R}^3$, si $\mathbb{F}$ es una recta, $\mathbb{G}$ es un plano transversal a ella y si $\mathbb{F}$ es un plano, $\mathbb{G}$ es una recta transversal (todo a través del origen). En general el teorema 37 nos dice que si $\dim(\mathbb{E}) = n$ y $\mathbb{F}$ subespacio de $\mathbb{E}$ tiene dimensión m, una base $(u_1, \ldots, u_m)$ de $\mathbb{F}$ puede ser ampliada a una base $(u_1, \ldots, u_m, \ldots, u_n)$ de $\mathbb{E}$.

Problema 102 *Probar que $\mathbb{G} = \lg(u_{m+1}, \ldots, u_n)$ es un suplementario de $\mathbb{F}$, luego todo subespacio de un espacio de dimensión finita tiene suplementarios.*

El que las sumas sean directas repercute en cómo son las dimensiones de los sumandos.

Teorema 46 *Sean $\mathbb{F}_1, \ldots, \mathbb{F}_k$ subespacios de dimensión finita de $\mathbb{E}$. Para que su suma sea directa es necesario y suficiente que* $\dim(\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_k) = \dim(\mathbb{F}_1) + \ldots + \dim(\mathbb{F}_k)$.

Demostración. En el problema 97 se vio que cuando la suma era directa, con las bases $\mathcal{U}_1, \ldots, \mathcal{U}_k$ se obtenía, yuxtaponiéndolas, la base $\mathcal{U}$, con una longitud como sucesión suma de las longitudes de las $\mathcal{U}_j$. Claramente, si la suma es directa, la dimensión del espacio suma es suma de las dimensiones de los sumandos.

Probamos la recíproca. Abreviamos $\mathbb{F} = \mathbb{F}_1 + \ldots + \mathbb{F}_k$, $n = \dim(\mathbb{F})$, $n_j = \dim(\mathbb{F}_j)$. Por hipótesis, $n = \sum_{h=1}^{k} n_h$. Si la suma no fuera directa, tendríamos un j tal que

$$[\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] \cap \mathbb{F}_j \neq 0, \quad \dim([\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] \cap \mathbb{F}_j) = d_j > 0.$$

(Ver teorema 44.) Con la hipótesis en $\overset{(1)}{=}$ y la fórmula de Grassmann en $\overset{(2)}{=}$,

$$\begin{aligned}\sum_{h=1}^{k} n_h &\overset{(1)}{=} n = \dim(\mathbb{F}) = \dim([\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k] + \mathbb{F}_j) \\ &\overset{(2)}{=} \dim([\mathbb{F}_1 + \ldots + \mathbb{F}_{j-1} + \mathbb{F}_{j+1} + \ldots + \mathbb{F}_k]) + \dim(\mathbb{F}_j) - d_j \\ &\leq (n_1 + \ldots + n_{j-1} + n_{j+1} + \ldots + n_k) + n_j - d_j = \sum_{h=1}^{k} n_j - d_j < \sum_{h=1}^{k} n_j = n\end{aligned}$$

porque $d_j > 0$. Es imposible que sea $n < n$ y esta contradicción implica que la suma es directa. ♣

Se pueden plantear muchos problemas numéricos donde nos dan $\mathbb{F}$ y $\mathbb{G}$ subespacios de $\mathbb{E}$, bien sea en paramétricas o en implícitas y nos piden la dimensión de $\mathbb{F}+\mathbb{G}$ o de $\mathbb{F}\cap\mathbb{G}$. Las reglas son estas

1. Si $\mathbb{F}$ y $\mathbb{G}$ están en paramétricas, digamos que $\mathbb{F} = \lg(a_1, \ldots, a_p)$ y $\mathbb{G} = \lg(a_{p+1}, \ldots, a_{p+q})$ se tiene que $\mathbb{F}+\mathbb{G} = \lg(a_1, \ldots, a_{p+q})$. Tomamos una base $\mathcal{U}$ y cada a_j tiene coordenadas en ella dada por el vector columna $\left(a_j^1, \ldots, a_j^n\right)$ con los que formamos una matriz $a \in \Bbbk^{n\times(p+q)}$. El rango de esta matriz es $\dim(\mathbb{F}+\mathbb{G})$.

2. Si $\mathbb{F}$ y $\mathbb{G}$ están en implícitas y digamos que, para una base $\mathcal{U}$ tienen ecuaciones

$$\begin{aligned}f^i\left(x^1, \ldots, x^n\right) &= a_1^i x^1 + \cdots + a_n^i x^n = 0,\ (1 \leq i \leq h), \\ g^j\left(x^1, \ldots, x^n\right) &= a_1^j x^1 + \cdots + a_n^j x^n = 0,\ (h+1 \leq j \leq k),\end{aligned}$$

   la intersección $\mathbb{F}\cap\mathbb{G}$ viene dada por $f^1 = \ldots = f^h = g^{h+1} = \ldots = g^{h+k} = 0$. Este sistema es $ax = 0$, siendo a la matriz cuyas filas son las diversas $(a_1^t, \ldots, a_n^t)$, de modo que $a \in \Bbbk^{(h+k)\times n}$. Como hemos dicho, $\mathbb{F}\cap\mathbb{G}$ es en coordenadas, el espacio de soluciones de $ax = 0$, luego $\dim(\mathbb{F}\cap\mathbb{G}) = n - \operatorname{rg}(a)$.

3. Si $\mathbb{F}$ y $\mathbb{G}$ están en paramétricas pero piden $\dim(\mathbb{F}\cap\mathbb{G})$, compensa calcular $\dim(\mathbb{F}+\mathbb{G})$ y usar la fórmula de Grassmann. Si $\mathbb{F}$ y $\mathbb{G}$ están en implícitas y se quiere calcular $\dim(\mathbb{F}\cup\mathbb{G})$ compensa calcular $\dim(\mathbb{F}\cap\mathbb{G})$ y usar la fórmula de Grassmann.

Problema 103 *Verificar lo que no parezca evidente y preparar uno o dos porblemas numéricos.* ♦

Solución. Preparamos un problema numérico. Tomamos en $\mathbb{E} = \mathbb{R}^4$ cuatro vectores

$$a_1 = \begin{pmatrix}1\\2\\-1\\2\end{pmatrix},\ a_2 = \begin{pmatrix}-1\\3\\2\\-1\end{pmatrix},\ a_3 = \begin{pmatrix}5\\-5\\-8\\7\end{pmatrix} = 2\begin{pmatrix}1\\2\\-1\\2\end{pmatrix} - 3\begin{pmatrix}-1\\3\\2\\-1\end{pmatrix},\ a_4 = \begin{pmatrix}1\\1\\1\\0\end{pmatrix},$$

Definimos $\mathbb{F} = \lg(a_1, a_2)$ y $\mathbb{G} = \lg(a_3, a_4)$, suponiendo que el lector no sabe que $2a_1 - 3a_2 = a_3$, lo que permite jugar con ventaja y saber que $\mathbb{F}+\mathbb{G} = \lg(a_1, a_2, a_4)$. En todo caso, bien sea calculando $\operatorname{rg}(a_1, a_2, a_4)$ o $\operatorname{rg}(a_1, a_2, a_3, a_4)$, que es 3, conocemos que $\dim(\mathbb{F}+\mathbb{G}) = 3$. Preguntamos al lector $\dim(\mathbb{F}\cap\mathbb{G})$, que es fácil porque

$$\dim(\mathbb{F}\cap\mathbb{G}) = -\dim(\mathbb{F}+\mathbb{G}) + \dim(\mathbb{F}) + \dim(\mathbb{G}) = -3 + 2 + 2 = 1,$$

dado que $\dim(\mathbb{F}) = \dim(\mathbb{G}) = 2$. La fórmula de Grassmann no permite escribir sin más $\mathbb{F}\cap\mathbb{G} = \lg(v)$ con v a determinar. No obstante, si conocemos que $2a_1 - 3a_2 = a_3$, queda claro que $a_3 \in \mathbb{F}\cap\mathbb{G}$ y, al ser unidimensional (aquí sí se usa esa fórmula), llegamos a que $\mathbb{F}\cap\mathbb{G} = \lg(a_3)$. ♦

2.7. Sobre subespacios afines

Hasta ahora hemos hablado casi exclusivamente de subespacios *vectoriales*, que se visualizaban como rectas y planos siempre a través del origen. Expresados en implícitas si $\mathbb{E} = \Bbbk^n$, eran espacios de soluciones de sistemas *homogéneos* $ax = 0$. ¿Cómo vamos a tratar el caso de planos o rectas que no pasan por el origen o espacios de soluciones de sistemas no homogéneos $ax = y$ con $y \neq 0$?

Elegido un vector $t \in \mathbb{E}$ definimos la **traslación (asociada a** t**)** como la función $T : \mathbb{E} \to \mathbb{E}$ definida por $T(x) = x + t$. Si tenemos un subespacio vectorial $\mathbb{F}$ de $\mathbb{E}$ y lo trasladamos; es decir, consideramos

$$\mathbf{A} = T(\mathbb{F}) = \{t + x \mid x \in \mathbb{F}\}$$

tenemos por definición un **subespacio afín**.[15] Los subespacios afines son trasladados de los subespacios vectoriales. Si $t = 0$ o, más generalmente, si $t \in \mathbb{F}$, se tiene que $T(\mathbb{F}) = \mathbb{F}$ luego los subespacios vectoriales son un caso particular de subespacios afines.

Es muy sencillo, una vez que sabemos manejar las ecuaciones de subespacios vectoriales, hacer lo mismo con subespacios afines de $\mathbb{E}$. Si $\mathbf{A} = T(\mathbb{F}) = \{t + v \mid v \in \mathbb{F}\}$ y $\mathbb{F} = \lg(a_1, \ldots, a_n)$, sabemos que cada $x \in \mathbf{A}$ se puede poner en la forma $x = t + \lambda^1 a_1 + \ldots + \lambda^n a_n$, si bien los parámetros λ^i pueden no ser únicos pero lo son si $(a_1, \ldots, a_n)$ es independiente. En todo caso se dice que $x = t + \lambda^1 a_1 + \ldots + \lambda^n a_n$, sobrentendiendo que $\lambda^1, \ldots, \lambda^n \in \Bbbk$, es una **ecuación paramétrica de A**. Si adicionalmente tenemos una base $\mathcal{U}$ de $\mathbb{E}$, supuesto de dimensión m (¡ojo a lo que son m y n!),[16] podemos asignar coordenadas a t y los a_i y entonces hay cuatro notaciones equivalentes (en la primera no aparece $\mathcal{U}$) para expresar $\mathbf{A}$ en paramétricas, que son,

$$x = t + \lambda^1 a_1 + \ldots + \lambda^n a_n, \qquad \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix} = \begin{pmatrix} t^1 \\ \vdots \\ t^m \end{pmatrix} + \lambda^1 \begin{pmatrix} a_1^1 \\ \vdots \\ a_1^m \end{pmatrix} + \ldots + \lambda^n \begin{pmatrix} a_n^1 \\ \vdots \\ a_n^m \end{pmatrix},$$

$$\begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix} = \begin{pmatrix} t^1 \\ \vdots \\ t^m \end{pmatrix} + \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^n \end{pmatrix}, \qquad \begin{cases} x^1 = t^1 + \lambda^1 a_1^1 + \ldots + \lambda^n a_n^1 \\ \vdots \\ x^m = t^m + \lambda^1 a_1^m + \ldots + \lambda^n a_n^m \end{cases}.$$

Otra posibilidad son las ecuaciones implícitas, que al no tener aún el concepto de forma lineal, describimos solo para $\mathbb{E} = \Bbbk^n$.[17] Tenemos una matriz $a \in \Bbbk^{m \times n}$ y $\alpha \in \Bbbk^m$ y definimos $\mathbf{A} = \{x \in \Bbbk^n \mid ax = \alpha\}$. Es muy fácil ver que, si $ax = \alpha$ es *compatible* (porque si no sería vacío) $\mathbf{A}$ es un subespacio afín. En efecto, al ser compatible hay al menos un $t \in \Bbbk^n$ tal que $at = \alpha$ y $a(x - t) = 0$. Como los v tales que $av = 0$ forman un subespacio vectorial $\mathbb{F}$ de $\Bbbk^n$, vemos que $x \in \mathbf{A}$ equivale a que $x + t \in \mathbb{F}$, lo que nos lleva a que $\mathbf{A} = t + \mathbb{F}$ y $\mathbf{A}$ es un subespacio afín. La ecuación compacta $ax = \alpha$ tiene formas equivalentes

$$\begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^m \end{pmatrix}, \qquad \begin{cases} a_1^1 x^1 + \ldots + a_n^1 x^n = \alpha^1 \\ \vdots \\ a_1^m x^1 + \ldots + a_n^m x^n = \alpha^m \end{cases},$$

que son las ecuaciones implícitas de $\mathbf{A}$. El lector comprobará enseguida que si un $t \in \Bbbk^n$ concreto cumple $at = \alpha$, la ecuación $a(x - t) = 0$ es la ecuación en implícitas de $\mathbf{A} = t + \mathbb{F}$, siendo $\mathbb{F} = \{v \in \Bbbk^n \mid av = 0\}$.

Todo esto se puede ampliar pero preferimos esperar al capítulo *Funciones lineales*. Vamos a añadir solamente que se puede pasar de un tipo a otro de ecuación con lo que ya sabemos de ecuaciones lineales. Si nos dan $\mathbf{A}$ en la forma $ax = \alpha$ resolvemos el sistema $ax = \alpha$ pasándolo a forma escalonada o escalonada reducida. La solución aparece en la forma $x = t + \lambda^1 v_1 + \cdots + \lambda^k v_k$ donde incluso, en el caso de escalonada reducida, se prueba que los v_i son independientes. Ya tenemos una ecuación paramétrica.

Problema 104 *Sea* $\mathbf{A} \subset \mathbb{R}^4$ *con ecuaciones implícitas* $ax = \alpha$*, siendo*

$$a = \begin{pmatrix} 1 & 3 & 0 & -4 \\ 2 & 4 & 2 & -4 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Dar ecuaciones paramétricas para $\mathbf{A}$. ♦

[15] Como $\mathbb{F}$ nunca es vacío, un subespacio afín nunca puede ser vacío.

[16] Estos cambios de la letra m o n para la dimensión de $\mathbb{E}$ o un subespacio $\mathbb{F}$ son un poco irritantes, pero tienen una lógica. Conviene que cuando vaya a aparecer una matriz a en la discusión, esta sea $m \times n$. El autor piensa que así se equivoca menos, pero también se podría mantener a rajatabla que $\dim(\mathbb{E}) = m$, $\dim(\mathbb{F}) = n$ y que sean las matrices como les toque ser.

[17] Para que la matriz a que pronto aparecerá sea $m \times n$, trabajamos en $\Bbbk^n = \mathbb{E}$, que tiene dimensión n, mientras que para las paramétricas, $\dim(\mathbb{E}) = m$.

Solución. No es nada nuevo. Se resuelve $ax = \alpha$ con operaciones

$$\left(\begin{array}{cccc|c} 1 & 3 & 0 & -4 & 1 \\ 2 & 4 & 2 & -4 & -1 \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 3 & 0 & -4 & 1 \\ 0 & -2 & 2 & 4 & -3 \end{array}\right)$$
$$\to \left(\begin{array}{cccc|c} 1 & 3 & 0 & -4 & 1 \\ 0 & 1 & -1 & -2 & \frac{3}{2} \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 0 & 3 & 2 & -\frac{7}{2} \\ 0 & 1 & -1 & -2 & \frac{3}{2} \end{array}\right).$$

La solución general es

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} -\frac{7}{2} - 3x^3 - 2x^4 \\ \frac{3}{2} + x^3 + x^4 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} -\frac{7}{2} \\ \frac{3}{2} \\ 0 \\ 0 \end{pmatrix} + \lambda^1 \begin{pmatrix} -3 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \lambda^2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

que nos informa que si t, v_1 y v_2 son los tres vectores a la derecha, $\mathbf{A}$ pasa por t, y $\mathbb{F} = \lg(v_1, v_2)$. ♦

El paso de paramétricas a implícitas es algo menos evidente, pero también conocido. Si en $\Bbbk^m$ tenemos $\mathbf{A}$ con ecuación paramétrica $x = t + \lambda^1 a_1 + \ldots + \lambda^n a_n$, hacemos $y = x - t$ y se trata de determinar los y que hacen $a\lambda = y$ un sistema compatible. (la incógnita es λ, no x como de costumbre). Llevamos $a\lambda = y$ a forma escalonada o escalonada reducida $b\lambda = z$ y si las fillas nulas de b son $b^{r+1}, \ldots, b^m$, la compatibilidad del sistema es $z^{r+1} = \ldots = z^m = 0$, que son las ecuaciones implícitas de $\mathbf{A}$.

Problema 105 *Dar ecuaciones implícitas para el subespacio afín de $\mathbb{R}^4$ con ecuaciones paramétricas*

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} + \lambda^1 \begin{pmatrix} 1 \\ 3 \\ 0 \\ -4 \end{pmatrix} + \lambda^2 \begin{pmatrix} 2 \\ 4 \\ 2 \\ -4 \end{pmatrix}. \blacklozenge$$

Solución. Para $y = x - t$ con $t = (1, -1, 1, -1)^\top$ hacemos operaciones

$$\left(\begin{array}{cc|c} 1 & 2 & y^1 \\ 3 & 4 & y^2 \\ 0 & 2 & y^3 \\ -4 & -4 & y^4 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & y^1 \\ 0 & -2 & y^2 - 3y^1 \\ 0 & 2 & y^3 \\ 0 & 4 & y^4 + 4y^1 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 2 & y^1 \\ 0 & -2 & y^2 - 3y^1 \\ 0 & 0 & y^3 + \left(y^2 - 3y^1\right) \\ 0 & 0 & y^4 + 4y^1 + 2\left(y^2 - 3y^1\right) \end{array}\right).$$

La compatibilidad del sistema equivale a

$$\left\{\begin{array}{l} -3y^1 + y^2 + y^3 = 0 \\ -2y^1 + 2y^2 + y^2 = 0 \end{array}\right., \text{ que equivale también a } \left\{\begin{array}{l} -3\left(x^1 - 1\right) + \left(x^2 + 1\right) + \left(x^3 - 1\right) = 0 \\ -2\left(x^1 - 1\right) + 2\left(x^2 + 1\right) + \left(x^4 + 1\right) = 0 \end{array}\right.,$$

y estas son las ecuaciones en implícitas del subespacio. Dejar las ecuaciones en función de $\left(x^i - t^i\right)$ con ceros a la derecha permite saber inmediatamente que $t = \left(t^1, \ldots, t^m\right)^\top$ está en el subespacio. ♦

Pasamos a cuestiones más teóricas sin ecuaciones. Al escribir $\mathbf{A} = T(\mathbb{F}) = \{t + x \mid x \in \mathbb{F}\}$ surge la pregunta de si t y $\mathbb{F}$, los "datos" de $\mathbf{A}$, quedan unívocamente determinados por $\mathbf{A}$, o sea, si puede suceder que $\mathbf{A} = T_1(\mathbb{F}_1) = T_1(\mathbb{F}_1)$ pudiendo ser $t_1 \neq t_2$ o $\mathbb{F}_1 \neq \mathbb{F}_2$. Se intuye que $\mathbf{A}$ determina $\mathbb{F}$ unívocamente pero que hay muchos t que cumplen $\mathbf{A} = T(\mathbb{F})$. Esta intuición es correcta. Con más precisión

Teorema 47 *Dados $\mathbf{F} = s + \mathbb{F}$ y $\mathbf{G} = t + \mathbb{G}$ se tiene que*

1. *$\mathbf{F} \subset \mathbf{G}$ si y solo si $s - t \in \mathbb{G}$ y $\mathbb{F} \subset \mathbb{G}$.*
2. *$\mathbf{F} = \mathbf{G}$ si y solo si $s - t \in \mathbb{G}$ y $\mathbb{F} = \mathbb{G}$.*
3. *$\mathbf{F} \cap \mathbf{G} \neq \emptyset$ si y solo si $s - t \in \mathbb{F} + \mathbb{G}$. En tal caso, escribiendo $s - t = -u + v$ con $u \in \mathbb{F}$ y $v \in \mathbb{G}$ se tiene para $r = s + u = t + v$ que $\mathbf{F} \cap \mathbf{G} = r + (\mathbb{F} \cap \mathbb{G})$.*

Demostración. Supongamos $\mathbf{F} \subset \mathbf{G}$ luego se puede poner $s + 0 = t + u$ con $u \in \mathbb{G}$ que implica $s - t \in \mathbb{G}$. Asimismo, si $u \in \mathbb{F}$, ha de ser $s + u = t + v$ para cierto $v \in \mathbb{G}$ y entonces $u = -(s - t) + v \in \mathbb{G}$. Recíprocamente, si $s - t \in \mathbb{G}$ y $\mathbb{F} \subset \mathbb{G}$, dado $x = s + u \in \mathbf{F}$ escribimos $x = t + (s - t) + u \in t + \mathbb{G} + \mathbb{G} = \mathbf{G}$. Probado **1**, **2** es consecuencia inmediata de **1**.

Si $\mathbf{F}\cap\mathbf{G}\neq\emptyset$ tomamos x en $\mathbf{F}\cap\mathbf{G}$ y lo escribimos $x=s+u=t+v$, con $u\in\mathbb{F}$ y $v\in\mathbb{G}$. Entonces $s-t=-u+v$. Recíprocamente, si tenemos $s-t=-u+v$ con $u\in\mathbb{F}$ y $v\in\mathbb{G}$, definimos $r=s+u=t+v$, que sin duda está en $\mathbf{F}\cap\mathbf{G}$. Además, $r-s=u\in\mathbb{F}$, luego por **2** es $\mathbf{F}=s+\mathbb{F}=r+\mathbb{F}$ e igualmente, $\mathbf{G}=t+\mathbb{G}=r+\mathbb{G}$. Al ser $\mathbb{F}\cap\mathbb{G}\subset\mathbb{F}$ se cumplirá $r+(\mathbb{F}\cap\mathbb{G})\subset r+\mathbb{F}$ e igualmente $r+(\mathbb{F}\cap\mathbb{G})\subset r+\mathbb{G}$, luego $r+(\mathbb{F}\cap\mathbb{G})\subset\mathbf{F}\cap\mathbf{G}$. Por otra parte, si $x\in\mathbf{F}\cap\mathbf{G}$ se puede escribir $x=r+u=r+v$ con $u\in\mathbb{F}$ y $v\in\mathbb{G}$. Así pues $x-r=u=v\in\mathbb{F}\cap\mathbb{G}$ y $x\in r+(\mathbb{F}\cap\mathbb{G})$. Queda probado todo **3**. ♣

El teorema prueba en **2** que $\mathbf{F}$ determina $\mathbb{F}$ unívocamente. Se llama a $\mathbb{F}$ la **dirección** de $\mathbf{F}$ y la **dimensión** de $\mathbf{F}$ es la de su dirección, luego $\dim\mathbf{F}=\dim\mathbb{F}$. El caso con $\dim\mathbf{A}=1$ y $\mathbf{F}=\mathbf{D}$ se llama **recta**, y el caso $\dim\mathbf{F}=n-1$ y $\mathbf{F}=\mathbf{H}$ se llama **hiperplano**, que puede llamarse también **plano** si $\dim\mathbf{F}=2$. En $\mathbb{k}^3$ un plano es un hiperplano y en $\mathbb{k}^2$ un hiperplano es una recta, cosa que suena muy rara pero es a lo que lleva la aplicación estricta de las definiciones. Puede enunciarse **2** como que un espacio afín queda determinado por su dirección y un vector s en él.

En el resto de la sección $\mathbf{F}=s+\mathbb{F}$ y $\mathbf{G}=t+\mathbb{G}$ serán dos subespacios afines.

Diremos que $X,Y\subset\mathbb{E}$ no vacíos son **paralelos** si existe una traslación tal que $T(X)\subset Y$ o bien $T(Y)\subset X$. Aplicaremos el concepto a subespacios donde hay una definición equivalente más familiar.

Problema 106 *Probar que $\mathbf{F}$ y $\mathbf{G}$ son paralelos si y solo si $\mathbb{F}\subset\mathbb{G}$ o $\mathbb{G}\subset\mathbb{F}$.*

Problema 107 *Dado un conjunto de subespacios afines $\mathbf{F}_i$ con $i\in I$, su intersección $\bigcap_{i\in I}\mathbf{F}_i$ o es vacía o es un subespacio afín. Si $s\in\bigcap_{i\in I}\mathbf{F}_i$ y $\mathbb{F}=\bigcap_{i\in I}\mathbb{F}_i$ se cumple que $s+\bigcap_{i\in I}\mathbb{F}_i=\bigcap_{i\in I}\mathbf{F}_i$ luego la intersección de las direcciones es la dirección de la intersección.*

En general, la unión $\mathbf{F}\cup\mathbf{G}$ de subespacios afines no es un subespacio afín. Definimos la **adjunción** de $\mathbf{F}$ y $\mathbf{G}$ (*joint* en inglés) como el mínimo subespacio que contiene a la unión $\mathbf{F}\cup\mathbf{G}$. Hemos pues de considerar todos los $\mathbf{H}_i$ tales que $\mathbf{H}_i\supset\mathbf{F}\cup\mathbf{G}$ y hacer su intersección, que es la adjunción, denotada por $\mathbf{F}\vee\mathbf{G}$. Por la propia definición, $\mathbf{F}\vee\mathbf{G}\supset\mathbf{F}\cup\mathbf{G}$ pero el contenido puede ser estricto. Analicemos qué supone probar que cierto $\mathbf{H}$ es $\mathbf{F}\vee\mathbf{G}$. Por una parte debemos mostrar que $\mathbf{F}\cup\mathbf{G}\subset\mathbf{H}$, lo que dará $\mathbf{F}\vee\mathbf{G}\subset\mathbf{H}$ puesto que $\mathbf{F}\vee\mathbf{G}$ es el espacio más pequeño de los que contienen a $\mathbf{F}\cup\mathbf{G}$. Por otra parte debemos tomar $\mathbf{K}$ arbitrario con $\mathbf{F}\cup\mathbf{G}\subset\mathbf{K}$ y probar que $\mathbf{H}\subset\mathbf{K}$, pues entonces, al ser $\mathbf{F}\vee\mathbf{G}$ la intersección de todos esos $\mathbf{K}$ tendremos $\mathbf{H}\subset\mathbf{F}\vee\mathbf{G}$. El problema que sigue pone a prueba si el lector ha comprendido esto.[18]

Problema 108 *Probar que*

1. *Sean como sean $\mathbf{F}$ y $\mathbf{G}$, si $p\in\mathbf{F}\cup\mathbf{G}$ se tiene $\mathbf{F}\vee\mathbf{G}=p+(\mathbb{F}+\mathbb{G}+\lg(s-t))$.* ♦
2. *Si $q\in\mathbf{F}\cap\mathbf{G}$ se tiene*
$$\mathbf{F}\vee\mathbf{G}=q+(\mathbb{F}+\mathbb{G}),\quad \dim(\mathbf{F}\vee\mathbf{G})+\dim(\mathbf{F}\cap\mathbf{G})=\dim\mathbf{F}+\dim\mathbf{G}.$$
3. *Si $\mathbf{F}\cap\mathbf{G}=\emptyset$, $\dim(\mathbf{F}\vee\mathbf{G})>\max\{\dim\mathbf{F},\dim\mathbf{G}\}$.*

Solución. Abreviamos $\mathbb{H}=(\mathbb{F}+\mathbb{G}+\lg(s-t))$, luego $\mathbf{H}=p+\mathbb{H}$. Como $p\in\mathbf{F}\cup\mathbf{G}$ digamos que $p\in\mathbf{F}$ y que $p=s+u_1$ con $u_1\in\mathbb{F}$. Es sencillo ver que $\mathbf{F}\subset\mathbf{H}$, pero también $\mathbf{G}\subset\mathbf{H}$ puesto que un $t+v\in\mathbf{G}$ cumple que
$$t+v=t-s+s+v=t-s+(p-u_1)+v\in p+\mathbb{H}$$
y por definición de $\mathbf{F}\vee\mathbf{G}$ se tiene $\mathbf{F}\vee\mathbf{G}\subset\mathbf{H}$. Queda probar que cualquier espacio $\mathbf{K}$ que contenga a $\mathbf{F}$ y $\mathbf{G}$ contiene también a $\mathbf{H}$. Dado que p también está en $\mathbf{K}$, elegimos expresar $\mathbf{K}=p+\mathbb{K}$. Sea $x=p+u+v+\lambda(s-t)\in\mathbf{H}$. Como $u\in\mathbb{F}$, $v\in\mathbb{G}$ y $\mathbb{F},\mathbb{G}\subset\mathbb{H}\subset\mathbb{K}$, tenemos $u+v\in\mathbb{K}$. Por otra parte $s,t\in\mathbf{H}\subset\mathbf{K}$ así que $s-t\in\mathbb{K}$. Queda claro que $u+v+(s-t)\in\mathbb{K}$ y $x\in\mathbf{K}$. Esto finaliza **1,** quedando **2** y **3** para el lector. ♦

Este problema es útil para expresar $\mathbf{F}\vee\mathbf{G}$ en paramétricas a partir de las paramétricas de $\mathbf{F}$ y $\mathbf{G}$ utilizando **1**. Si en $\mathbb{R}^4$ se tiene
$$\mathbf{F}:\begin{pmatrix}x^1\\x^2\\x^3\\x^4\end{pmatrix}=\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}+\lambda\begin{pmatrix}0\\1\\1\\0\end{pmatrix},\qquad \mathbf{G}:\begin{pmatrix}x^1\\x^2\\x^3\\x^4\end{pmatrix}=\begin{pmatrix}1\\2\\0\\3\end{pmatrix}+\mu\begin{pmatrix}0\\1\\1\\2\end{pmatrix},$$

[18] Ayuda imaginar la situación en $\mathbb{R}^3$ y, quizás, examinar los ejemplos numéricos tras el problema.

que son dos rectas en $\mathbb{R}^4$, el subespacio $\mathbf{F} \vee \mathbf{G}$ se expresa en paramétricas como

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix} + \theta \left(\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \end{pmatrix} \right).$$

El que tenga rango 3 la matriz

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

nos dice que $\dim(\mathbf{F} \vee \mathbf{G}) = 3$.

Problema 109 *Sean* $\mathbf{F}$ *y* $\mathbf{G}$ *de la misma dimensión* m *pero disjuntos. Probar que* $\dim(\mathbf{F} \vee \mathbf{G}) = m+1$ *equivale al paralelismo de* $\mathbf{F}$ *y* $\mathbf{G}$*. Indicación: para* $\Longrightarrow$ *probar que* $\dim(\mathbb{F} \cap \mathbb{G}) = \dim(\mathbb{F})$.

Problema 110 *Probar que* $\dim(\mathbf{F} \vee \mathbf{G}) - \dim(\mathbb{F} + \mathbb{G})$ *es* 0 *o* 1*, siendo* $\dim(\mathbf{F} \vee \mathbf{G}) = \dim(\mathbb{F} + \mathbb{G})$ *equivalente a* $\mathbf{F} \cap \mathbf{G} \neq \emptyset$.

Problema 111 *En un espacio tridimensional* $\mathbb{E}$ *se tienen dos rectas* $\mathbf{F} = s + \lg(u)$ *y* $\mathbf{G} = t + \lg(v)$ *con* (u, v) *independientes. Suponemos que no se cortan y por ello hay dos planos paralelos pero disjuntos* $\mathbf{F}_1 = s + \lg(u, v)$ *y* $\mathbf{G}_1 = t + \lg(u, v)$*. Probar que si* p *no está en esos planos hay una recta* $\mathbf{D}$ *única que pasa por* p *y corta a* $\mathbf{F}$ *y* $\mathbf{G}$ *exactamente en un punto en cada una. (Puede usarse que la recta que pasa por* y *y* z *es el conjunto de puntos de la forma* $x = (1-\theta)\, y + \theta z$ *con* $\theta \in \Bbbk$.*)*

Capítulo 3

Funciones lineales

3.1. Cuestiones generales

En todo el capítulo $\mathbb{E}$ y $\mathbb{F}$ representarán espacios vectoriales sobre el mismo cuerpo $\Bbbk$. Una **función o aplicación lineal,** u **homomorfismo** (de $\mathbb{E}$ en $\mathbb{F}$) es una función $L : \mathbb{E} \to \mathbb{F}$ que cumple que para todo $x, y \in \mathbb{E}$ y $\lambda \in \Bbbk$ se tiene $L(x+y) = L(x) + L(y)$ y $L(\lambda x) = \lambda L(x)$. Como muchas definiciones matemáticas, lo que indica es que ciertas operaciones pueden hacerse de modo diferente pero dando el mismo resultado. Si vemos una función como una "regla de transporte" de $\mathbb{E}$ en $\mathbb{F}$, la condición $L(x+y) = L(x) + L(y)$ dice que se consigue lo mismo sumando en $\mathbb{E}$ y transportando a $\mathbb{F}$ que transportando a $\mathbb{F}$ y sumando allí. Igualmente $L(\lambda x) = \lambda L(x)$ dice que "es lo mismo multiplicar y transportar que transportar y multiplicar".

El ejemplo estándar es el caso $\mathbb{E} = \Bbbk^n$, $\mathbb{F} = \Bbbk^m$ y $L(x) = ax$, siendo $a \in \Bbbk^{m\times n}$ (ojo al orden de los índices). Cada matriz $a \in \Bbbk^{m\times n}$ nos permite enviar vectores x de $\Bbbk^n$ en otros ax de $\Bbbk^m$ multiplicando a por x. Diremos que L es la **función lineal asociada a la matriz** $a \in \Bbbk^{m\times n}$, entendiendo por defecto que L va de $\Bbbk^n$ a $\Bbbk^m$. La comprobación de la linealidad es inmediata con las propiedades del producto de matrices. Veremos que muchos conceptos que se definirán para L general tienen una descripción familiar si $L(x) = ax$. Le advertimos al lector que hay muchos más ejemplos de función lineal, que debe intentar asimilar y que requieren ideas que van más allá de puro manejo de matrices.

La linealidad supone que las combinaciones lineales, fundamentales en la definición de conceptos (independencia, generación, bases, etc.) se conservan cuando el transporte es lineal. En efecto, si

$$x = \lambda^1 a_1 + \ldots + \lambda a_n, \text{ entonces } L(x) = \lambda^1 L(a_1) + \ldots + \lambda^n L(a_n). \tag{3.1}$$

En el caso en el que $\mathbb{F} = \Bbbk = \Bbbk^{1\times 1}$ las funciones lineales se llaman también **formas lineales**.

Problema 112 *Probar por inducción (3.1).*

Hay que tener cuidado sin embargo porque $(a_1, \ldots, a_n)$ puede ser independiente o generadora de $\mathbb{E}$ y ser falso que la sucesión transportada $(L(a_1), \ldots, L(a_n))$ a $\mathbb{F}$ lo sea. Esto y otras propiedades elementales de uso continuo están en el siguiente teorema.

Teorema 48 *Una función o aplicación lineal $L : \mathbb{E} \to \mathbb{F}$ verifica*

1. *Envía el cero de $\mathbb{E}$ al cero de $\mathbb{F}$ y el opuesto en $\mathbb{E}$ al opuesto en $\mathbb{F}$; abreviadamente, $L(0) = 0$ y $L(-x) = -L(x)$.*
2. *Sea $L(a_i) = b_i$, $1 \le i \le n$. Si $(a_1, \ldots, a_n)$ es dependiente, también lo es $(b_1, \ldots, b_n)$ y si $(b_1, \ldots, b_n)$ es independiente también lo es $(a_1, \ldots, a_n)$.*[1]
3. *Sean $\mathbb{S}$ y $\mathbb{T}$ respectivamente subespacios de $\mathbb{E}$ y $\mathbb{F}$. Entonces los subconjuntos*

$$\begin{cases} L(\mathbb{S}) = \{y \in \mathbb{F} \mid \text{existe } x \in \mathbb{E} \text{ tal que } y = L(x)\} \\ L^{-1}(\mathbb{T}) = \{x \in \mathbb{E} \mid \text{existe } y \in \mathbb{F} \text{ tal que } y = L(x)\} \end{cases}$$

[1] Abreviadamente, L preserva la dependencia en sentido directo y la independencia en sentido inverso.

son subespacios vectoriales de $\mathbb{F}$ *y* $\mathbb{E}$ *respectivamente.*[2]

Demostración. Usando $L(\lambda x) = \lambda L(x)$ y particularizando a $\lambda = 0$ y $x = 0$ resulta $L(0) = 0$. Si $\lambda = -1$, resula $L(-x) = L(-1 \cdot x) = (-1) \cdot L(x) = -L(x)$.

Supongamos en **2** que $(b_1, \ldots, b_n)$ independiente y $0 = \lambda^1 a_1 + \ldots + \lambda a_n$. Con (3.1),

$$0 = L(0) = L\left(\lambda^1 a_1 + \ldots + \lambda a_n\right) = \lambda^1 L(a_1) + \ldots + \lambda^n L(a_n) = \lambda^1 b_1 + \ldots + \lambda^n b_n,$$

y la independencia de $(b_1, \ldots, b_n)$ da $\lambda^1 = \ldots = \lambda^n = 0$, que supone la independencia de $(a_1, \ldots, a_n)$. La otra parte de **2** queda para el lector.

Para ver que $L(\mathbb{S})$ es subespacio tomamos $y_1 = L(x_1)$ e $y_2 = L(x_2)$ debiéndose mostrar que $y_1 + y_2 \in L(\mathbb{S})$. Esto está claro porque $y_1 + y_2 = L(x_1) + L(x_2) = L(x_1 + x_2)$. Igualmente, si $y = L(x)$ y $\lambda \in \Bbbk$ tenemos $\lambda y = \lambda(L(x)) = L(\lambda x)$, luego $\lambda y \in L(\mathbb{S})$. Para la imagen inversa se toman x_1, x_2 en $L^{-1}(\mathbb{T})$, luego los elementos $y_1 = L(x_1)$ e $y_2 = L(x_2)$, al ser sumados, están es $\mathbb{T}$. Debemos mostrar que $x_1, x_2 \in L^{-1}(\mathbb{T})$ pero esto es así porque $L(x_1 + x_2) = L(x_1) + L(x_2) = y_1 + y_2 \in \mathbb{T}$. La otra propiedad que debe cumplir $L^{-1}(\mathbb{T})$ queda para el lector. ♣

Basados en **3** destacamos dos subespacios asociados a L de suma importancia. La **imagen de** L es $L(\mathbb{E})$ (como en cualquier función). Se ha tomado pues $\mathbb{S} = \mathbb{E}$. El **núcleo de** L es $L^{-1}(0)$ y se ha tomado $\mathbb{T} = 0$. Por influencia del inglés y el alemán se denota el núcleo ("kernel" y "kern") por

$$\ker(L) = \{x \in \mathbb{E} \,|L(x) = 0\}.$$

La imagen es $\operatorname{im}(L)$. Si $L(\mathbb{E}) = \operatorname{im}(L) = \mathbb{F}$ estamos diciendo que L es suprayectiva. Será muy importante en los espacios de dimensión finita medir el "tamaño" de L (o de $L(\mathbb{E})$ si se prefiere ver así) por la dimensión de $L(\mathbb{E}) = \operatorname{im}(L)$. A las dimensiones de los subespacios imagen y núcleo de la aplicación lineal L se les llama el **rango de** L y la **nulidad de** L y se denotan por $\operatorname{rg}(L)$ y $\operatorname{nul}(L)$.

Si $L : \Bbbk^n \to \Bbbk^m$ está asociada a la matriz $a \in \Bbbk^{m \times n}$, es inmediato que $\ker(L)$ es el espacio de soluciones del sistema homogéneo $ax = 0$ y que $\operatorname{im}(L) = \{y = ax \mid x \in \Bbbk^n\}$. Alternativamente,

$$\operatorname{im}(L) = \left\{y = x^1 a_1 + x^2 a_2 + \ldots + x^n a_n \mid x^1, x^2, \ldots, x^n \in \Bbbk\right\},$$

e $\operatorname{im}(L) = \lg(a_1, \ldots, a_m)$ es el subespacio de $\Bbbk^m$ generado por las columnas $a_1, \ldots, a_m$ de a. Los textos están llenos de problemas que piden "calcular" $\ker(L)$ o $\operatorname{im}(L)$ en el sentido de "dar ecuaciones de…". Pues bien, en el caso en que L esté asociado a la matriz $a \in \Bbbk^{m \times n}$, $\ker(L)$ está dado en implícitas por $ax = 0$ e $\operatorname{im}(L)$ en paramétricas como $\operatorname{im}(L) = \lg(a_1, \ldots, a_m)$. Si necesitamos las ecuaciones paramétricas de $\ker(L)$ o las implícitas de $\operatorname{im}(L)$ hay que usar los procedimientos de los capítulos anteriores. Pueden pedir también una base de $\ker(L)$ o $\operatorname{im}(L)$ pero esto se ha visto ya como hacerlo. Para $\ker(L)$ es hallar una base del espacio de soluciones del sistema homogéneo $ax = 0$ y para $\operatorname{im}(L)$ supone quedarse en $(a_1, \ldots, a_m)$ con una subsucesión independiente que siga generando L. Finalmente, $\operatorname{rg}(L) = \dim(\operatorname{im}(L))$ es $\dim(\lg(a_1, \ldots, a_m))$ es, como ya sabemos, el rango de a, calculable por filas o columnas según convenga.

Si $\mathbb{E}$ tiene dimensión n, hemos visto que a cada base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$ se le puede asociar un isomorfismo $\Phi : \mathbb{E} \to \Bbbk^n$ que permite identificar $x = \sum_{i=1}^n x^i u_i \in \mathbb{E}$ con $\left(x^1, \ldots, x^n\right) = \operatorname{mat}^{\mathcal{U}}(x)$, la matriz de sus coordenadas o componentes en esa base. De este modo los entes matemáticos de $\mathbb{E}$ se correspondían de modo biunívoco con entes similares en $\Bbbk^n$, con la ventaja de que en $\Bbbk^n$ hay muchos recursos de cálculo. Con las funciones lineales $L : \mathbb{E} \to \mathbb{F}$ sucede algo parecido cuando $\mathbb{E}$ y $\mathbb{F}$ tienen dimensiones finitas n y m, porque, una vez elegidas bases $\mathcal{U}$ y $\mathcal{V}$ en $\mathbb{E}$ y $\mathbb{F}$, se puede asociar a L una matriz $a \in \Bbbk^{m \times n}$ que permite conocer L a través de su matriz a. No decimos aún cómo se define a, pero sí que tenemos la buena noticia que, *hasta cierto punto*, estudiar L entre espacios de dimensión finita, es como estudiar L asociado a la matriz a. *De todos modos, interesa que el lector se familiarice con el concepto de función lineal en su forma general y abstracta.*

Teorema 49 *Para que* $L : \mathbb{E} \to \mathbb{F}$ *sea inyectiva es necesario y suficiente que sea* $\ker(L) = 0$; *dicho verbalmente, que solo* $x = 0$ *en* $\mathbb{E}$ *sea aplicado sobre* $0 \in \mathbb{F}$. *Más generalmente, para que* $L(x_1) = L(x_2)$ *es necesario y suficiente que sea* $x_1 - x_2 \in \ker(L)$.

[2] Si $f : X \to Y$ se ve como una regla de transporte (totalmente arbitraria; no hace falta linealidad) y tenemos $S \subset X$ y $T \subset Y$ los conjuntos $f(S) \subset Y$ y $f^{-1}(T)$ son están respectivamente formados por los elementos de Y que han sido transportados desde S y los de X que *serían* transportados a T.

Demostración. Supongamos L inyectiva. Para toda L es $L(0)=0$ y si otro $x\in\mathbb{E}$ cumple $L(x)=0$, la inyectividad de L implica $x=0$, luego $\ker(L)=0$. Supongamos recíprocamente $\ker(L)=0$. Sean x_1,x_2 en $\mathbb{E}$ tales que $L(x_1)=L(x_2)$. Por linealidad, $0=L(x_1)-L(x_2)=L(x_1-x_2)$, luego $x_1-x_2\in\ker(L)$. Como $\ker(L)=0$ tiene que ser $x_1=x_2$. Así se ha probado por completo la primera afirmación y parte de la segunda. Solo queda ver que $x_1-x_2\in\ker(L)$ implica que $L(x_1)=L(x_2)$. Por hipótesis, $x_1-x_2=x_3\in\ker(L)$, de donde $L(x_1)=L(x_2+x_3)=L(x_2)+L(x_3)=L(x_2)$. ♣

Damos una colección de ejemplos de funciones lineales además del asociado a matrices.

1. La función $0:\mathbb{E}\to\mathbb{F}$ que lleva todo vector de $\mathbb{E}$ en el cero de $\mathbb{F}$ es lineal. Lo mismo pasa con las funciones, llamadas **homotecias**, de $\mathbb{E}$ en $\mathbb{E}$, por definición de la forma $L(x)=\kappa x$ siendo pues la multiplicación por un $\kappa\in\Bbbk$ fijo. Si $\kappa=1$ resulta $\mathrm{id}_{\mathbb{E}}$, la **identidad** de $\mathbb{E}$.

2. Una descomposición de $\mathbb{E}$ en suma directa $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$ da ejemplos sencillos pero importantes de homomorfismos. Definimos la **proyección (sobre $\mathbb{F}$)** y la **simetría (respecto de $\mathbb{F}$)**. Se toma $x\in\mathbb{E}$ y se escribe $x=y+z$ con $y\in\mathbb{F}$ y $z\in\mathbb{G}$ *de manera única.* La proyección y simetría se definen por
$$P,S:\mathbb{E}\to\mathbb{E},\qquad P(x)=y,\qquad S(x)=y-z.$$
Podemos también definirlas respecto de $\mathbb{G}$ siendo $Q(x)=z$, $T(x)=-y+z$.[3] Aunque se diga "proyección sobre $\mathbb{F}$" y parezca que P solo depende de $\mathbb{F}$, depende en realidad de la descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$ elegida. Pueden tenerse dos descomposiciones $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}_1=\mathbb{F}\oplus\mathbb{G}_2$ y ser $P_1\neq P_2$.

3. Sea $\mathbb{E}=\mathcal{F}(I,\Bbbk)$, el espacio de las funciones de I arbitrario en $\Bbbk$. Tenemos funciones lineales $\Phi:\mathcal{F}(I,\Bbbk)\to\Bbbk$ **por evaluación**. Se fija un punto $p\in I$ y se define $\Phi(f)=f(p)$. Por supuesto, si se toma una sucesión $(p_1,\ldots,p_h)$ de puntos de I se tiene $\Phi:\mathcal{F}(I,\Bbbk)\to\Bbbk^h$ por $\Phi(f)=(f(p_1),\ldots,f(p_h))$, también lineal. Si en vez de $\mathbb{E}=\mathcal{F}(I,\Bbbk)$ tomamos como $\mathbb{E}$ un espacio más concreto de funciones (polinomios, funciones continuas, etc.) tenemos también $\Phi(f)=(f(p_1),\ldots,f(p_h))$ lineal de $\mathbb{E}$ en $\Bbbk^h$.

4. En el Análisis aparecen operaciones lineales con la ayuda de la derivada y la integral. Si $\mathbb{E}$ es el espacio de funciones derivables de I (intervalo) en $\mathbb{R}$, $D(f)(x)=f'(x)$ es una función lineal con valores en $\mathcal{F}(I,\mathbb{R})$. Igualmente, si $\mathbb{E}$ es el espacio de las funciones continuas sobre $I=[0,1]$ con valores en $\mathbb{R}$, $\Phi(f)=\int_0^1 f(x)\,dx$ es lineal.

5. Hay muchos ejemplos de funciones lineales entre espacios de matrices. Elegida $a\in\Bbbk^{m\times n}$ se define $L:\Bbbk^{n\times p}\to\Bbbk^{m\times p}$ por $L(v)=av$. Otro ejemplo es la **trasposición** $a\to a^\top$ de $\Bbbk^{m\times n}$ en $\Bbbk^{n\times m}$.

Si $\mathbb{E}$ es un espacio de funciones no tenemos garantizado que las funciones lineales que partan de $\mathbb{E}$ o lleguen a $\mathbb{E}$ puedan estudiarse con matrices. Esta es una razón poderosa para que haya que considerar las definiciones generales. Las proyecciones y simetrías del ejemplo **2**, son ejemplos interesantes pues tienen una definición relativamente concreta sin matrices y hay que prestar atención a los problemas donde aparecen. El siguiente problema es muy fácil pero deben recordarse los resultados.

Problema 113 *Probar las siguientes propiedades de funciones lineales de los ejemplos anteriores:*

1. *El núcleo e imagen de la función $0:\mathbb{E}\to\mathbb{F}$ son los subespacios cero de $\mathbb{E}$ y $\mathbb{F}$ respectivamente. En las homotecias con $\kappa\neq 0$ se tiene* $\ker(L)=0$, $\mathrm{im}(L)=\mathbb{E}$.

2. *Si a la matriz $a\in\Bbbk^{m\times n}$ le asociamos $L(x)=ax$ de $\Bbbk^n$ en $\Bbbk^m$ se tiene que* $\ker(L)$ *es el espacio de soluciones de $ax=0$ y la imagen es el espacio de columnas de a.*

3. *En el ejemplo **5**,* $\ker(L)$ *está formado por las matrices v tales que $av=0$. Si a es cuadrada e invertible, $L(v)=av$ es inyectiva.*

4. *Si P,S son la proyección y simetría, sobre o respecto de $\mathbb{F}$, asociadas a la descomposición en suma directa $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$, tenemos que* $\ker(P)=\mathbb{G}$, $\mathrm{im}(P)=\mathbb{F}$, $\ker(S)=0$, $\mathrm{im}(S)=\mathbb{E}$.

5. *Si Φ es una evaluación con las notaciones del ejemplo **3**, tenemos que* $\ker(\Phi)$ *está formado por las f que se anulan a la vez en todos los p_i. ¿Es siempre* $\mathrm{im}(\Phi)=\Bbbk^h$*?*

[3] Si $\Bbbk$ tiene característica 2, al ser $-z=z$ resulta ser S la identidad. La definición no tiene interés en este caso.

6. *Sean D y Φ como en el ejemplo **4**. Se tiene que $\ker(D)$ está formado por las funciones constantes si bien es complicada una definición alternativa de $\operatorname{im}(D)$. Para Φ es complicada una definición alternativa de $\ker(\Phi)$ pero $\operatorname{im}(\Phi)=\mathbb{R}$. Aunque no tengamos definición alternativa de $\operatorname{im}(D)$ o $\ker(\Phi)$ podemos decir que los polinomios (funciones polinomiales) están en $\operatorname{im}(D)$ y que si f derivable cumple que $f(0)=f(1)$, entonces $D(f)\in\ker(\Phi)$.*

A partir de funciones lineales se pueden construir otras nuevas por suma, producto por escalares o composición.

Teorema 50 *Sean $L,M:\mathbb{E}\to\mathbb{F}$ y $\kappa\in\Bbbk$. Entonces las funciones*

$$L+M:\mathbb{E}\to\mathbb{F},\quad (L+M)(x)=L(x)+M(x),\qquad \kappa L:\mathbb{E}\to\mathbb{F},\quad (\kappa L)(x)=\kappa L(x)$$

son lineales. Si tenemos $L:\mathbb{E}\to\mathbb{F}$ y $M:\mathbb{F}\to\mathbb{G}$, también es lineal la composición

$$M\circ L:\mathbb{E}\to\mathbb{G},\qquad (M\circ L)(x)=M(L(x)).$$

Demostración. Hacemos solo la última parte (la primera es más fácil). Verificamos que

$$(M\circ L)(x+y)\overset{1}{=}M(L(x+y))\overset{2}{=}M(L(x)+L(y))\overset{2}{=}M(L(x))+M(L(y))\overset{1}{=}(M\circ L)(x)+(M\circ L)(y),$$

$$(M\circ L)(\lambda x)\overset{1}{=}M(L(\lambda x))\overset{2}{=}M(\lambda L(x))\overset{2}{=}\lambda M(L(x))\overset{1}{=}\lambda(M\circ L)(x).$$

En los pasos $\overset{1}{=}$ se han usado las definiciones y en $\overset{2}{=}$ la linealidad de L o M. ♣

Se pueden construir muchos ejemplos con los anteriores. Por ejemplo, si $\mathbb{E}=\mathbb{R}[X]$, el espacio de los polinomios reales (los veremos como funciones), la función

$$L:\mathbb{E}=\mathbb{R}[X]\longrightarrow\mathbb{R},\qquad L(f)=f'(0)+2f(1)$$

es lineal. Quizás sea más rápido una comprobación directa, pero la función $f\to f'$ (derivada de f) es lineal, compuesta con la evaluación da $f\to f'(0)$ lineal. Al multiplicar por 2 otra evaluación $f\to f(1)$, sale otra función lineal. Sumando las dos funciones lineales anteriores obtenemos L.

Problema 114 *Sean P,S son la proyección y simetría respecto de $\mathbb{F}$ asociadas a $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$ (ejemplo **4**) y sean Q y T la proyección y simetría respecto de $\mathbb{G}$ para la misma descomposición.*

1. *Calcular $P\circ Q$, $Q\circ P$, $S\circ T$ y $T\circ S$.*
2. *Probar que $P\circ P=P$ y $S\circ S=\mathrm{id}$.*
3. *Sea $L:\mathbb{E}\to\mathbb{E}$ una función lineal tal que $L\circ L=L$. Probar que si definimos $\mathbb{K}=\ker(L)$ y $\mathbb{J}=\operatorname{im}(L)$ se cumple que $\mathbb{E}=\mathbb{J}\oplus\mathbb{K}$ siendo la proyección P asociada a $\mathbb{E}=\mathbb{J}\oplus\mathbb{K}$ la propia L.*
4. *Sea $M:\mathbb{E}\to\mathbb{E}$ una función lineal tal que $M\circ M=\mathrm{id}$ y $\Bbbk$ con característica distinta de 2. Probar que si definimos*
$$\mathbb{J}=\{x\in\mathbb{E}\,|M(x)=x\},\qquad \mathbb{I}=\{x\in\mathbb{E}\,|M(x)=-x\}$$
tenemos una descomposición $\mathbb{E}=\mathbb{J}\oplus\mathbb{I}$, siendo la simetría S asociada a esta descomposición la propia M. Indicación: se tiene
$$x=\frac{x+M(x)}{2}+\frac{x-M(x)}{2}.$$

Los apartados **3** y **4** del problema son interesantes pues dicen que las propiedades $L\circ L=L$ y $M\circ M=\mathrm{id}$ indican que, aunque no haya referencia aparente a una suma directa $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$, son de hecho una proyección y una simetría y tenemos recursos para construir esa suma directa. *Hay pues dos definiciones equivalentes de proyección y simetría, si bien no se adivina el sentido geométrico de $L\circ L=L$ y $M\circ M=\mathrm{id}$.*

Los homomorfismos biyectivos se llaman **isomorfismos**. Vimos que si $\dim(\mathbb{E})=n$ entonces $\mathbb{E}$, tras elegir una base $\mathcal{U}$, es isomorfo a $\Bbbk^n$ por medio de $x\to\mathrm{mat}^{\mathcal{U}}(x)$. Un isomorfismo de $\mathbb{E}$ en sí mismo es un **automorfismo**.

Problema 115 *Pedimos probar o estudiar las cuestiones propuestas.*

1. *Si $a \in \Bbbk^{n\times n}$ es invertible, $A:\Bbbk^n \to \Bbbk^n$ dada por $A(x) = ax$ es un automorfismo de $\mathbb{E} = \Bbbk^n$.*
2. *Si $L:\mathbb{E}\to\mathbb{F}$ es un isomorfismo, $L^{-1} = M:\mathbb{F}\to\mathbb{E}$ es lineal.*
3. *Determinar en la lista de ejemplos **1-5** qué funciones son isomorfismos. ¿Cuáles son las inversas?*
4. *Sea $\mathbb{E} = \mathbb{F}\times\mathbb{F}$ y $L:\mathbb{E}\to\mathbb{E}$, $L(x,y) = (y,\kappa x)$ para $\kappa\in\Bbbk$ fijo. Determinar los valores de κ (si existen) tales que L sea un isomorfismo. Cuando lo sea, explicitar la inversa.*

3.2. Funciones lineales y matrices

Aparte de las fórmulas "sin coordenadas" de funciones lineales vamos a ver ahora que en espacios de dimensión finita se pueden obtener funciones lineales con solo conocer los valores de L sobre una base, y también fijarlas de modo único si se conocen los valores de ellas sobre generadores.

Teorema 51 *Sean $(a_1,\dots,a_p)$ una sucesión generadora y $\mathcal{U} = (u_1,\dots,u_n)$ una base de $\mathbb{E}$. Entonces*

1. *Si L, M son funciones lineales de $\mathbb{E}$ en $\mathbb{F}$ tales que $L(a_j) = M(a_j)$, $j = 1,\dots,p$, se tiene, $L = M$.*
2. *Dada una sucesión $(b_1,\dots,b_n)$ de vectores de otro espacio $\mathbb{F}$ de longitud $n = \dim(\mathbb{E})$ (pero con posibles repeticiones), hay una función lineal $L:\mathbb{E}\to\mathbb{F}$ y solo una cumpliendo $L(u_i) = b_i$, $i = 1,\dots,n$. La fórmula de L es*

$$si\ x = \sum_{i=1}^{n} x^i u_i \in \mathbb{E},\ entonces\ L(x) = \sum_{i=1}^{n} x^i b_i \tag{3.2}$$

Demostración. Cada $x\in\mathbb{E}$ se puede escribir al menos de una forma como $x = \sum_{i=1}^{p}\lambda^i a_i$. Entonces

$$L(x) = L\left(\sum_{i=1}^{p}\lambda^i a_i\right) \overset{1}{=} \sum_{i=1}^{p}\lambda^i L(a_i), \qquad M(x) = M\left(\sum_{i=1}^{p}\lambda^i a_i\right) \overset{1}{=} \sum_{i=1}^{p}\lambda^i M(a_i)$$

y las condiciones $L(a_j) = M(a_j)$ dan $L(x) = M(x)$. En $\overset{1}{=}$ se ha usado la linealidad.

Definimos L por (3.2), cosa posible y correcta pues las coordenadas $\left(x^1,\dots,x^n\right)$ vienen unívocamente fijadas por x. Ha de verse que L es lineal. Si tenemos $x = \sum_{i=1}^{n} x^i u_i$ e $y = \sum_{i=1}^{n} y^i u_i$, se sigue que

$$L(x+y) = L\left(\sum_{i=1}^{n}\left(x^i+y^i\right)u_i\right) = \sum_{i=1}^{n}\left(x^i+y^i\right)b_i, \quad L(x)+L(y) = \sum_{i=1}^{n}x^i b_i + \sum_{i=1}^{n}y^i b_i = \sum_{i=1}^{n}\left(x^i+y^i\right)b_i$$

y $L(x+y) = L(x)+L(y)$. De modo análogo, $L(\lambda x) = \lambda L(x)$. Las coordenadas de u_h son

$$\left(0,\dots,0,\overset{(h)}{1},0,\dots,0\right), \quad \text{o de otro modo,} \quad u_h^r = \delta_h^r,\ r = 1,\dots,n.$$

Como consecuencia, $L(u_h) = \sum_{i=1}^{n}\delta_h^i b_i = b_h$. Esto prueba que hay al menos una L lineal con las condiciones dichas. La unicidad (hay una como máximo) se sigue de **1**. ♣

Si es por ejemplo $\mathbb{E} = \mathbb{R}^2$ y $\mathcal{U} = \mathcal{E}$, la base estándar, al elegir en $\mathbb{F}$, el espacio de las funciones de $\mathbb{R}$ en $\mathbb{R}$, $b_1 = \cos t$, $b_2 = \sin t$, tenemos una función lineal L tal que $L(e_1) = \cos t$, $L(e_2) = \sin t$ y, por ejemplo, $L(1,-3)^\top = L(e_1 - 3e_2) = \cos t - 3\sin t$.

En el teorema no se presupone que $\mathbb{F}$ tenga una base, pero si la tiene, podemos asociar a L una matriz en un teorema esencial. Quizás el lector prefiera ver primero los comentarios y ejemplos que van tras la demostración antes que la propia demostración, que es directa pero muy formal.

Teorema 52 *Sean $\mathcal{U} = (u_1,\dots,u_n)$ y $\mathcal{V} = (v_1,\dots,v_m)$ bases de $\mathbb{E}$ y $\mathbb{F}$, y $L:\mathbb{E}\to\mathbb{F}$ lineal definida por las condiciones $L(u_j) = b_j$, $j = 1,\dots,n$. Si los vectores $L(u_j) = b_j$ se expresan en la base $\mathcal{V}$ por las ecuaciones*

$$b_j = L(u_j) = L_j^1 v_1 + L_j^2 v_2 + \ldots + L_j^m v_m = \sum_{i=1}^{m} L_j^i v_i\ , \quad 1\le j\le n,$$

la matriz, que llamaremos la ***matriz de*** L ***para las bases*** $\mathcal{U}$ *y* $\mathcal{V}$,

$$a = \begin{pmatrix} L_1^1 & L_2^1 & \cdots & L_{n-1}^1 & L_n^1 \\ L_1^2 & L_2^2 & \cdots & L_{n-1}^2 & L_n^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ L_1^{m-1} & L_2^{m-1} & \cdots & L_{n-1}^{m-1} & L_n^{m-1} \\ L_1^m & L_2^m & \cdots & L_{n-1}^m & L_n^m \end{pmatrix}$$

determina las coordenadas en $\mathcal{V}$ *de* $L(x)$ *en función de las de* x *en* $\mathcal{U}$ *por la fórmula*

$$\begin{pmatrix} L(x)^1 \\ L(x)^2 \\ \vdots \\ L(x)^{m-1} \\ L(x)^m \end{pmatrix} = \begin{pmatrix} L_1^1 & L_2^1 & \cdots & L_{n-1}^1 & L_n^1 \\ L_1^2 & L_2^2 & \cdots & L_{n-1}^2 & L_n^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ L_1^{m-1} & L_2^{m-1} & \cdots & L_{n-1}^{m-1} & L_n^{m-1} \\ L_1^m & L_2^m & \cdots & L_{n-1}^m & L_n^m \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^{n-1} \\ x^n \end{pmatrix}. \tag{3.3}$$

Demostración. Dado $x = \sum_{j=1}^n x^j u_j \in \mathbb{E}$ tenemos que $\left(x^1, \ldots, x^n\right)^\top$ es la matriz de coordenadas de x en $\mathcal{U}$. Por (3.2),

$$L(x) = \sum_{j=1}^n x^j b_j = \sum_{j=1}^n x^j \left(\sum_{i=1}^m L_j^i v_i \right) = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} x^j L_j^i v_i = \sum_{i=1}^m \left(\sum_{j=1}^n L_j^i x^j \right) v_i,$$

con las reglas usuales de manipulación de sumatorios. Las $\mathcal{V}$-coordenadas de $L(x)$ son los coeficientes de las v_i; o sea

$$L(x)^i = \sum_{j=1}^n L_j^i x^j, \qquad 1 \le i \le m. \tag{3.4}$$

Estas ecuaciones con sumatorios equivalen a la ecuación matricial (3.3). ♣

Este teorema se usa en los dos sentidos. A veces tenemos la expresión de L "sin coordenadas" y se busca la matriz para tener L "con coordenadas" y otras veces se hace al revés. *Es fundamental para escribir la matriz de* L *el observar que la columna* j *está formada por las coordenadas de* $L(u_j)$ *en la base* $\mathcal{V}$*, y por tanto lo natural es escribir la matriz de* L *por columnas y no por filas.*

Por ejemplo, si $\mathbb{E} = \mathbb{F} = \mathbb{R}_2[X]$, los polinomios o funciones polinomiales de grado ≤ 2, y $\mathcal{U} = \mathcal{V} = \mathcal{E}$ la base estándar, calculamos la matriz de $L = D$, la derivada, con

$$\begin{cases} D(1) = 0 = 0 \cdot 1 + 0 \cdot X + 0 \cdot X^2 \\ D(X) = 1 = 1 \cdot 1 + 0 \cdot X + 0 \cdot X^2 \\ D\left(X^2\right) = 2X = 0 \cdot 1 + 2 \cdot X + 0 \cdot X^2 \end{cases} \text{ y la matriz es } a = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Por supuesto, si cambian las bases, cambia la matriz. Mantengamos $\mathcal{V} = \mathcal{E}$ y L pero sea, aceptando que es base, $\mathcal{U} = \left(1, 1+2X, 1+2X+3X^2\right)$. La nueva matriz b de $L = D$ se calcula por

$$\begin{cases} D(1) = 0 = 0 \cdot 1 + 0 \cdot X + 0 \cdot X \\ D(1+2X) = 2 = 2 \cdot 1 + 0 \cdot X + 0 \cdot X^2 \\ D\left(1+2X+3X^2\right) = 2 + 6X = 2 \cdot 1 + 6 \cdot X + 0 \cdot X^2 \end{cases} \text{ y la matriz es } b = \begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix}.$$

Recordamos que para $x \in \mathbb{E}$ con la base $\mathcal{U} = (u_1, \ldots, u_n)$, se representa por $\mathrm{mat}^{\mathcal{U}}(x)$ a la matriz en $\Bbbk^n$ dada por las coordenadas $x^1, \ldots, x^n$ de x en $\mathcal{U}$. La **matriz de** L **para las bases** $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$, denotada por a en el teorema, se denotará también por $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$. Esta notación es fea pero ayuda a recordar fórmulas como (3.3), que se reescribe

$$\mathrm{mat}^{\mathcal{V}}(L(x)) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) \cdot \mathrm{mat}^{\mathcal{U}}(x). \tag{3.5}$$

Es importante observar que se elige la posición de $\mathcal{U}$ y $\mathcal{V}$ para que funcione una especie de "ley de cancelación de fracciones" $V = \frac{V}{U} U$. Si volvemos al ejemplo de $L = D : \mathbb{R}_2[X] \to \mathbb{R}_2[X]$ con las bases $\mathcal{U} = \left(1, 1+2X, 1+2X+3X^2\right)$ y $\mathcal{V} = \mathcal{E}$ y queremos calcular "en coordenadas" $L\left(X^2\right)$ hay que empezar por conocer las coordenadas de X^2 en $\mathcal{U}$, que se ven enseguida pues

$$X^2 = \frac{1}{3}\left((1+2X+3X) - (1+2X)\right) = \frac{1}{3}(u_3 - u_2), \text{ luego son } (0, -1/3, 1/3),$$

y las coordenadas de $L\left(X^2\right)$ en la base estándar son

$$\begin{pmatrix} 0 & 2 & 2 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -\frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \text{ equivalente a } \; L\left(X^2\right) = 2X.$$

Por supuesto, para saber que la derivada de X^2 es $2X$ nadie se complica la vida metiendo en liza a la extraña base $\mathcal{U}$, pero aquí se trata de ver cómo, si se necesita otra base que no sea la estándar, se pueden realizar cálculos. Luego daremos procedimientos más sistemáticos.

Teorema 53 *Sean $\mathcal{U}$ y $\mathcal{V}$ bases respectivas de $\mathbb{E}$ y $\mathbb{F}$. Entonces,*

1. *Si $L, M : \mathbb{E} \to \mathbb{F}$ son funciones lineales y $\lambda \in \Bbbk$ se tiene*

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L+M) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) + \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(M), \qquad \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\lambda L) = \lambda\, \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L). \tag{3.6}$$

2. *Si $L : \mathbb{E} \to \mathbb{F}$ y $M : \mathbb{F} \to \mathbb{G}$ son funciones lineales y $\mathcal{W} = (w_1, \ldots, w_p)$ es base de $\mathbb{G}$ se tiene que*[4]

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(M \circ L) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(M) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L). \tag{3.7}$$

Demostración. Hacemos **2** que es lo más complejo. Sean $b = \mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(M)$ y $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$. Debemos probar que $b \cdot a$ es la matriz c de $M \circ L$ para las bases $\mathcal{U}$ (inicial) y $\mathcal{V}$ (final). Calculamos

$$M(L(u_j)) \overset{1}{=} M\left(\sum_{i=1}^{m} a_j^i v_i\right) \overset{2}{=} \sum_{i=1}^{m} a_j^i M(v_i) \overset{1}{=} \sum_{i=1}^{m} a_j^i \left(\sum_{k=1}^{p} b_i^k w_k\right) \overset{3}{=} \sum_{i=1}^{m}\sum_{k=1}^{p} a_j^i b_i^k w_k$$

con sustituciones de datos o definiciones en $\overset{1}{=}$, uso de la linealidad en $\overset{2}{=}$, y del axioma de distributividad en $\overset{3}{=}$. Recordemos que j está fijo luego hay una suma de productos (de hecho, vectores) $P_{ik} = a_j^i b_i^k w_k$. Imaginémoslos en una tabla de modo que en la fila i y columna k está P_{ik}. El orden de los sumatorios indica que para cada i se suma en k y tenemos la suma de la *fila i*; viniendo a continuación la suma de todas las *filas*. Lo mismo da, y es lo esencial en la demostración, hacer primero la suma de cada *columna* k y luego sumar todas las columnas. Esto es lo que justifica $\overset{4}{=}$ en la parte final del cálculo

$$M(L(u_j)) = \sum_{i=1}^{m}\sum_{k=1}^{p} a_j^i b_i^k w_k = \sum_{i=1}^{m}\sum_{k=1}^{p} P_{ik} \overset{4}{=} \sum_{k=1}^{p}\sum_{i=1}^{m} P_{ik} = \sum_{k=1}^{p}\left(\sum_{i=1}^{m} b_i^k a_j^i\right) w_k \overset{1}{=} \sum_{k=1}^{p} (b \cdot a)_j^k\, w_k.$$

Comparando con $M(L(u_j)) = \sum_{k=1}^{p} c_j^k w_k$ queda $c_j^k = (b \cdot a)_j^k$ como queríamos demostrar. ♣

Problema 116 *Calcular la matriz de la derivada $D : \mathbb{R}_n[X] \to \mathbb{R}_n[X]$ con las bases $\mathcal{U}$ y $\mathcal{V}$ iguales a la estándar $\mathcal{E} = \left(1, X, X^2, \ldots, X^n\right)$.*

Problema 117 *Para la descomposición $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ consideramos la proyección P y la simetría S basadas en $\mathbb{F}$ y una base $\mathcal{U} = (u_1, \ldots u_n, u_{n+1}, \ldots, u_{n+p})$ tal que $(u_1, \ldots u_n)$ es base de $\mathbb{F}$ y $(u_{n+1}, \ldots, u_{n+p})$ es base de $\mathbb{G}$. Calcular las matrices de P y S en esta base.*

Problema 118 *Para $\mathbb{E} = \mathbb{C}_n[X]$ los polinomios complejos de grado $\leq n$ consideramos $L : \mathbb{C}_n[X] \to \mathbb{C}^2$ dada por $L(f(X)) = (f(0), f(i))$. Calcular para $\mathcal{U}$ y $\mathcal{V}$ bases estándar de $\mathbb{C}_n[X]$ y $\mathbb{C}^2$ la matriz de L.*

Problema 119 *Sea $f : \mathbb{E} \to \Bbbk$ una forma (o función) lineal y $\mathcal{U} = (u_1, \ldots u_n)$ una base de $\mathbb{E}$. Si en $\mathbb{F} = \Bbbk$ se toma la base estándar, probar que la matriz de f con estas bases es $(f(u_1), f(u_2), \ldots, f(u_n)) \in \Bbbk^{1\times n}$.*

Problema 120 *Sean $a \in \Bbbk^{n\times n}$ y $L(x) = ax$ función lineal asociada a a. Probar que a es triangular superior si y solo si para la base estándar de $\Bbbk^n$, $L(e_j) = \sum_{i=1}^{j} a_j^i e_i$. ¿Cuál es la proposición que corresponde a "triangular inferior"?*

[4] Verbalmente: la matriz de la *composición* de funciones es el *producto* de sus matrices. Para que esta fórmula tan útil sea cierta es para lo que hay que definir el producto de matrices con la fórmula que inicialmente pudo parecer extraña.

3.3. Rangos y dimensiones

Se definió el rango de L como $\dim(\operatorname{im}(L))$. Lo relacionamos con el rango de $\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$.

Teorema 54 *Sea $L:\mathbb{E}\to\mathbb{F}$ lineal y las bases $\mathcal{U}=(u_1,\ldots,u_n)$ y $\mathcal{V}=(v_1,\ldots,v_m)$ de $\mathbb{E}$ y $\mathbb{F}$. Entonces*

$$\operatorname{rg}(L)=\operatorname{rg}\left(\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)\right) \tag{3.8}$$

Demostración. Tenemos un isomorfismo Φ entre $\mathbb{F}$ y $\Bbbk^m$ dado por la base $\mathcal{V}$ que lleva cada $y=\sum_{i=1}^m y^i v_i$ a $\left(y^1,\ldots,y^m\right)^\top\in\Bbbk^m$. Claramente para $x=\sum_{j=1}^n x^j u_j\in\mathbb{E}$,

$$L(x)=\sum_{j=1}^n x^j L(u_j)\quad\text{ por lo que }\quad \operatorname{im}(L)=\operatorname{lg}(L(u_1),\ldots,L(u_n)).$$

Conocer la dimensión de $\operatorname{lg}(L(u_1),\ldots,L(u_n))\subset\mathbb{F}$, que es el rango por definición, es conocer la dimensión en $\Bbbk^m$ de $\operatorname{lg}(\Phi(L(u_1)),\ldots,\Phi(L(u_n)))$. Ya comentamos que $\Phi(L(u_j))$ es la columna j de la matriz $\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$, luego

$$\operatorname{rg}(L)=\dim\operatorname{lg}(L(u_1),\ldots,L(u_n))=\dim\operatorname{lg}(\Phi(L(u_1)),\ldots,\Phi(L(u_n)))=\operatorname{rg}\left(\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)\right),$$

como queríamos demostrar. ♣

Este teorema nos abre muchas posibilidades para calcular el rango de L. Lo primero (es fácil pasarlo por alto) es que la definición de $\operatorname{rg}(L)$ como $\dim(\operatorname{im}(L))$ no hace referencia a bases, luego es independiente de las que se puedan elegir para su cálculo. Buscaremos las bases para que en (3.8) el cálculo del rango matricial sea lo más sencillo posible. El teorema 54 dice que un objeto geométrico, el "tamaño" de $\operatorname{im}(L)$, se puede conseguir con cálculos algebraicos sobre matrices. Sin embargo, se puede recorrer el camino inverso porque si se quieren encontrar relaciones entre rangos de matrices a y b se pueden obtener estudiando las imágenes de $A(x)=ax$ y $B(x)=bx$, las funciones lineales asociadas a las matrices a y b. Luego veremos ejemplos de esto.

Problema 121 *Calcular los rangos de las funciones de los problemas 116-119.*

El teorema que sigue es de los más importantes al tratar rangos y dimensiones.

Teorema 55 (del rango-nulidad) *Sea $L:\mathbb{E}\to\mathbb{F}$ con $\dim(\mathbb{E})=n$. Entonces*

$$n=\dim(\mathbb{E})=\dim(\ker(L))+\dim(\operatorname{im}(L))=\operatorname{nul}(L)+\operatorname{rg}(L). \tag{3.9}$$

Demostración. Tomamos una base $\mathcal{K}=(u_1,\ldots,u_k)$ que, por el teorema 37 puede ampliarse hasta una base $\mathcal{U}=(u_1,\ldots,u_k,u_{k+1},\ldots,u_n)$de $\mathbb{E}$. Afirmamos que $\mathcal{L}=(L(u_{k+1}),\ldots,L(u_n))$ es una base de $\operatorname{im}(L)$. Si admitimos esto provisionalmente es fácil concluir porque $\dim(\operatorname{im}(L))=\operatorname{rg}(L)=n-k=n-\dim(\ker(L))$.

Veamos que $\mathcal{L}$ es una base de $\operatorname{im}(L)$. Sea $y=L(x)\in\operatorname{im}(L)$, Escribimos con $\mathcal{U}$

$$x=\sum_{i=1}^k x^i u_i+\sum_{j=k+1}^n x^j u_j,\qquad y=L(x)=\sum_{i=1}^k x^i L(u_i)+\sum_{j=k+1}^n x^j L(u_j)=\sum_{j=k+1}^n x^j L(u_j),$$

ya que al estar $u_1,\ldots,u_k$ en $\ker(L)$, serán $L(u_1)=\ldots=L(u_k)=0$, y queda y como combinación lineal de $\mathcal{L}$ con coeficientes $\left(x^{k+1},\ldots,x^n\right)$. Queda mostrar la independencia de $\mathcal{L}$. Supongamos que sea

$$0=\sum_{j=k+1}^n \lambda^j L(u_j),\quad\text{que da}\quad 0=L\left(\sum_{j=k+1}^n \lambda^j u_j\right),\quad \sum_{j=k+1}^n \lambda^j u_j\in\ker(L).$$

Al ser $\mathcal{K}$ base de $\ker(L)$ existirán $\mu^1,\ldots,\mu^k$ de modo que

$$\sum_{j=k+1}^n \lambda^j u_j=\sum_{i=1}^k \mu^i u_i,\qquad \sum_{i=1}^k \mu^i u_i-\sum_{j=k+1}^n \lambda^j u_j=0.$$

Hay pues una combinación de $\mathcal{U}$ que da 0, luego $\mu^1 = \ldots = \mu^k = \lambda^{k+1} = \ldots = \lambda^n = 0$. ♣

La aplicación más directa se tiene para cuando L está asociada a la matriz $a \in \Bbbk^{m\times n}$; o sea, $L : \Bbbk^n \to \Bbbk^m$ es $L(x) = ax$. En tal caso, $\ker(L)$ es el espacio de soluciones del sistema homogéneo $ax = 0$ e $\operatorname{im}(L)$ es el espacio de columnas de la matriz. La fórmula (3.9) dice que n, el número de incógnitas del sistema $ax = 0$, es la suma de la dimensión del espacio de soluciones (claramente la nulidad) y del rango de L, que es el de a. Se reencuentra la fórmula "la dimensión del espacio de soluciones es igual al número de incógnitas menos el rango de la matriz" derivada de $\operatorname{nul}(L) = \dim(\mathbb{E}) - \operatorname{rg}(L) = n - \operatorname{rg}(a)$. Hay algo añadido y es que antes $\operatorname{rg}(a)$ era solo calculable por un procedimiento (transformarla en b escalonada) pero ahora hay más, aunque en muchos casos el primer procedimiento sea el más efectivo.

Problema 122 *Dar un ejemplo de un sistema de 4 ecuaciones con 7 incógnitas $ax = 0$ de manera que de los 28 coeficientes de a sean no nulos y con un espacio de soluciones de dimensión 5. (¡No hay que resolverlo!)* ♦

Dar un ejemplo de un sistema de m ecuaciones con m arbitrario pero solo dos incógnitas tal que siempre que haya solución, esta sea única. No se permiten coeficientes cero para las incógnitas.

Solución. Tomamos

$$a = \begin{pmatrix} 1 & 9 & 1 & 2 & 3 & 4 & 5 \\ 9 & 1 & 1 & 2 & 3 & 4 & 5 \\ 10 & 10 & 2 & 4 & 6 & 8 & 10 \\ 19 & 11 & 3 & 6 & 9 & 12 & 15 \end{pmatrix}.$$

Las filas a^1 y a^2 son independientes y se tiene $a^3 = a^1 + a^2$ y $a^4 = a^1 + 2a^2$. Por tanto $\lg\left(a^1, a^2, a^3, a^4\right) = \lg\left(a^1, a^2\right)$ y $\operatorname{rg}(a) = 2$. Para $L(x) = ax$ se tiene $2 = \operatorname{rg}(a) = \dim(\operatorname{im}(L))$ y el espacio de soluciones de $ax = 0$ es $\ker(L)$ que tendrá dimensión $7 - 2 = 5$. ♦

Problema 123 *Sea $\mathbb{E} = \mathbb{R}_n[X]$ el espacio de los polinomios de grado $\leq n$. Definimos*

$$f : \mathbb{R}_n[X] \to \mathbb{R}, \quad f(P(X)) = \int_0^1 P(X)\, dX$$

y $\mathbb{F} = \{P(X) \mid f(P(X)) = 0\}$. Probar que $\mathbb{F}$ es un subespacio calculando su dimensión y una base. ♦

Solución. Es fácil ver que f es lineal y $\mathbb{F} = \ker(f)$. Como $\operatorname{im}(f) = \mathbb{R}$, (3.9) nos da $\dim(\mathbb{F}) = (n+1) - 1 = n$. Conseguir una base da más trabajo. Tomamos $\mathcal{E} = (1, X, \ldots, X^n)$ la base estándar de $\mathbb{R}_n[X]$ y (1) como base estándar de $\Bbbk$. Vimos en el problema 119 que la matriz de f en esas bases es

$$\left(f(1), f(X), f\left(X^2\right), \ldots, f\left(X^n\right)\right) = \left(\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n+1}\right) = a$$

porque $f\left(X^k\right) = \left[X^{k+1}/(k+1)\right]_{X=0}^{X=1}$. Nos están pidiendo una base del espacio de soluciones del sistema con una única ecuación y $n+1$ incógnitas $\xi^0, \xi^1, \ldots, \xi^n$, dado por

$$\frac{1}{1}\xi^0 + \frac{1}{2}\xi^1 + \frac{1}{3}\xi^2 + \ldots + \frac{1}{n+1}\xi^n = 0, \quad \xi^0 = -\frac{1}{2}\xi^1 - \frac{1}{3}\xi^2 - \ldots - \frac{1}{n+1}\xi^n$$

con variables libres $\xi^1, \ldots, \xi^n$. La solución general es

$$\begin{pmatrix} \xi^0 \\ \xi^1 \\ \vdots \\ \xi^n \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\xi^1 - \frac{1}{3}\xi^2 - \ldots - \frac{1}{n+1}\xi^n \\ \xi^1 \\ \vdots \\ \xi^n \end{pmatrix} = \xi^1 \begin{pmatrix} -\frac{1}{2} \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \xi^n \begin{pmatrix} -\frac{1}{n+1} \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Los polinomios $B_1(X) = -\frac{1}{2} + X$, $B_2(X) = -\frac{1}{3} + X^2$, …, $B_n(X) = -\frac{1}{n+1} + X^n$ sirven como base de $\mathbb{F}$. Hay que observar que si bien $\dim(\mathbb{F}) = n$ tiene respuesta inmediata, construir la base es laborioso. ♦

Problema 124 *Hacer un problema similar con $g : \mathbb{R}_n[X] \to \mathbb{R}$ dada por $g(P(X)) = P(1)$.*

El teorema 55 no pide que $\dim(\mathbb{F})$ sea finita, pero si hay bases $\mathcal{U}=(u_1,\ldots,u_n)$ y $\mathcal{V}=(v_1,\ldots,v_m)$ en $\mathbb{E}$ y $\mathbb{F}$ hemos visto en el teorema 54 que $\dim(\operatorname{im}(L))$ es el rango de $\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$. Sabemos cómo calcular este rango y, con (3.9), tenemos la dimensión de $\ker(L)$. Puede darse el caso de que sea más fácil calcular $\dim\ker(L)$ y entonces se derivaría el valor de $\dim\operatorname{im}(L)$.

Teorema 56 *Sea $L:\mathbb{E}\to\mathbb{F}$ función lineal entre dos espacios* de la misma dimensión n. *Son equivalentes las propiedades, que nos permiten decir, si se dan, que L es un isomorfismo.*

1. *L es inyectiva.*
2. *L es suprayectiva.*
3. *L es biyectiva.*
4. *Existen bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$ en las que la matriz a de L tiene rango n.*
5. *Para todo par de bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$, la matriz a de L tiene rango n.*
6. *L lleva toda base $\mathcal{U}$ de $\mathbb{E}$ en otra de $\mathbb{F}$.*
7. *L lleva alguna base $\mathcal{U}$ de $\mathbb{E}$ en otra de $\mathbb{F}$.*

Demostración. La fórmula 3.9 nos dice que $\dim(\ker(L))=0$ equivale $\dim(\operatorname{im}(L))=n$. Pero $\dim(\ker(L))=0$ equivale a L inyectiva (teorema 49) y $\dim(\operatorname{im}(L))=n$ equivale a $\operatorname{im}(L)=\mathbb{F}$, la suprayectividad de L. Por tanto, **1** equivale a **2** y, como **3** es la conjunción de **1** y **2**, vemos que **1**,**2** y **3** son equivalentes. Se ha visto en el teorema 54 que $\dim(\operatorname{im}(L))$ es $\operatorname{rg}(a)$ con $a=\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$, luego, en particular, $\operatorname{rg}(a)$ no dependerá de las bases elegidas. El que L sea suprayectiva equivale a $\dim(\operatorname{im}(L))=\dim(\mathbb{F})=n$ y por tanto $\mathbf{2}\iff\mathbf{4}$ y $\mathbf{2}\iff\mathbf{5}$. Hasta aquí tenemos las equivalencias $\mathbf{1}\iff\cdots\iff\mathbf{5}$. Es evidente que $\mathbf{6}\implies\mathbf{7}$, luego para concluir basta ver que alguna de las condiciones $\mathbf{1}-\mathbf{5}$ implica **6** y que **7** implica alguna de las condiciones $\mathbf{1}-\mathbf{5}$. Lo hacemos. Primero, $\mathbf{1}\implies\mathbf{6}$. Si $\mathcal{U}=(u_1,\ldots,u_n)$ es base de $\mathbb{E}$, consideramos la combinación

$$0=\sum_{i=1}^{n}\lambda^i L(u_i)=L\left(\sum_{i=1}^{n}\lambda^i u_i\right).$$

Por la inyectividad de L, $\sum_{i=1}^{n}\lambda^i u_i=0$ y, por ser $\mathcal{U}$ independiente, $\lambda^1=\ldots=\lambda^n=0$. Vemos pues que en $\mathbb{F}$ de dimensión n es $(L(u_1),\ldots,L(u_n))$ independiente así que será una base. Acabamos probando $\mathbf{7}\implies\mathbf{4}$. La matriz de L en las bases $\mathcal{U}$ y $L(\mathcal{U})$ es la unidad que tiene rango n y se cumple **4**. ♣

Problema 125 *Sea $\mathbb{E}=\mathbb{R}_n[X]$, los polinomios de grado $\le n$. Nos dan $L:\mathbb{E}\to\mathbb{E}$, $L(P(X))=P(X)+XP'(X)$. Estudiar si L es un isomorfismo y, en caso afirmativo calcular $M=L^{-1}$.* ♦

Solución. Lo natural es calcular la matriz a de L en la base estándar. Se tiene $L(1)=1$ y $L\left(X^k\right)=(1+k)X^k$ luego la matriz a es diagonal con elementos $(1,2,3,\ldots,1+n)$. Como su rango es $n+1=\dim(\mathbb{E})$, L es un isomorfismo y M tendrá matriz $b=a^{-1}$ con diagonal $(1/1,1/2,1/3,\ldots,1/(1+n))$. Si se quiere una fórmula concreta

$$L^{-1}\left(\alpha_0+\alpha_1X+\alpha_2X^2+\ldots+\alpha_nX^n\right)=\frac{\alpha_0}{1}+\frac{\alpha_1}{2}X+\frac{\alpha_2}{3}X^2+\ldots+\frac{\alpha_n}{1+n}X^n.$$

Una idea que no es normal que se le ocurra a un principiante (o no tan principiante) es observar que $D(XP(X))=L(P(X))$ (D es la derivada). Con esto es fácil ver que L es suprayectiva porque si se tiene $Q(X)$ se resuelve $L(P(X))=Q(X)$ integrando $Q(X)$, digamos a $R(X)$, con constante de integración que haga que $R(X)$ tenga la forma $R(X)=\beta_1X+\beta_2X^2+\ldots+\beta_{n+1}X^{n+1}$ y, $XP(X)=R(X)$ se sigue que $P(X)=\beta_1+\beta_2X+\ldots+\beta_{n+1}X^n$. Obsérvese que el teorema nos da, al haber probado la suprayectividad, "gratis" la inyectividad, luego $P(X)+XP'(X)=0$ implica $P(X)=0$. ♦

Problema 126 *En $\mathbb{E}=\mathbb{R}_n[X]$ elegimos $n+1$ elementos* distintos *$t_0,t_1,\ldots,t_n$ de $\mathbb{R}$ y definimos*

$$L:\mathbb{E}\to\mathbb{R}^{n+1},\quad L(P(X))=(P(t_0),P(t_1),\ldots,P(t_n))^\top.$$

Probar que es un isomorfismo.

Este problema es instructivo porque se ve la importancia de contar gracias al teorema 56 con 7 condiciones equivalentes, ya que en este caso unas son mucho más fáciles de comprobar que otras. Si se pretende calcular a, la matriz de L para las bases estándar de $\mathbb{E}$ y $\Bbbk^{n+1}$, se obtiene una matriz invertible pero que cuesta ver de modo directo que lo es. La vía clásica es obtener que $\det(a) \neq 0$, pues es el determinante de Vandermonde (se verá en el capítulo siguiente). Sin embargo, cuesta poco ver que L es inyectiva. ¿Cuántas raíces tiene como máximo un polinomio de grado n?

El siguiente problema tiene un propósito similar al del teorema 56 que es el de dar diversas propiedades para probar que L es inyectiva o suprayectiva. Conviene tomar nota porque es una útil herramienta.

Problema 127 *Sea $L : \mathbb{E} \to \mathbb{F}$ función lineal entre dos espacios de dimensiones n y m. Consideramos* posibles *propiedades de L. Probar que las propiedades **1-4** son equivalentes y lo mismo pasa con las propiedades **5-7.***

1. *L es inyectiva.*
2. *Existen bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$ en las que la matriz a de L tiene rango $n = \dim(\mathbb{E})$.*
3. *Para todo par de bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$ la matriz a de L tiene rango $n = \dim(\mathbb{E})$.*
4. *L lleva vectores independientes en vectores independientes.*
5. *L es suprayectiva.*
6. *Existen bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$ en las que la matriz a de L tiene rango $m = \dim(\mathbb{F})$.*
7. *Para todo par de bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$ y $\mathbb{F}$ la matriz a de L tiene rango $m = \dim(\mathbb{F})$.*
8. *L lleva sucesiones generadoras en sucesiones generadoras.*

Problema 128 *Tomamos $\mathbb{E} = \mathbb{F} = \mathbb{R}^{2\times 2}$ y $L : \mathbb{E} \to \mathbb{E}$ dada por $L(x) = x - \operatorname{tr}(x) I$. (¡Ojo! x representa una matriz.) Recordamos que la traza de una matriz cuadrada es la suma de los términos de la diagonal principal. Estudiar si L es respectivamente inyectiva o suprayectiva.*

El problema que sigue, aunque no lo parezca, es "casi" una generalización del anterior. En un espacio $\mathbb{E}$ de dimensión n tenemos una forma lineal $f : \mathbb{E} \to \Bbbk$ no nula. Sabemos que $H = \ker(f)$ es un hiperplano y tiene dimensión $n - 1$. Consideramos

$$L : \mathbb{E} \to \mathbb{E}, \quad L(x) = x + f(x)\, c, \quad c \in \mathbb{E} \text{ es un vector fijo no nulo.}$$

Si $x \in \mathbb{H}$ se tiene $L(x) = x$ luego L fija el $\mathbb{H}$ y mueve (¿cómo?) los vectores fuera de $\mathbb{H}$.

Problema 129 *Con las notaciones precedentes distinguimos los casos **(1)** $c \notin \mathbb{H}$ y digamos, para simplificar, que $f(c) = 1$ y **(2)** $c \in \mathbb{H}$, luego $f(c) = 0$. Determinar si L es inyectiva o suprayectiva, y L^{-1} si L es biyectiva. Indicación: todo será más fácil con bases que amplíen una de $\mathbb{H}$.*

Hay ciertas fórmulas que relacionan rangos de *dos* funciones lineales. Lo interesante es que requieren ver el rango como dimensión de la imagen y usar cosas tan obvias como que $\mathbb{F} \subset \mathbb{G}$ implica $\dim(\mathbb{F}) \leq \dim(\mathbb{G})$. Las demostraciones no son largas, pero requieren cierto ingenio.

Teorema 57 *Sean $L : \mathbb{E} \to \mathbb{F}$ lineal y $P : \mathbb{E} \to \mathbb{E}$, $Q : \mathbb{F} \to \mathbb{F}$ isomorfismos. Entonces* $\operatorname{rg}(Q \circ L \circ P) = \operatorname{rg}(L)$. *Si $M : \mathbb{F} \to \mathbb{G}$ es otra función lineal,* $\operatorname{rg}(M \circ L) \leq \min\{\operatorname{rg}(L), \operatorname{rg}(M)\}$.

Demostración. Tenemos que $(Q \circ L \circ P)(\mathbb{E}) = (Q \circ L)(\mathbb{E})$ porque $P(\mathbb{E}) = \mathbb{E}$. Por otra parte, un isomorfismo preserva la dimensión de subespacios (es consecuencia fácil del teorema 56). Por tanto,

$$\operatorname{rg}(Q \circ L \circ P) = \dim[(Q \circ L \circ P)(\mathbb{E})] = \dim[(Q \circ L)(\mathbb{E})] = \dim(L(\mathbb{E})) = \operatorname{rg}(L).$$

Para la segunda parte, es obvio que $L(\mathbb{E}) \subset \mathbb{F}$ implica $M(L(\mathbb{E})) \subset M(\mathbb{F})$. Con esto,

$$\operatorname{rg}(M \circ L) = \dim[M(L(\mathbb{E}))] \leq \dim[M(\mathbb{F})] = \operatorname{rg}(M),$$

$$\operatorname{rg}(M \circ L) = \dim[M(L(\mathbb{E}))] \leq \dim[L(\mathbb{E})] = \operatorname{rg}(L),$$

y $\operatorname{rg}(M \circ L) \leq \min\{\operatorname{rg}(L), \operatorname{rg}(M)\}$. ♣

Como consecuencia de este teorema, si tenemos $a \in \Bbbk^{m\times n}$ y hacemos operaciones fila y operaciones columna, la nueva matriz b resultante es $b = qap$ siendo q y p producto de matrices elementales de dimensiones respectivas $m \times m$ y $n \times n$. Si asociamos a q, a, p y b funciones lineales Q, L, P y M, tenemos $M = Q \circ L \circ P$ de donde se sigue que a y b tienen el mismo rango. El poder modificar a tanto con filas como con columnas puede ayudar a calcular su rango.

Problema 130 *Sean $L, M : \mathbb{E} \to \mathbb{F}$ lineales. Probar que* $\operatorname{rg}(L+M) \leq \operatorname{rg}(L) + \operatorname{rg}(M)$.

Problema 131 *Probar que si $L, M : \mathbb{E} \to \mathbb{E}$ y $M \circ L = 0$ (la composición es el homomorfismo cero) entonces* $\operatorname{rg}(L) + \operatorname{rg}(M) \leq \dim(\mathbb{E}) = n$. *Indicación: ¿Relación entre núcleos e imágenes?*

A veces, tratando a como la función $L : \Bbbk^n \to \Bbbk^m$ se obtienen propiedades algebraicas que requieren mucha labor si se trabaja con índices. Damos una ilustración. Si en $\Bbbk^n$ definimos $\mathbb{E}_j = \lg(e_1, \ldots, e_j)$, el subespacio generado por los j primeros elementos de la base estándar, es muy fácil comprobar que el que a sea triangular superior equivale a que se cumpla $L(\mathbb{E}_j) \subset \mathbb{E}_j$ para $j = 1, \ldots, n$.

Problema 132 *Sea $a \in \Bbbk^{n\times n}$ triangular superior e invertible. Probar que $b = a^{-1}$ es también triangular superior. Indicación: en general, si $f : X \to Y$ es biyectiva, la condición $A \subset B$ equivale a $f(A) \subset f(B)$.*

Si el lector toma matrices a, b cuadradas triangulares superiores pero con ceros en la diagonal principal, observará que al multiplicarlas aparecen cada vez más ceros. Por ejemplo, en sucesivas potencias de a, los coeficientes no nulos se agrupan cada vez más en la esquina superior derecha hasta que por fin se llega a una potencia $a^k = 0$. Generalizamos el concepto de matriz triangular superior definiendo que a es **p−triangular superior** si L cumple $L(\mathbb{E}_j) \subset \mathbb{E}_{j-p}$ entendiéndose que $\mathbb{E}_k = 0$ si $k \leq 0$. Las matrices triangulares superiores usuales son 0−triangulares con esta definición.

Problema 133 *Probar que si $a, b \in \Bbbk^{n\times n}$ son respectivamente p y q triangulares, se tiene que ab es $(p+q)-$triangular. Si a es $p-$triangular, ¿cuál es el mínimo k tal que $a^k = 0$?*

Problema 134 *Sea $\mathbb{F}$ y $\mathbb{G}$ subespacios de dimensión finita de $\mathbb{E}$. Definimos $L : \mathbb{F} \times \mathbb{G} \to \mathbb{F} + \mathbb{G}$ por $L(x, y) = x + y$. Probar que* $\ker(L)$ *es isomorfo a $\mathbb{F} \cap \mathbb{G}$ y dar otra demostración de la fórmula de Grassmann.*

3.4. Espacios de funciones lineales

Ya advertimos que muchos espacios vectoriales tienen como vectores objetos que son funciones. La primera parte del teorema que sigue se podía haber dado inmediatamente tras el teorema 50.

Teorema 58 *El conjunto $\mathcal{L}(\mathbb{E}, \mathbb{F})$ de las funciones lineales entre dos espacios vectoriales $\mathbb{E}$ y $\mathbb{F}$ es un subespacio vectorial del espacio de todas las funciones de $\mathbb{E}$ en $\mathbb{F}$.*

Si $\mathbb{E}$ y $\mathbb{F}$ tienen dimensiones finitas n y m, entonces $\mathcal{L}(\mathbb{E}, \mathbb{F})$ tiene dimensión finita nm. Elegidas bases $\mathcal{U}$ y $\mathcal{V}$ en $\mathbb{E}$ y $\mathbb{F}$, la función $\Phi : \mathcal{L}(\mathbb{E}, \mathbb{F}) \longrightarrow \Bbbk^{m\times n}$, $\Phi(L) = \operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$, es un isomorfismo.

Demostración. La primera parte es obvia pues el teorema 50 dice que la suma de funciones lineales y el producto de ellas por escalares siguen siendo lineales.

Las fórmulas (3.6) nos dicen que Φ es una función lineal entre los espacios vectoriales $\mathcal{L}(\mathbb{E}, \mathbb{F})$ y $\Bbbk^{m\times n}$. El segundo espacio tiene dimensión nm, pues la base estándar tiene nm elementos. Si vemos que Φ es biyectiva, tendremos un isomorfismo y por tanto será también $\dim(\mathcal{L}(\mathbb{E}, \mathbb{F})) = nm$. Φ es inyectiva porque $\Phi(L) = 0$ da que L tiene matriz 0, luego es 0. Es también suprayectiva porque dada $a \in \Bbbk^{m\times n}$ definimos L, usando **2** en el teorema 51, por $L(u_1) = \sum_{i=1}^m a_1^i v_i, \ldots, L(u_n) = \sum_{i=1}^m a_n^i v_i$ y este teorema dice que estos datos son suficientes para asegurar la existencia de L. Es inmediato que $\Phi(L) = a$. ♣

El caso particular en que $\mathbb{F} = \Bbbk = \Bbbk^{1\times 1}$ es muy interesante. Se suele denotar $\mathcal{L}(\mathbb{E}, \Bbbk)$ por $\mathbb{E}^*$ y se le llama el **espacio dual (de** $\mathbb{E}$). Sus elementos son las formas lineales $f : \mathbb{E} \to \Bbbk$ (reservaremos para ellas letras tipo f, g, h en lo posible). De momento el teorema 58 nos dice que las formas lineales (un tipo *particular* de función de $\mathbb{E}$ en $\Bbbk$) forman un espacio vectorial con las definiciones "obvias" de $f + g$ y λf.

Por el teorema 58, $\dim(\mathbb{E}^*) = n \cdot 1 = n = \dim(\mathbb{E})$ y podemos identificar, *una vez elegida una base* $\mathcal{U}$ de $\mathbb{E}$, cada $f \in \mathbb{E}^*$ con la matriz

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(f) = (f(u_1), \ldots, f(u_n)) \in \Bbbk^{1\times n}.$$

En $\mathbb{F} = \Bbbk^{1\times 1}$ se toma siempre por defecto la base estándar (1) y es por ello por lo que se pone $\mathrm{mat}_{\mathcal{U}}(f)$ en vez de $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(f)$.[5] Obsérvese que si los vectores se escriben en coordenadas con matrices columna, las formas lineales se escriben con matrices fila.

Hay un tipo muy importante de formas lineales, que son las **funciones coordenadas asociadas a una base** $\mathcal{U}$ de $\mathbb{E}^*$. Hasta ahora hemos hablado de la coordenada j de $x \in \mathbb{E}$ una vez elegida una base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$. La **j-ésima función coordenada (asociada a $\mathcal{U}$)** es la *función* u^j que asigna a $x \in \mathbb{E}$ su coordenada x^j. Por la propia definición $x = \sum_{j=1}^n u^j(x)\, u_j$. Estas funciones coordenadas, que sin duda son funciones de $\mathbb{E}$ en $\Bbbk$, son lineales; es decir, son elementos de $\mathbb{E}^*$. Esto se deriva de modo inmediato de $(x+y)^j = x^j + y^j$ y $(\lambda x)^j = \lambda u^j$ (la coordenada j de la suma es la suma de las coordenadas, etc.) porque se reescribe en la forma $u^j(x+y) = u^j(x) + u^j(y)$ y $u^j(\lambda x) = \lambda u^j(x)$, precisamente las condiciones de linealidad. Con todo esto, a la base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$ le podemos asociar una sucesión $(u^1, \ldots, u^n)$ en $\mathbb{E}^*$. Probaremos enseguida que esta sucesión, que se denotará por $\mathcal{U}^*$ para indicar que deriva de $\mathcal{U}$ y está en $\mathbb{E}^*$, es una base de $\mathbb{E}^*$ (no olvidemos que $\mathbb{E}^*$ es tan espacio vectorial como $\mathbb{E}$), que se llama la **base dual de $\mathcal{U}$**. Esto es muy sencillo y se puede probar de varios modos. El primero es que como $\mathbb{E}^*$ tiene dimensión n, bastaría verificar la independencia. Si se tiene $\sum_{i=1}^n \lambda_i u^i = 0$ (aquí 0 denota la función cero, que asigna a cada $x \in \mathbb{E}$ el escalar $0 \in \Bbbk$) se aplica esta función a cada u_j y se tiene

$$\Bbbk \ni 0 = \left(\sum_{i=1}^n \lambda_i u^i\right)(u_j) \overset{1}{=} \sum_{i=1}^n \lambda_i u^i(u_j) \overset{2}{=} \sum_{i=1}^n \lambda_i \delta_j^i \overset{3}{=} \lambda_j$$

y la independencia lineal. Se justifica $\overset{1}{=}$ porque $(\lambda f + \mu g)(x) = \lambda f(x) + \mu g(x)$; $\overset{2}{=}$ porque la coordenada j de u_i es 1 si $i = j$ y 0 si $i \neq j$ y esto es exactamente la delta de Kronecker δ_j^i; y $\overset{3}{=}$ porque en la suma todos los sumandos son nulos excepto si $i = j$. Otro modo de ver que $(u^1, \ldots, u^n)$ es base es, utilizando de nuevo que $\dim(\mathbb{E}^*) = n$, el mostrar que genera $\mathbb{E}^*$. Esto también es fácil porque los coeficientes de $f \in \mathbb{E}^*$ son los $f(u_j) \in \Bbbk$. Con más detalle, se cumple que $f = \sum_{j=1}^n f(u_j)\, u^j$. En efecto, tanto f como $\sum_{j=1}^n f(u_j)\, u^j$ están en $\mathbb{E}^*$, y dos funciones lineales son la misma si coinciden al evaluarlas en una base (teorema 51). Pues esto sucede porque

$$\left(\sum_{j=1}^n f(u_j)\, u^j\right)(u_i) \overset{1}{=} \sum_{j=1}^n f(u_j)\, u^j(u_i) \overset{2}{=} \sum_{j=1}^n f(u_j)\, \delta_i^j \overset{3}{=} f(u_i).$$

Se justifica $\overset{1}{=}$ porque $(\lambda f + \mu g)(x) = \lambda f(x) + \mu g(x)$ y $\overset{2}{=}$ y $\overset{3}{=}$ como en la demostración de independencia. Se ha probado un teorema de uso continuo.

Teorema 59 *La sucesión de funciones coordenadas $(u^1, \ldots, u^n)$ asociadas a una base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$, es una base de $\mathbb{E}^*$. Las coordenadas de $f \in \mathbb{E}^*$ en esa base son $(f(u_1), \ldots, f(u_n))$; es decir, la matriz de las coordenadas de f en $\mathcal{U}^*$ es precisamente la matriz de $f : \mathbb{E} \to \Bbbk$ para la base $\mathcal{U}$ de $\mathbb{E}$ y la estándar (1) de $\Bbbk$. Los vectores u_j y las formas u^i se relacionan por $u^i(u_j) = \delta_j^i$.*

Es probable que el lector esté un poco liado con la posición de los índices, que unas veces son subíndices en $(u_1, \ldots, u_n)$, otras son superíndices en $(x^1, \ldots, x^n)$ y $(u^1, , \ldots, u^n)$ y, como las coordenadas de f en $\mathcal{U}^*$ son $f_j = f(u_j)$, tendremos f con coordenadas $(f_1, \ldots, f_n)$ que con subíndices. Esto parece peligroso porque visto aisladamente $(H^1, \ldots, H^n)$ sin saber qué es H, no podedems afirmar si H^j es la coordenada j de un vector H de $\mathbb{E}$ (y tendríamos una sucesión de n escalares) o el elemento j de la base dual de otra $(H_1, \ldots, H_n)$ de $\mathbb{E}$ (y tendríamos una sucesión de formas). Inevitablemente se sale de dudas por el contexto. La elección de la posición de índices, como ya dijimos a principio del curso, es para tener la posición de índices que permite recordar fórmulas donde aparecen sumas de productos. Se busca colocar

[5] Este es un caso en el que la propia simplicidad complica. Una base es una sucesión $(u_1, \ldots, u_n)$ con $u_j \in \mathbb{E}$. Si $\mathbb{E} = \Bbbk$ una base sería una sucesión con un único elemento u_1, que a su vez sería un escalar. La base estándar de $\mathbb{E} = \Bbbk$ sería pues (1) y en esa base, $x \in \mathbb{E} = \Bbbk$ tiene el doble rol de vector y número y $x = x \cdot 1$; luego la única coordenada del *vector* x en $\mathcal{E}$ es el *escalar* x.

los índices en los factores de modo que el índice de suma aparezca en un factor como subíndice y en otro como superíndice. Por ejemplo,

$$\sum_{j=1}^{n} a_j^i x^j \text{ se prefiere a } \sum_{j=1}^{n} a_{ij} x_j \text{ porque se podría confundir con } \sum_{j=1}^{n} a_{ji} x_j.$$

Esto que se conoce como convención de Einstein, que no se adopta en general en el Álgebra Lineal pero es imprescindible en el Cálculo Tensorial donde la profusión de índices es abrumadora.

Damos unos problemas de puro cálculo. Si el lector prefiere saber cuánto antes por qué se introduce el espacio dual y su relación con los sistemas lineales, tiene una explicación tras estos problemas.

Problema 135 *En el espacio $\mathbb{E} = \mathbb{R}_3[X]$ de polinomios de grado ≤ 3 definimos tres formas lineales*

$$f(P(X)) = P(2), \quad g(P(X)) = P'(3), \quad h(P(X)) = \int_0^1 P(X)\, dX.$$

Dar sus componentes en $\mathcal{E}^*$, la base dual de la estándar, estudiar si son independientes, y, si lo son, añadir una cuarta forma $k \in \mathbb{E}^*$ de modo que (f, g, h, k) sea una base. Dar todos los polinomios $P(X) \in \mathbb{E}$ que cumplan $f(P(X)) = 0$, $g(P(X)) = -1$, $h(P(X)) = 2$. ¿Hay alguno? ¿Hay infinitos? ♦

Solución. Para la base estándar $\mathcal{E} = \left(1, X, X^2, X^3\right)$ tenemos

$$f(1) = 1, \quad f(X) = 2, \quad f\left(X^2\right) = 4, \quad f\left(X^3\right) = 8,$$

$$g(1) = 0, \quad g(X) = 1(3) = 1, \quad g\left(X^2\right) = 2X(3) = 6, \quad g\left(X^3\right) = 3X^2(3) = 27,$$

$$h(1) = [X]_0^1 = 1, \quad h(X) = \left[\frac{X^2}{2}\right]_0^1 = \frac{1}{2}, \quad h\left(X^2\right) = \left[\frac{X^3}{3}\right]_0^1 = \frac{1}{3}, \quad h\left(X^3\right) = \left[\frac{X^4}{4}\right]_0^1 = \frac{1}{4}.$$

De aquí se sigue que las coordenadas son

$$\mathrm{mat}_{\mathcal{U}}(f) = (1, 2, 4, 8), \quad \mathrm{mat}_{\mathcal{U}}(g) = (0, 1, 6, 27), \quad \mathrm{mat}_{\mathcal{U}}(f) = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}\right).$$

Para una posible base es suficiente añadir a la matriz a cuyas tres primeras filas son estas tres matrices una cuarta fila que de una matriz 4×4 invertible. Por ejemplo,

$$a = \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 6 & 27 \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

En efecto, si $\mathrm{rg}(a) = 4$, las tres primeras filas son independientes, y lo es (f, g, h). Admitido provisionalmente que $\mathrm{rg}(a) = 4$, la base (f, g, h, k) de $\mathbb{E}$ tiene como $k \in \mathbb{E}^*$ la forma dada por $k(1) = k(X) = k\left(X^2\right) = 0$ y $k\left(X^3\right) = 1$. Otro modo de expresarlo es con $k\left(\alpha_0 + \alpha_1 X + \alpha_2 X^2 + \alpha_3 X\right) = \alpha_3$. El que $\mathrm{rg}(a) = 4$ sale con operaciones fila Hemos quitado para ahorrar espacio la última fila de a)

$$\begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 6 & 27 \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 6 & 27 \\ 0 & -\frac{3}{2} & -\frac{11}{3} & -\frac{31}{4} \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 6 & 27 \\ 0 & 0 & \frac{16}{3} & \frac{131}{4} \end{pmatrix}$$

Al no haber filas nulas, el rango de la última matriz es 3 y el de a es 4.

Para la última pregunta se ha de resolver el sistema

$$\begin{pmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 6 & 27 \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}.$$

Nos valen las operaciones fila de antes pero han de afectar también a la matriz ampliada,

$$\left(\begin{array}{cccc|c} 1 & 2 & 4 & 8 & 0 \\ 0 & 1 & 6 & 27 & -1 \\ 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 2 \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 2 & 4 & 8 & 0 \\ 0 & 1 & 6 & 27 & -1 \\ 0 & -\frac{3}{2} & -\frac{11}{3} & -\frac{31}{4} & 2 \end{array}\right)$$
$$\to \left(\begin{array}{cccc|c} 1 & 2 & 4 & 8 & 0 \\ 0 & 1 & 6 & 27 & -1 \\ 0 & 0 & \frac{16}{3} & \frac{131}{4} & \frac{7}{2} \end{array}\right) \to \left(\begin{array}{cccc|c} 1 & 2 & 4 & 8 & 0 \\ 0 & 1 & 6 & 27 & -1 \\ 0 & 0 & 1 & \frac{393}{64} & \frac{21}{32} \end{array}\right)$$

El cálculo se va complicando porque solo hemos llegado a escalonar la matriz. La solución general es (seamos honrados, la ha calculado el ordenador)

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} \frac{11}{4} - \frac{25}{8}x^4 \\ -\frac{25}{16} + \frac{315}{32}x^4 \\ \frac{3}{32} - \frac{393}{64}x^4 \\ x^4 \end{pmatrix} = \begin{pmatrix} \frac{11}{4} \\ -\frac{25}{16} \\ \frac{3}{32} \\ 1 \end{pmatrix} + x^4 \begin{pmatrix} -\frac{25}{8} \\ \frac{315}{32} \\ -\frac{393}{64} \\ 1 \end{pmatrix}.$$

Resultan las soluciones al variar t en $\mathbb{R}$. La más sencilla es $P(X) = \frac{11}{4} - \frac{25}{16}X + \frac{3}{32}X^2$ con $x^4 = 0$. ♦

Problema 136 *Planteamos un problema similar al anterior algo menos laborioso. Se toma* $\mathbb{E} = \mathbb{C}_2[X]$, *los de polinomios de grado* ≤ 2 *pero (¡cuidado!),* $\Bbbk = \mathbb{C}$. *Definimos dos formas lineales*

$$f(P(X)) = P(1), \quad g(P(X)) = P(i).$$

Dar sus componentes en $\mathcal{E}^*$ *la base dual de la estándar, estudiar si son independientes, y, en caso afirmativo, añadir una tercera forma* $h \in \mathbb{E}^*$ *de modo que* (f, g, h) *sea una base. Dar todos los polinomios* $P(X) \in \mathbb{E}$ *que cumplan* $f(P(X)) = i$, $g(P(X)) = 1$. *¿Hay alguno? ¿Hay infinitos?*

El espacio dual es, hablando informalmente, el "espacio que se necesita para generalizar las ecuaciones lineales". Sea $\mathbb{E}$ arbitrario, $f^1, \ldots, f^m \in \mathbb{E}^*$ e $y^1, \ldots, y^m \in \Bbbk$. Un **sistema lineal** es un conjunto de ecuaciones $f^1(x) = y^1, \ldots, f^m(x) = y^m$. El sistema es **homogéneo** si $y^1 = \ldots = y^m = 0$. Una **solución del sistema** es un vector $x \in \mathbb{E}$ que verifica las ecuaciones. Es fácil probar que el conjunto de todas las soluciones de un sistema homogéneo es un subespacio vectorial. Los **sistemas compatibles** son los que tienen al menos una solución y los **incompatibles** los que no tienen solución. No hemos supuesto que $\mathbb{E}$ tenga dimensión finita, pero vamos a hacerlo en adelante tomando una base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$ y su base dual $\mathcal{U}^* = \left(u^1, \ldots, u^n\right)$. Si se hace todo con coordenadas

$$f^i(x) = \left(f_1^i u^1 + \ldots + f_n^i u^n\right)(x) = f_1^i u^1(x) + \ldots + f_n^i u^n(x) = f_1^i x^1 + \ldots + f_n^i x^n.$$

El sistema, sea con sumatorios o con matrices toma la forma

$$\left\{\begin{array}{l} \sum_{j=1}^n f_j^1 x^j = f_1^1 x^1 + \ldots + f_n^1 x^n = y^1 \\ \vdots \\ \sum_{j=1}^n f_j^m x^j = f_1^m x^1 + \ldots + f_n^m x^n = y^m \end{array}\right., \qquad \begin{pmatrix} f_1^1 & \cdots & f_n^1 \\ \vdots & \ddots & \vdots \\ f_1^m & \cdots & f_n^m \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix}.$$

Son ecuaciones ya familiares donde la matriz a se ha transformado en la matriz $f = \left(f_j^i\right) \in \Bbbk^{m \times n}$. Los sistemas lineales del principio del curso salen con $\mathbb{E} = \Bbbk^n$ y $\mathcal{U} = \mathcal{E} = (e_1, \ldots, e_n)$ la base estándar. En efecto, en esta situación, $x = \left(x^1, \ldots, x^n\right)$ y $e^p(x) = x^p$ y "no se nota" que haya bases subyacentes. Sabemos cómo resolver el sistema matricial $f(x) = y$ o si es incompatible. Una solución $\left(x^1, \ldots, x^n\right)$ da la matriz de las coordenadas de una solución del sistema abstracto $f^1(x) = y^1, \ldots, f^m(x) = y^m$ en la base $\mathcal{U}$; o sea, $\left(x^1, \ldots, x^n\right) = \mathrm{mat}^{\mathcal{U}}(x)$.

Problema 137 *Sea* $\mathbb{E} = \mathbb{R}_3[X]$. *Determinar los polinomios* $P(X)$ *que verifican*

$$[P(X)]_{X=1} = \left[X^2 P''(X)\right]_{X=1}, \qquad [P(X)]_{X=2} = \left[X^2 P''(X)\right]_{X=2}.$$

Para un polinomio arbitrario se usa la notación $[Q(X)]_{X=\gamma} = Q(\gamma)$ *(valor de* $Q(X)$ *en* $X = \gamma$*). Probar que forman un subespacio* $\mathbb{S}$ *de* $\mathbb{E}$, *determinando una base y su dimensión.* ♦

Solución. Tenemos dos formas lineales

$$f\left(P\left(X\right)\right)=\left[P\left(X\right)-X^{2}P''\left(X\right)\right]_{X=1},\qquad g\left(P\left(X\right)\right)=\left[P\left(X\right)-X^{2}P''\left(X\right)\right]_{X=2}.$$

Y debemos resolver un sistema homogéneo de dos ecuaciones $f\left(P\left(X\right)\right)=g\left(P\left(X\right)\right)=0$ en $\mathbb{E}=\mathbb{R}_{3}\left[X\right]$ que tiene dimensión 4. Para la base estándar $\mathcal{E}$ de $\mathbb{R}_{3}\left[X\right]$ se tienen como matrices de f y g y matriz a del sistema

$$\mathrm{mat}_{\mathcal{U}}\left(f\right)=\left(1,1,-1,-4\right),\quad \mathrm{mat}_{\mathcal{U}}\left(g\right)=\left(2,-2,-4,-40\right),\quad a=\begin{pmatrix}1 & 1 & -1 & -4\\ 2 & -2 & -4 & -40\end{pmatrix}.$$

Resolvemos $f\left(P\left(X\right)\right)=g\left(P\left(X\right)\right)=0$ con operaciones fila

$$\begin{pmatrix}1 & 1 & -1 & -4\\ 2 & -2 & -4 & -40\end{pmatrix}\rightarrow\begin{pmatrix}1 & 1 & -1 & -4\\ 0 & -4 & -2 & -32\end{pmatrix}\rightarrow\begin{pmatrix}1 & 1 & -1 & -4\\ 0 & 1 & \frac{1}{2} & 8\end{pmatrix}\rightarrow\begin{pmatrix}1 & 0 & -\frac{3}{2} & -12\\ 0 & 1 & \frac{1}{2} & 8\end{pmatrix},$$

que dan la solución

$$\begin{pmatrix}x^{0}\\ x^{1}\\ x^{2}\\ x^{3}\end{pmatrix}=\begin{pmatrix}-\frac{3}{2}x^{2}+12x^{3}\\ -\frac{1}{2}x^{2}-8x^{3}\\ x^{2}\\ x^{3}\end{pmatrix}=x^{2}\begin{pmatrix}-\frac{3}{2}\\ -\frac{1}{2}\\ 1\\ 0\end{pmatrix}+x^{3}\begin{pmatrix}12\\ -8\\ 0\\ 1\end{pmatrix}$$

(es conveniente numerar las incógnitas $x^{0},\ldots,x^{3}$ en vez de $x^{1},\ldots,x^{4}$). Hay dos variables libres y el espacio de polinomios dado tiene dimensión 2. Forman base los polinomios cuyas coordenadas en la base estándar son los vectores que van multiplicados por x^{2} y x^{3}; es decir

$$\left(-\frac{3}{2}-\frac{1}{2}X+X^{2},12-8X+X^{3}\right)$$

sirve como base. ♦

Problema 138 *Admitamos que en el espacio de las funciones de $\mathbb{R}$ en $\mathbb{R}$ infinitamente diferenciables, $\left(\cos t,\sin t,\cos^{2}t,\sin^{2}t\right)$ son independientes y sea $\mathbb{E}\subset C^{\infty}\left(\mathbb{R},\mathbb{R}\right)$ el subespacio que generan. Decir si hay funciones $F\in\mathbb{E}$ tales que $F\left(\pi\right)=F\left(2\pi\right)=1$ y, si existen dar la fórmula general de F.*

Problema 139 *Hallar la dimensión y una base del subespacio $\mathbb{S}$ de polinomios $P\left(X\right)$ de grado ≤ 3 que cumplen la condición de que $P\left(X\right)$ y $\left(X+\alpha\right)P'\left(X\right)$ valen lo mismo en $X=\alpha$. Se supone $\alpha\in\mathbb{R}$ prefijado. ¿Depende la respuesta del valor de α?*

Hay una forma alternativa y muy importante de ver los sistemas $f^{1}\left(x\right)=y^{1},\ \ldots\ f^{m}\left(x\right)=y^{m}$.

Teorema 60 *Sea $\mathbb{E}$ de dimensión n y $f^{1},\ldots,f^{m}\in\mathbb{E}^{*}$. Construimos la función lineal*

$$L:\mathbb{E}\rightarrow\Bbbk^{m},\qquad L\left(x\right)=\left(f^{1}\left(x\right),\ldots,\ f^{m}\left(x\right)\right)^{\top}$$

y elegimos una base $\mathcal{U}$ de $\mathbb{E}$. Entonces

1. *La matriz $a=\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}\left(L\right)$ tiene como fila i a $\mathrm{mat}_{\mathcal{U}}\left(f^{i}\right)=\left(f^{i}\left(u_{1}\right),\ldots,f^{i}\left(u_{n}\right)\right)$ y su rango, que es el de L, es también la dimensión del subespacio de $\mathbb{E}^{*}$ que generan $\left(f^{1},\ldots,f^{m}\right)$.*
2. *Decir que $\left(x^{1},\ldots,x^{n}\right)^{\top}$ es solución de $ax=y=\left(y^{1},\ldots,y^{m}\right)^{\top}\in\Bbbk^{m}$ equivale a decir que $x\in\mathbb{E}$ con coordenadas $\mathrm{mat}^{\mathcal{U}}\left(x\right)=\left(x^{1},\ldots,x^{n}\right)$ verifica $L\left(x\right)=y$. En particular, si $y=0$, las soluciones $\left(x^{1},\ldots,x^{n}\right)$ de $ax=0$ dan todas las $\mathrm{mat}^{\mathcal{U}}\left(x\right)$ con $x\in\ker\left(L\right)$.*

Demostración. La columna j de $a=\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}\left(L\right)$, está formada por las coordenadas de $L\left(u_{j}\right)$ en la base $\mathcal{E}$. Claramente a es como se dice porque

$$L\left(u_{j}\right)=\begin{pmatrix}f^{1}\left(u_{j}\right)\\ \vdots\\ f^{m}\left(u_{j}\right)\end{pmatrix},\qquad a=\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}\left(L\right)=\begin{pmatrix}f^{1}\left(u_{1}\right) & \cdots & f^{1}\left(u_{n}\right)\\ \vdots & \ddots & \vdots\\ f^{m}\left(u_{1}\right) & \cdots & f^{m}\left(u_{n}\right)\end{pmatrix}.$$

El rango de L es el de a (teorema 54) y $\dim\left(\lg\left(f^1,\ldots,f^m\right)\right)$ el de una matriz que tenga por filas las $\mathrm{mat}_{\mathcal{U}}\left(f^i\right)$ y para esto nos sirve a, luego $\mathrm{rg}\,(L)=\mathrm{rg}\,(a)=\dim\left(\lg\left(f^1,\ldots,f^m\right)\right)$. El punto **2** es obvio porque se tienen condiciones equivalentes

$$L(x)=y,\ \mathrm{mat}^{\mathcal{E}}(L(x))=\mathrm{mat}^{\mathcal{E}}(y),\ \mathrm{mat}^{\mathcal{E}}_{\mathcal{U}}(L)\,\mathrm{mat}^{\mathcal{U}}(x)=\mathrm{mat}^{\mathcal{E}}(y),\ a\left(x^1,\ldots,x^n\right)^{\top}=\mathrm{mat}^{\mathcal{E}}(y).$$

La última afirmación es obvia también. ♣

El espacio dual nos permite enunciar las condiciones en que dos sucesiones de formas $\left(f^1,\ldots,f^m\right)=\mathbf{f}$ y $\left(g^1,\ldots,g^p\right)=\mathbf{g}$ dan lugar al mismo espacio de soluciones para los sistemas *homogéneos*

$$\mathcal{S}_{\mathbf{f}}: f^1(x)=\ldots=f^m(x)=0 \quad \text{y} \quad \mathcal{S}_{\mathbf{g}}: g^1(x)=\ldots=g^p(x)=0.$$

Teorema 61 *Es necesario y suficiente para que $\mathcal{S}_{\mathbf{f}}$ y $\mathcal{S}_{\mathbf{g}}$ tengan el mismo espacio de soluciones $\mathbb{S}\subset\mathbb{E}$ que las sucesiones $\mathbf{f}$ y $\mathbf{g}$ generen en $\mathbb{E}^*$ el mismo subespacio; o sea,* $\lg\left(f^1,\ldots,f^m\right)=\lg\left(g^1,\ldots,g^p\right)$.

Demostración. Supongamos que $\lg\left(f^1,\ldots,f^m\right)=\lg\left(g^1,\ldots,g^p\right)$. Para cada g^i existen $\sigma^i_j\in\Bbbk$ tales que $g^i=\sum_{j=1}^m\sigma^i_jf^j$. Si x es solución de $\mathcal{S}_{\mathbf{f}}$ debe cumplirse que $f^1(x)=\ldots=f^m(x)=0$, luego $g^i(x)=\sum_{j=1}^m\sigma^i_jf^j(x)=0$. Así pues, $\mathcal{S}_{\mathbf{f}}\subset\mathcal{S}_{\mathbf{g}}$ y de modo simétrico se prueba que $\mathcal{S}_{\mathbf{g}}\subset\mathcal{S}_{\mathbf{f}}$.

Sea ahora $\mathbb{S}$ el espacio común de soluciones de los dos sistemas y $s=\dim(\mathbb{S})$. Por el teorema 60 los dos espacios $\lg\left(f^1,\ldots,f^m\right)$ y $\lg\left(g^1,\ldots,g^p\right)$ tienen la misma dimensión $n-s=r$ con $n=\dim(\mathbb{E})$. Renumerando si hiciera falta las f y g podemos suponer que

$$\lg\left(f^1,\ldots,f^r\right)=\lg\left(f^1,\ldots,f^m\right) \quad \text{y} \quad \lg\left(g^1,\ldots,g^r\right)=\lg\left(g^1,\ldots,g^p\right);$$

o sea, que los r primeros elementos forman base respectiva de $\lg\left(f^1,\ldots,f^m\right)$ y $\lg\left(g^1,\ldots,g^p\right)$.

Se ha visto en el teorema 37 que se puede ampliar una base $(u_1,\ldots,u_s)$ de $\mathbb{S}$ hasta una base $(u_1,\ldots,u_s,u_{s+1},\ldots,u_{s+r})$ de $\mathbb{E}$. Interesa renumerar esta base en la forma

$$\mathcal{V}=(v_1,\ldots,v_r,v_{r+1},\ldots,v_{r+s})=(u_{s+1},\ldots,u_{s+r},u_1,\ldots,u_s)$$

poniendo los $v_k\in\mathbb{S}$ al final, luego $(v_{r+1},\ldots,v_{r+s})$ es base de $\mathbb{S}$. Usamos $\mathcal{V}^*$ para escribir $f^1,\ldots,f^r$ y las evaluamos en los v_k con $k>r$; o sea, los que están en $\mathbb{S}$. Tenemos

$$f^i=\sum_{j=1}^n\lambda^i_jv^j, \qquad 0=f^i(v_k)=\sum_{j=1}^n\lambda^i_jv^j(v_k)=\sum_{j=1}^n\lambda^i_j\delta^j_k=\lambda^i_k.$$

El que sean nulos los λ^i_k para $k>r$ nos da sucesivamente

$$f^i=\sum_{j=1}^r\lambda^i_jv^j \text{ para } 1\leq i\leq r, \quad \lg\left(f^1,\ldots,f^r\right)\subset\lg\left(v^1,\ldots,v^r\right), \quad \lg\left(f^1,\ldots,f^r\right)=\lg\left(v^1,\ldots,v^r\right).$$

Se pasa del contenido a la igualdad porque $\lg\left(f^1,\ldots,f^r\right)$ y $\lg\left(v^1,\ldots,v^r\right)$ tienen igual dimensión r. De modo similar se prueba que $\lg\left(g^1,\ldots,g^r\right)=\lg\left(v^1,\ldots,v^r\right)$. Entonces,

$$\lg\left(f^1,\ldots,f^m\right)=\lg\left(f^1,\ldots,f^r\right)=\lg\left(v^1,\ldots,v^r\right)=\lg\left(g^1,\ldots,g^r\right)=\lg\left(g^1,\ldots,g^p\right)$$

y se acaba la demostración. ♣

Dado el sistema lineal $\mathcal{S}_{\mathbf{f}}: f^1(x)=\ldots=f^m(x)=0$ tenemos la matriz a con $a^i=\mathrm{mat}_{\mathcal{U}}\left(f^i\right)=\left(f^i(u_1),\ldots,f^i(u_n)\right)$, cuyo rango es r. El número r determina el "tamaño" de $\mathbb{S}$, espacio de soluciones, medido como $\dim(\mathbb{S})=n-r$, que es el número mínimo de parámetros imprescindibles para expresar la solución general de $\mathcal{S}_{\mathbf{f}}$. Pero también, y esto es lo que añade el teorema 61, nos da el número mínimo r de formas lineales para expresar $\mathbb{S}$ como espacio de soluciones de $\mathcal{S}_{\mathbf{f}}$.

Problema 140 *En el sistema homogéneo $ax=0$ con*

$$a=\begin{pmatrix}
1 & 2 & 3 & \cdots & n-2 & n-1 & n\\
2 & 3 & 4 & \cdots & n-1 & n & n+1\\
3 & 4 & 5 & \cdots & n & n+1 & n+2\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\
n-2 & n-1 & n & \cdots & n+(n-5) & n+(n-4) & n+(n-3)\\
n-1 & n & n+1 & \cdots & n+(n-4) & n+(n-3) & n+(n-2)\\
n & n+1 & n+2 & \cdots & n+(n-3) & n+(n-2) & n+(n-1)
\end{pmatrix}$$

sobran muchas ecuaciones. ¿Cuántas? ♦

Decir si sobra alguna ecuación en los sistemas

$$\begin{cases} x^1+x^2+\ldots+x^{n-1}+x^n=0 \\ x^2+\ldots+x^{n-1}+x^n=0 \\ \vdots \\ x^{n-1}+x^n=0 \end{cases}, \qquad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

¿Se pueden las ecuaciones por otras sin alterar las soluciones pero con menos ecuaciones? ♠

Solución. Si u es el vector fila $(1,1,,\ldots,1,1)$ (todo unos) queda claro que

$$a^2=a^1+u,\ a^3=a^1+2u,\ldots,a^k=a^1+(k-1)\,u,\ldots,a^n=a^1+(n-1)\,u.$$

Si las formas $f^i:\mathbb{R}^n\to\mathbb{R}$ vienen dadas por las matrices fila a^i; o sea, $f^i(x)=a^ix^\top=\sum_{j=1}^n a^i_jx^j$ y g por la matriz fila u; o sea, $g(x)=x^1+\ldots+x^n$, Se tiene que

$$\lg\left(f^1,\ldots,f^n\right)\subset\lg\left(f^1,g\right)\quad\text{y}\quad\lg\left(f^1,g\right)\subset\lg\left(f^1,\ldots,f^n\right),$$

esto último porque $g=f^2-f^1$. El sistema $ax=0$, equivalente a $f^1(x)=\ldots=f^n(x)=0$ equivale a su vez al sistema con solo dos ecuaciones $f^1(x)=0$, $g(x)=0$. ♦

El problema práctico de determinar para $\Bbbk$ un polinomio $P(X)$ que cumpla $P(\tau_i)=p^i$ $i=0,1,\ldots,n$ se expresa de una forma curiosa en términos de bases y bases duales en $\mathbb{E}=\Bbbk_n[X]$, los polinomios de grado $\leq n$. Sean $\tau^0,\tau^1,\ldots,\tau^n$ elementos *distintos de* $\Bbbk$. Definimos formas lineales t^i en $\mathbb{E}^*$ y elementos $T_j(X)$ (polinomios) en $\mathbb{E}$ por

$$t^i(P(X))=P\left(\tau^i\right)\in\Bbbk,\qquad T_j(X)=\frac{\left(X-\tau^0\right)\cdots\widehat{(X-\tau^j)}\cdots(X-\tau^n)}{\left(\tau^j-\tau^0\right)\cdots\widehat{(\tau^j-\tau^j)}\cdots\left(\tau^j-\tau^n\right)},\quad i,j=0,1,\ldots,n.$$

Observemos que $t^i(T_j(X))=T_j\left(\tau^i\right)=\delta^i_j$, luego si $(T_0(X),T_1(X),\ldots,T_n(X))$ fuera base de $\Bbbk_n[X]$ (enseguida veremos que lo es) tendría a $\left(t^0,t^1,\ldots,t^n\right)$ como base dual. Cuesta poco verlo porque

$$\text{si } 0=\sum_{j=0}^n\lambda^jT_j,\ \text{entonces } 0=\left(\sum_{j=0}^n\lambda^jT_j\right)\left(\tau^i\right)=\sum_{j=0}^n\lambda^jT_j\left(\tau^i\right)=\sum_{j=0}^n\lambda^j{\delta_j}^i=\lambda^i.$$

Se ha probado que los $T_j(X)$ son independientes y, al ser $\dim(\mathbb{E})=\dim(\Bbbk_n[X])=n+1$, forman una base. En cualquier espacio $\mathbb{E}$ con base $\mathcal{U}$ y base dual $\mathcal{U}^*$, los coeficientes de x en $\mathcal{U}$ cumplen $x=\sum_{j=1}^n u^j(x)\,u_j$. En nuestro caso,

$$P(X)=\sum_{j=0}^n t^j(P(X))\,T_j(X)=\sum_{j=0}^n P\left(\tau^j\right)\frac{\left(X-\tau^0\right)\cdots\widehat{(X-\tau^j)}\cdots(X-\tau^n)}{\left(\tau^j-\tau^0\right)\cdots\widehat{(\tau^j-\tau^j)}\cdots\left(\tau^j-\tau^n\right)},$$

que da una descomposición interesante de $P(X)$. Con esto resolvemos el problema inicialmente planteado.

Teorema 62 (interpolación de Lagrange) *Sean* $\tau^0,\tau^1,\ldots,\tau^n$ *elementos* distintos de $\Bbbk$. *Dada una sucesión* $\left(p^0,p^1,\ldots,p^n\right)$ *con posibles repeticiones, hay un polinomio y solo uno de grado* $\leq n$ *que cumple* $P\left(\tau^i\right)=p^i$, $i=0,1,\ldots,,n$ *y es*

$$P(X)=\sum_{j=0}^n p^jT_j(X)=\sum_{j=0}^n p^j\frac{\left(X-\tau^0\right)\cdots\widehat{(X-\tau^j)}\cdots(X-\tau^n)}{\left(\tau^j-\tau^0\right)\cdots\widehat{(\tau^j-\tau^j)}\cdots\left(\tau^j-\tau^n\right)}. \tag{3.10}$$

Demostración. Si un polinomio $P(X)$ cumple $P\left(\tau^i\right)=p^i$, en la base $(T_0(X),T_1(X),\ldots,T_n(X))$ sus coeficientes son $\left(p^0,p^1,\ldots,p^n\right)$, luego, si $P(X)$ existe, es único y ha de ser el dado en (3.10). Por otra parte, este polinomio cumple $P\left(\tau^i\right)=p^i$, $i=0,1,\ldots,,n$ y se tiene probada la existencia. ♣

Los ejemplos son facilísimos de tratar. Dados $1, i, -1$, el polinomio

$$P(X) = 4\frac{(X-i)(X+1)}{(1-i)(1-(-1))} + 4\frac{(X-1)(X+1)}{(i-1)(i+1)} + i\frac{(X-1)(X-i)}{(-1-1)(-1-i)}$$

es el único polinomio complejo de grado ≤ 2 tal que $P(1) = P(i) = 4$ y $P(-1) = i$. Otro ejemplo ilustrativo. Si $P(X) = 2\alpha(X-1)(X-2) - \beta X(X-2) + 2\gamma X(X-1)$ es visto como polinomio de $\mathbb{Z}_3[X]$, su función asociada f lleva $(0,1,2)$ en (α,β,γ) pudiendo haber repeticiones en (α,β,γ).

Al principio de este curso distinguimos entre el polinomio $P(X) = a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ y la función polinómica (o polinomial) $f : \Bbbk \to \Bbbk$, $f(t) = P(t)$, pero dijimos que si $\Bbbk$ era infinito no había razón para distinguir porque cada función polinomial proviene de un único polinomio.

Teorema 63 *La función $\Phi(P(X)) = P(t)$ de $\Bbbk[X]$ en el espacio $\mathbb{P}$ de las funciones polinomiales de $\Bbbk$ en $\Bbbk$ es un isomorfismo cuando $\Bbbk$ es infinito.*

Demostración. Es fácil ver que Φ es lineal y suprayectiva. Si $\Phi(P(X)) = 0$ elegimos $\tau^0, \tau^1, \ldots, \tau^n$ elementos *distintos de* $\Bbbk$, que es posible por grande que sea n. Evidentemente, (3.10) dice que $P(X) = 0$. Nota: La demostración muestra que si $\Bbbk$ tiene al menos n elementos, la función $\Phi_n : \Bbbk_{n-1}[X] \to \Bbbk$ restricción de Φ, es inyectiva. ♣

3.5. Efectos de cambios de base

Nunca hasta ahora hemos manejado a la vez dos bases $\mathcal{U}$ y $\mathcal{V}$ en un mismo espacio $\mathbb{E}$, pero veremos en el futuro con frecuencia situaciones donde los datos se dan usando la base $\mathcal{U}$, y la base $\mathcal{V}$ aparece después como una base "buena" porque informa mejor sobre el problema que se trata o lo expresa con sencillez. Son fórmulas pesadas pero esenciales y donde es fácil equivocarse. En toda la sección $\mathbb{E}$ tendrá dimensión finita n y usaremos (3.7) que copiamos otra vez: *si $L : \mathbb{E} \to \mathbb{F}$ y $M : \mathbb{F} \to \mathbb{G}$ son funciones lineales y $\mathcal{U}, \mathcal{V}$ y $\mathcal{W}$ son bases respectivas de $\mathbb{E}, \mathbb{F}$ y $\mathbb{G}$ se tiene que* $\mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(M \circ L) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(M) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$. Hay una cuestión esencial, que es interpretar $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = c$ para bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$. En general, si tenemos $L : \mathbb{E} \to \mathbb{E}$, se calcula $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})$ por columnas poniendo en la columna j las coordenadas de $L(u_j)$ en la base $\mathcal{V}$. Si $L = \mathrm{id}_{\mathbb{E}}$, la columna j de $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = c$ vendrá dada por $\mathrm{id}_{\mathbb{E}}(u_j) = u_j = c_j^1 v_1 + \ldots + c_j^n v_n$. Y esta es la observación esencial: *los coeficientes c_j^i que expresan la base $\mathcal{U}$ en función de la base $\mathcal{V}$ constituyen la matriz* $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = c = \left(c_j^i\right)$. Las matrices $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$ y $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})$ se llaman **matrices de cambio de base**. Frecuentemente, pondremos id en vez de $\mathrm{id}_{\mathbb{E}}$ si se sobrentiende $\mathbb{E}$.

Teorema 64 *Sean $\mathcal{U}$ y $\mathcal{V}$ bases de $\mathbb{E}$ y las matrices $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$ y $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})$ de cambio de base. Entonces,*

1. *Son matrices inversas una de otra; o sea,* $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = I_n$.
2. *Para $x \in \mathbb{E}$, las matrices de coordenadas en esas bases se relacionan por*

$$\mathrm{mat}^{\mathcal{U}}(x) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}^{\mathcal{V}}(x) \quad y \quad \mathrm{mat}^{\mathcal{V}}(x) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}^{\mathcal{U}}(x).$$

Demostración. Todo es trivial porque

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = I_n$$

y, con la fórmula $\mathrm{mat}^{\mathcal{V}}(L(x)) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) \cdot \mathrm{mat}^{\mathcal{U}}(x)$ (3.5) que relaciona coordenadas, aplicada a $\mathrm{id}_{\mathbb{E}}$,

$$\mathrm{mat}^{\mathcal{U}}(x) = \mathrm{mat}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}(x)) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}^{\mathcal{V}}(x).$$

Naturalmente, $\mathrm{mat}^{\mathcal{V}}(x) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}^{\mathcal{U}}(x)$ se obtiene de modo análogo. ♣

Las cálculos en $\Bbbk^n$ son sencillos conceptualmente pero es inevitable la tediosa labor de multiplicar e invertir matrices. Si $\mathcal{U} = (u_1, \ldots, u_n)$ es arbitraria y $\mathcal{V} = \mathcal{E}$, la matriz $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})$ tiene por columnas las coordenadas de cada u_j en la base estándar, que son $u_j^1, \ldots, u_j^n$. Así pues (¡observación esencial!)

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n}) = \begin{pmatrix} u_1^1 & \cdots & u_n^1 \\ \vdots & \ddots & \vdots \\ u_1^n & \cdots & u_n^n \end{pmatrix}.$$

Dicho con palabras: yuxtaponiendo las columnas de los u_j se obtiene $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n})$. Si se necesitara, y muchas veces se necesita, conocer $\mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id}_{\Bbbk^n})$, hay que usar que $\mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id}_{\Bbbk^n}) = \left[\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n})\right]^{-1}$ (**1** en el teorema 64). Si la situación involucra otra base $\mathcal{V} = (v_1, \ldots, v_n)$ en vez de $\mathcal{E}$, se usa que

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\Bbbk^n}) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id}_{\Bbbk^n}) \cdot \mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n}) = \left[\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n})\right]^{-1} \cdot \mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id}_{\Bbbk^n}).$$

Problema 141 *En $\mathbb{E} = \mathbb{R}^2$ tenemos bases*

$$\mathcal{U} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right) \quad y \quad \mathcal{V} = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right).$$

El vector x tiene coordenadas $(1,2)^\top$ en $\mathcal{E}$ e y tiene coordenadas $(3,-1)^\top$ en $\mathcal{U}$. Dar las coordenadas de x en $\mathcal{U}$ y de y en $\mathcal{V}$. ♦

Solución. Todo es una simple aplicación de las fórmulas. Primeramente,

$$\mathrm{mat}^{\mathcal{U}}(x) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id})\,\mathrm{mat}^{\mathcal{E}}(x) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id})^{-1}\,\mathrm{mat}^{\mathcal{E}}(x) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \end{pmatrix}.$$

A continuación,

$$\mathrm{mat}^{\mathcal{V}}(y) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id})\,\mathrm{mat}^{\mathcal{U}}(y) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{V}}(\mathrm{id})\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id})\,\mathrm{mat}^{\mathcal{U}}(y) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id})^{-1}\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id})\,\mathrm{mat}^{\mathcal{U}}(y).$$

Esto de concreta en

$$\mathrm{mat}^{\mathcal{V}}(y) = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ -2 \end{pmatrix}.$$ ♦

Problema 142 *Sean $h, k \in \mathbb{R}$ fijos no nulos y las bases*

$$\mathcal{U} = \left(\begin{pmatrix} 1 \\ h \\ k \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right) \quad y \quad \mathcal{V} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} h \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} k \\ 0 \\ 1 \end{pmatrix} \right).$$

Determinar los vectores $v \in \mathbb{R}^3$ que tengan las mismas coordenadas en $\mathcal{U}$ y $\mathcal{V}$.

Problema 143 *En $\mathbb{R}^2$ nos dicen que los vectores $x = (1,2)^\top$ e $y = (1,-1)^\top$ tienen en una base $\mathcal{V} = (v_1, v_2)$ desconocida coordenadas $(0,1)^\top$ y $(2,-1)^\top$. Determinar $\mathcal{V}$ y $\mathrm{mat}^{\mathcal{V}}(z)$ para $z = (4,3)^\top$.*

La notación $\mathrm{mat}_{\bullet}^{\bullet}(\bullet)$ ayuda mucho, pero hay que advertir sobre un error fácil de cometer. Si nos dan la relación entre bases con una colección de n ecuaciones expresando los v como combinación de los u, ¿cómo se obtiene la matriz $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$? La respuesta correcta es

$$\left\{ \begin{array}{c} v_1 = c_1^1 u_1 + \ldots + c_1^n u_n \\ \vdots\!\vdots \\ v_n = c_n^1 u_1 + \ldots + c_n^n u_n \end{array} \right. \text{ equivale a } (v_1, \ldots, v_n) = (u_1, \ldots, u_n) \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix}. \tag{3.11}$$

Entendemos la parte matricial como una multiplicación formal porque las "matrices" $(u_1, \ldots, u_n)$ y $(v_1, \ldots, v_n)$ tienen como coeficientes vectores y no escalares. Pero se puede confundir la expresión matricial a la derecha de (3.11) con otras muy parecidas, *que son falsas,*

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad (v_1, \ldots, v_n) = (u_1, \ldots, u_n) \begin{pmatrix} c_1^1 & \cdots & c_1^n \\ \vdots & \ddots & \vdots \\ c_n^1 & \cdots & c_n^n \end{pmatrix}.$$

Ambas dan $v_1 = c_1^1 u_1 + c_2^1 u_2 + c_3^1 u_3 + \ldots + c_1^1 u_n$ como primera ecuación, que debería ser sin embargo $v_1 = c_1^1 u_1 + c_1^2 u_2 + c_1^3 u_3 + \ldots + c_1^n u_n$. Lo ilustramos con un ejemplo en $\Bbbk^2$ con bases $\mathcal{V} = (u, v)$ y $\mathcal{U} = (p, q)$ relacionadas por

$$\left\{ \begin{array}{l} u = p + q \\ v = q \end{array} \right. \text{, equivalente a } (u, v) = (p, q) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Las formas matriciales no equivalentes al par de ecuaciones en la llave son

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \quad (u,v) = (p,q) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ y ambas dan } \begin{cases} u = p \\ v = p+q \end{cases}$$

Podemos sintetizar las advertencias (en realidad es repetir el análisis de qué es $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$) diciendo que si se dispone de la ecuación $v_j = c_j^1 u_1 + \ldots + c_j^n u_n$, la sucesión $\left(c_j^1, \ldots, c_j^n\right)$ *no es la fila* j *de* $c = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$) *sino la columna* j.

Otra advertencia es que si $c = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$, la matriz c da la base $\mathcal{V}$ en función de la base $\mathcal{U}$, pero $\mathrm{mat}^{\mathcal{V}}(x) = c\,\mathrm{mat}^{\mathcal{U}}(x)$ *es falsa, siendo lo verdadero* $\mathrm{mat}^{\mathcal{U}}(x) = c\,\mathrm{mat}^{\mathcal{V}}(x)$. Como se ve, se traspone la posición de $\mathcal{U}$ con $\mathcal{V}$ en el aserto. Se tiene pues, aparte de (3.11), que

$$(v_1, \ldots, v_n) = (u_1, \ldots, u_n) \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \text{ equivale a } \mathrm{mat}^{\mathcal{U}}(x) = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \mathrm{mat}^{\mathcal{V}}(x),$$

sin olvidar que $\mathrm{mat}^{\mathcal{U}}(x)$ y $\mathrm{mat}^{\mathcal{V}}(x)$ son vectores columna.

Problema 144 *Sea* $\mathbb{E} = \Bbbk_n[X]$ *y* $\left(\tau^0, \tau^1, \ldots, \tau^n\right)$ *una sucesión de elementos distintos de* $\Bbbk$. *Probar que la base* $(T_0(X), T_1(X), \ldots, T_n(X))$ *asociada a esta sucesión, que se construyó al tratar la interpolación de Lagrange, se relaciona con la base estándar* $\mathcal{E} = (1, X, \ldots, X^n)$ *por*

$$(1, X, \ldots, X^n) = (T_0(X), T_1(X), \ldots, T_n(X)) \begin{pmatrix} \left(\tau^0\right)^0 & \left(\tau^0\right)^1 & \cdots & \left(\tau^0\right)^j & \cdots & \left(\tau^0\right)^n \\ \left(\tau^1\right)^0 & \left(\tau^1\right)^1 & \cdots & \left(\tau^1\right)^j & \cdots & \left(\tau^1\right)^{n-1} \\ \vdots & \vdots & & \vdots & & \vdots \\ \left(\tau^{n-1}\right)^0 & \left(\tau^{n-1}\right)^1 & \cdots & \left(\tau^{n-1}\right)^j & \cdots & \left(\tau^{n-1}\right)^n \\ \left(\tau^n\right)^0 & \left(\tau^n\right)^1 & \cdots & \left(\tau^n\right)^j & \cdots & \left(\tau^n\right)^n \end{pmatrix}.$$

Como consecuencia la matriz $(n+1) \times (n+1)$ *que aparece es invertible por ser una matriz de cambio de base. Nota: este tipo de matrices, llamadas* ***matrices de Vandermonde****, tienen interés. Su determinante es famoso y vale* $\prod_{0 \le i < j \le n} \left(\tau^j - \tau^j\right)$.

Hemos usado la notación $\mathrm{mat}_{\bullet}^{\bullet}(\bullet)$ para relacionar las fórmulas de cambio de base y de coordenadas. No obstante, lo tradicional es hacerlo con índices. Damos un resumen de lo que se hace, aunque en el fondo vamos a repetir el trabajo con una presentación diferente. Hay una simplificación importante que ayuda a detectar que el índice de los sumatorios respecto al que se suma, es el correcto. Si $\mathcal{U} = (u_1, \ldots, u_n)$ es la primera base, conviene denotar a la segunda por $\mathcal{U}' = (u_{1'}, \ldots, u_{n'})$, marcando la diferencia en el índice. Lo importante no es usar primas; podrían ser estrellas o colores diferentes, lo que importa es que la diferencia se refleje en el índice. Las coordenadas de $x \in \mathbb{E}$ en $\mathcal{U}$ y $\mathcal{U}'$ serán $\left(x^1, \ldots, x^n\right)$ y $\left(x^{1'}, \ldots, x^{n'}\right)$, marcando también la diferencia en el índice. Expresamos $\mathcal{U}'$ en función de $\mathcal{U}$ por

$$\begin{cases} u_{1'} = c_{1'}^1 u_1 + \ldots + c_{1'}^n u_n \\ \vdots\vdots \\ u_{n'} = c_{n'}^1 u_1 + \ldots + c_{n'}^n u_n \end{cases} \text{ equivalente a } (u_{1'}, \ldots, u_{n'}) = (u_1, \ldots, u_n) \begin{pmatrix} c_{1'}^1 & \cdots & c_{n'}^1 \\ \vdots & \ddots & \vdots \\ c_{1'}^n & \cdots & c_{n'}^n \end{pmatrix}.$$

Expresada $\mathcal{U}'$ respecto de $\mathcal{U}$ (segunda base respecto a la primera), podemos expresar las coordenadas x^i respecto a las $x^{j'}$ (primeras coordenadas respecto a las segundas) comparando

$$x = \sum_{i=1}^n x^i u_i \text{ con } x = \sum_{j'=1}^n x^{j'} u_{j'} = \sum_{j'=1}^n x^{j'} \left(\sum_{i=1}^n c_{j'}^i u_i \right) = \sum_{i,j'=1}^n c_{j'}^i x^{j'} u_i = \sum_{i=1}^n \left(\sum_{j'=1}^n c_{j'}^i x^{j'} \right) u_i$$

El coeficiente de cada u_i es único ya que $\mathcal{U}$ es base. Por consiguiente,

$$x^i = \sum_{j'=1}^n c_{j'}^i x^{j'},\ 1 \le i \le n, \text{ equivalente a } \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} c_{1'}^1 & \cdots & c_{n'}^1 \\ \vdots & \ddots & \vdots \\ c_{1'}^n & \cdots & c_{n'}^n \end{pmatrix} \begin{pmatrix} x^{1'} \\ \vdots \\ x^{n'} \end{pmatrix},$$

que es la fórmula (3.11) que ya conocíamos. La convención de Einstein ayuda a recordar cualquiera de las fórmulas $u_{j'} = \sum_{i=1}^{n} c_{j'}^{i} u_i$ o $x^i = \sum_{j'=1}^{n} c_{j'}^{i} x^{j'}$ porque se suma siempre respecto a un par subíndice-superíndice ambos con o sin primas. Es fácil conjeturar que si queremos expresar la base $\mathcal{U}$ respecto de $\mathcal{U}'$, la fórmula será $u_j = \sum_{i'=1}^{n} d_j^{i'} u_{i'}$ y que las coordenadas $x^{i'}$ (segundas coordenadas) se expresaran respecto a las x^i (primeras coordenadas) por $x^{i'} = \sum_{j=1}^{n} d_j^{i'} x^j$. Por supuesto, $c = \mathrm{mat}_{\mathcal{U}'}^{\mathcal{U}}(\mathrm{id})$ y $d = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}'}(\mathrm{id})$. También el que sean c y d inversas una de otra (ya visto en el teorema 64) se comprueba con índices calculando

$$u_{i'} = \sum_{j=1}^{n} c_{i'}^{j} u_j \overset{1}{=} \sum_{j=1}^{n} c_{i'}^{j} \left(\sum_{q'=1}^{n} d_j^{q'} u_{q'} \right) \overset{2}{=} \sum_{q'=1}^{n} \left(\sum_{j=1}^{n} c_{i'}^{j} d_j^{q'} \right) u_{q'} \overset{3}{=} \sum_{q'=1}^{n} (dc)_{i'}^{q'} u_{q'}.$$

En (1) se sustituye u_j por su expresión en la base $\mathcal{U}'$, en (2) cambia el orden de precedencia de los sumatorios (en vez de sumar primero en q' y luego en j, se hace al revés), y en (3) se observa que al sumar en j aparece el coeficiente de la matriz dc (producto de las dos matrices). Esta ecuación nos da dos expresiones de $u_{i'}$ en la base $\mathcal{U}'$, pero la expresión de un vector en una base es única, luego tiene que ser $(dc)_{i'}^{q'} = \delta_{i'}^{q'}$; o sea, $dc = I$. De modo análogo, $cd = I$.

Problema 145 *Hacer con índices las comprobaciones que faltan.*

Teorema 65 *Sea $L : \mathbb{E} \to \mathbb{F}$ con bases $\mathcal{U}$ y $\mathcal{U}'$ en $\mathbb{E}$ y $\mathcal{V}$ y $\mathcal{V}'$ en $\mathbb{F}$. Entonces,*

$$\mathrm{mat}_{\mathcal{U}'}^{\mathcal{V}'}(L) = \mathrm{mat}_{\mathcal{U}'}^{\mathcal{V}'}(\mathrm{id}_{\mathbb{F}} \circ L \circ \mathrm{id}_{\mathbb{E}}) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{V}'}(\mathrm{id}_{\mathbb{F}}) \, \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) \, \mathrm{mat}_{\mathcal{U}'}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}).$$

En el caso particular (el más frecuente) que sea $\mathbb{E} = \mathbb{F}$ y tengamos dos bases $\mathcal{U}$ y $\mathcal{V}$ en $\mathbb{E}$, la relación entre $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$ y $\mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L)$ es

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \, \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) \, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = \left(\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right)^{-1} \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) \, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}).$$

Demostración. Lo primero es obvio con $\mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(M \circ L) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(M) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$ y $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = \left(\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right)^{-1}$ en el caso particular. ♣

Problema 146 *Nos dan en $\mathbb{R}^2$ y $\mathbb{R}^3$ bases*

$$\mathcal{U} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right), \quad \mathcal{V} = \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right)$$

y $L : \mathbb{R}^2 \to \mathbb{R}^3$ que en las bases estándar tiene matriz

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Calcular $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$.

El lector puede ponerse muchos problemas numéricos cuya dificultad solo radica en el cálculo de inversas y multiplicaciones de matrices. *Lo que sigue hasta el final de la sección es material optativo con información esencial sobre el cambio de coordenadas con bases duales y paso de una base a otra.*

Hay un tipo de cálculo teórico muy accesible con la notación $\mathrm{mat}_{\bullet}^{\bullet}(\bullet)$. Sean $\mathcal{U}$ y $\mathcal{V}$ bases en $\mathbb{E}$ y $\mathcal{U}^*$ y $\mathcal{V}^*$ sus correspondientes bases duales. Se suponen conocidas $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$ y $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id})$; es decir, las expresiones de $\mathcal{V}$ en función de $\mathcal{U}$ y $\mathcal{U}$ en función de $\mathcal{V}$. Nos gustaría saber lo análogo para $\mathcal{U}^*$ y $\mathcal{V}^*$ y las relaciones entre las coordenadas de $f \in \mathbb{E}^*$ en $\mathcal{U}^*$ y $\mathcal{V}^*$. Esto último es lo más fácil de responder. Recordamos que $f : \mathbb{E} \to \Bbbk$ tiene una matriz para $\mathcal{U}$ y $\mathcal{E} = (1)$, la base estándar de $\Bbbk$. Esta matriz $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(f) = \mathrm{mat}_{\mathcal{U}^*}(f)$, que por la teoría general es $(f(u_1), \ldots, f(u_n))$, es una matriz fila, y es la que da las coordenadas de f en $\mathcal{U}^*$. Tenemos dos relaciones simétricas

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(f) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(f) \, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathcal{E}}) \quad \text{y} \quad \mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(f) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(f) \, \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathcal{E}}).$$

Resumiendo con palabras: si $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathcal{E}})$ expresa la segunda base $\mathcal{V}$ respecto de la primera $\mathcal{U}$, esta misma matriz expresa las coordenadas de f en $\mathcal{V}^*$ (las segundas coordenadas) en términos de las coordenadas de f en $\mathcal{U}^*$ (las primeras coordenadas). El lector deberá observar, o habrá observado, que ahora no hay trasposición de las palabras "primera" y "segunda" al expresar verbalmente las fórmulas.

Problema 147 *En* $\mathbb{E}=\Bbbk_2[X]$ *nos dan* $f:\Bbbk_2[X]\to\Bbbk$, $f(P(X))=P(1)$. *Relacionar las coordenadas de* $P(X)$ *en* $\mathcal{E}=\left(1,X,X^2\right)$ *y* $\mathcal{U}=\left(1,1+X,1+X+X^2\right)$. ♦

Solución. En realidad se pueden calcular las coordenadas directamente porque

$$\mathrm{mat}_{\mathcal{E}_3}(f)=\left(f(1),f(X),f\left(X^2\right)\right)=(1,1,1),\quad \mathrm{mat}_{\mathcal{U}}(f)=\left(f(1),f(1+X),f\left(1+X+X^2\right)\right)=(1,2,3).$$

No obstante, calculamos las matrices de cambio de coordenadas directamente porque

$$\left(1,1+X,1+X+X^2\right)=\left(1,X,X^2\right)\begin{pmatrix}1&1&1\\0&1&1\\0&0&1\end{pmatrix},\ \text{luego}\ \mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_3}(\mathrm{id})=\begin{pmatrix}1&1&1\\0&1&1\\0&0&1\end{pmatrix}.$$

¡Cuidado! Distinguimos $\mathcal{E}_1=(1)$, la base estándar de $\Bbbk$ y $\mathcal{E}_3=\left(1,X,X^2\right)$, la base estándar de $\Bbbk_2[X]$. Verificamos efectivamente que

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_1}(f)=\mathrm{mat}_{\mathcal{E}_3}^{\mathcal{E}_1}(f)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_3}(\mathrm{id})\ \text{porque es}\ (1,2,3)=(1,1,1)\begin{pmatrix}1&1&1\\0&1&1\\0&0&1\end{pmatrix}.$$

Si tuviéramos necesidad de conocer $\mathrm{mat}_{\mathcal{E}_3}^{\mathcal{U}}(\mathrm{id})$, para calcular por ejemplo $\mathrm{mat}_{\mathcal{E}_3}^{\mathcal{E}_1}(g)$ en función de $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_1}(f)$, siendo g otra forma lineal, habría que calcular la inversa de $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_3}(\mathrm{id})$. ♦

Teorema 66 *Sean* $\mathcal{U}$ *y* $\mathcal{V}$ *bases de* $\mathbb{E}$ *y* $c=\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$, *luego* $v_j=\sum_{i=1}^n c_j^i u_i$. *Las bases duales están relacionadas por las dos formas equivalentes*

$$u^i=\sum_k c_k^i v^k,\ 1\le i\le n,\qquad \begin{pmatrix}u^1\\\vdots\\u^n\end{pmatrix}=\begin{pmatrix}c_1^1&\cdots&c_n^1\\\vdots&\ddots&\vdots\\c_1^n&\cdots&c_n^n\end{pmatrix}\begin{pmatrix}v^1\\\vdots\\v^n\end{pmatrix}.$$

Demostración. Si se quiere conocer cómo se expresan las u^i en función de las v^j, que son elementos de $\mathcal{U}^*$ y $\mathcal{V}^*$, lo más sencillo es poner unos coeficientes indeterminados p en $u^i=\sum_k p_k^i v^k$, y debería ser $p_k^i=c_k^i$. Aplicamos ambos lados (no olvidemos que son *funciones* sobre $\mathbb{E}$) a v_j y

$$u^i(v_j)=u^i\left(\sum_k c_j^k u_k\right)=\sum_k c_j^k u^i(u_k)=\sum_k c_j^k\delta_k^i=c_j^i,\quad \sum_k p_k^i v^k(v_j)=\sum_k p_k^i\delta_j^k=p_j^i,$$

y el teorema es inmediato. ♣

Recordamos que para los teoremas 29 y 32 se definieron tres tipos de operaciones elementales sobre sucesiones $(a_1,\ldots,a_k)$ de vectores de $\mathbb{E}$ de forma análoga a como se ha hecho con vectores filas o columna. Para $k=n=\dim(\mathbb{E})$, estos teoremas probaban que si $\mathcal{U}=(u_1,\ldots,u_n)$ es una base y se transforma en $\mathcal{V}=(v_1,\ldots,v_n)$ con operaciones elementales, entonces $\mathcal{V}$ es también base de $\mathbb{E}$. Se verifica la recíproca.

Teorema 67 *Si* $\mathcal{U}=(u_1,\ldots,u_n)$ *y* $\mathcal{V}=(v_1,\ldots,v_n)$ *son bases de* $\mathbb{E}$, *hay una sucesión de operaciones elementales que transforman* $\mathcal{U}$ *en* $\mathcal{V}$.

Demostración. Hemos usado en 3.11 ,y otros sitios más, expresiones formales de multiplicación como

$$(v_1,\ldots,v_n)=(u_1,\ldots,u_n)\begin{pmatrix}c_1^1&\cdots&c_n^1\\\vdots&\ddots&\vdots\\c_1^n&\cdots&c_n^n\end{pmatrix}\ \text{para indicar}\ v_j=c_j^1u_i+\ldots c_j^nu_n,\ 1\le j\le n.$$

Resumimos esto como $\mathcal{V}=\mathcal{U}\cdot c$. Si F es una matriz elemental y $\mathcal{V}=\mathcal{U}\cdot F$ es fácil aunque pesado comprobar que $\mathcal{V}$ es la base que resulta al aplicar a $\mathcal{U}$ la correspondiente operación elemental sobre bases. Por ejemplo, si

$$(z_1,\ldots,z_n)=(w_1,\ldots,w_n)F,\ \text{siendo}\ F=\begin{pmatrix}1&&&\\&\ddots&&\\\cdots&\lambda&\ddots&\\&\vdots&&1\end{pmatrix}$$

(todo 1 en la diagonal y fuera de ella todo 0 salvo λ en fila i columna j), vemos que $z_k = w_k$ si $k \neq j$ y $z_j = w_i + \lambda w_j$; es decir, $(z_1, \ldots, z_n)$ se obtiene a partir de $(w_1, \ldots, w_n)$ por la operación elemental "sustituir z_j por $z_j = w_i + \lambda w_j$". Con el teorema 17 factorizamos c invertible c como producto de matrices elementales $c = F_1 F_2 \cdots F_h$. Denotamos también con F_j a la operación homóloga con bases. Pues bien, si $\mathcal{V} = \mathcal{U} \cdot c$, obtenemos $\mathcal{V}$ aplicando a $\mathcal{U}$ la operacion F_1, luego a $\mathcal{U} \cdot F_1$ la operación $F_2, \ldots$, llegando a $\mathcal{V}$ tras aplicar F_h a $\mathcal{U} \cdot F_1 \cdots F_{h-1}$. ♣

El teorema tiene una idea subyacente muy importante para el futuro. Muchos teoremas dicen que dado un objeto matemático, digamos un endomorfismo, que se expresa con una base $\mathcal{U}$, podemos encontrar otra base $\mathcal{V}$ en donde L tenga una expresión "mejor" (se la puede llamar expresión estándar, canónica, normalizada, etc.). Este teorema 67 implica que la nueva base "buena" $\mathcal{V}$ será alcanzable realizando a bases previas operaciones elementales y una estrategia atractiva será investigar cómo se altera la expresión de este objeto cuando la base se va alterando a su vez por operaciones elementales.

3.6. La traza de un endomorfismo

El teorema 65 dice que si tenemos un *endomorfismo* (luego $\mathbb{E} = \mathbb{F}$) y solo consideramos matrices del tipo $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = a$, su relación con otra matriz del mismo tipo $\mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L) = b$ es por conjugación; es decir, existe una matriz invertible c (que es matriz de cambio de base) tal que $a = c^{-1}bc$. Esto nos permite asociar a L ciertos números a través de su matriz para una base $\mathcal{U}$ *sin que la aparente dependencia de la base se de en la realidad.* Un excelente ejemplo es la **traza del endomorfismo** L. Se define así: se toma una base cualquiera $\mathcal{U}$, se escribe su matriz $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = a$ y se define $\mathrm{tr}(L) = \mathrm{tr}(a)$. De haber elegido otra base $\mathcal{V}$ se tendría $\mathrm{tr}(a) = \mathrm{tr}\left(c^{-1}bc\right) \overset{*}{=} \mathrm{tr}\left(cc^{-1}b\right) = \mathrm{tr}(b)$, luego no influye la base que se elija. Se ha usado en $\overset{*}{=}$ algo ya sabido para matrices, que es $\mathrm{tr}(ab) = \mathrm{tr}(ba)$. De hecho las propiedades del problema 27 se reflejan para endomorfismos L y M de $\mathbb{E}$ pues

$$\mathrm{tr}(L+M) = \mathrm{tr}(L) + \mathrm{tr}(M), \quad \mathrm{tr}(\lambda L) = \lambda \,\mathrm{tr}(L), \quad \mathrm{tr}(LM) = \mathrm{tr}(ML).$$

si bien no podemos decir que sea $\mathrm{tr}\left(L^{\top}\right) = \mathrm{tr}(L)$ porque no se ha definido la traspuesta de una función lineal (requiere espacios duales).

Problema 148 *Calcular las trazas de la simetría S y proyección P para una descomposición $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ y de $L : \mathbb{E} \to \mathbb{E}$ de la forma $L(x) = x + f(x)\,c$, siendo $f : \mathbb{E} \to \Bbbk$ lineal con $f(c) \neq 0$.*

Problema 149 *Si tomamos $L = D : \mathbb{R}_n[X] \to \mathbb{R}_n[X]$ como la derivada, se ha calculado que $\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(D)$ tiene todo ceros en la diagonal, luego $\mathrm{tr}(D) = 0$. Por ejemplo, si $n = 2$,*

$$a = \mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(D) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Dar una base $\mathcal{U}$ tal que $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(D)$ no tenga todo ceros en la diagonal pero (¡como tiene que ser!) cumpla $\mathrm{tr}\left(\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(D)\right) = 0$.

3.7. Referencias y funciones afines

Aunque casi toda la teoría es válida en un cuerpo $\Bbbk$ arbitrario (puede haber algún problema con la característica 2) vamos a desarrollar toda la teoría con $\Bbbk = \mathbb{R}$. En adelante, $\mathbb{E}$ será un espacio real de dimensión n. Se puede visualizar una base $\mathcal{U} = (u_1, \ldots, u_n)$ como unos ejes coordenados, siendo el eje j la línea que genera u_j. Si $x = \sum x^i u_i$, medimos x^j proyectando x sobre $\lg(u_j)$ en paralelo a $\lg(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n)$ y, si la proyección es y, se sabe que $y = x^j u_j$. Los ejes de todas las bases tienen un origen común, que es $0 \in \mathbb{E}$. Interesa disponer de ejes cuyo origen esté en otro $p \neq 0$ y poder asociar coordenadas a $x \in \mathbb{E}$. Para ello disponemos de las llamadas **referencias**, obtenidas tomando un $p \in \mathbb{E}$ y una base $\mathcal{U} = (u_1, \ldots, u_n)$ de $\mathbb{E}$. Dada $\mathcal{R} = (p : \mathcal{U}) = (p : u_1, \ldots, u_n)$, asignamos a x las **coordenadas afines** $\left(x^1, \ldots, x^n\right)$ dadas por

$$x - p = \sum_{i=1}^{n} x^i u_i \text{ y escribiremos } \mathrm{mat}^{\mathcal{R}}(x) = \begin{pmatrix} 1 \\ x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} 1 \\ \mathrm{mat}^{\mathcal{U}}(x-p) \end{pmatrix}.$$

Luego veremos las ventajas de introducir esa 0−coordenada ficticia 1, la misma para todo x. Hay un peligro con la notación y es que podemos confundir x^i, la coordenada *vectorial* i de x en $\mathcal{U}$, con la coordenada *afín* i de x en $\mathcal{R}$, que enseguida se ve con ejemplos que pueden ser muy distintas. Por ello haremos hincapié hablando de "coordenadas (vectoriales) en $\mathcal{U}$" y "coordenadas (afines) en $\mathcal{R}$".[6] Si en cualquier $\mathbb{E}$ se tiene que el origen de $\mathcal{R}$ es 0, las coordenadas de x en $\mathcal{R}$ y $\mathcal{U}$ son las mismas. Protege contra las confusiones el forzarse a escribir las matrices de coordenadas de x en $\mathbb{E}$ de dimensión n como matrices $(n+1)\times 1$ con el primer coeficiente 1. Esto será importante al tratar las funciones afines.

La **referencia estándar en** $\mathbb{R}^n$ es $(0:\mathcal{E})$. Si $\mathcal{R}=(p:u_1,\ldots,u_n)$ es otra referencia en $\mathbb{R}^n$, las coordenadas de $x\in\mathbb{R}^n$ en $\mathcal{R}$ son los ξ^i solución del sistema

$$\begin{pmatrix} u_1^1 & \cdots & u_n^1 \\ \vdots & \ddots & \vdots \\ u_1^n & \cdots & u_n^n \end{pmatrix}\begin{pmatrix} \xi^1 \\ \vdots \\ \xi^n \end{pmatrix}=\begin{pmatrix} x^1-p^1 \\ \vdots \\ x^n-p^n \end{pmatrix}.$$

Se sorprenderá quizás el lector al ver la letra ξ (equivalente a x en el alfabeto griego) pero es que la propia simplicidad de $\mathbb{R}^n$ complica la explicación. Por definición, $x\in\mathbb{R}^n$ es $x=\left(x^1,\ldots,x^n\right)$ y hay que buscar otras letras para las coordenadas de x en $\mathcal{R}$.

Problema 150 *En* $\mathbb{E}=\mathbb{R}^2$ *se considera* $\mathcal{R}=(p:u_1,u_2)$ *y* x, *dados por*

$$p=\begin{pmatrix}1\\0\end{pmatrix},\ u_1=\begin{pmatrix}1\\1\end{pmatrix},\ u_2=\begin{pmatrix}1\\2\end{pmatrix},\ x=\begin{pmatrix}2\\4\end{pmatrix},$$

Calcular las coordenadas de x *en* $\mathcal{R}$. ♦

Solución. Hay que resolver $x-p=\xi^1u_1+\xi^2u_2$, que se concreta en

$$\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}\xi^1\\\xi^2\end{pmatrix}=\begin{pmatrix}2-1\\4-0\end{pmatrix},$$

y $\left(\xi^1,\xi^2\right)^\top=(-2,3)^\top$ distinto de $(-2,3)^\top\neq(2,4)^\top=x$. Brevemente, $\mathrm{mat}^{\mathcal{R}}(x)=(1,-2,3)^\top$.♦

Dejamos para más adelante las fórmulas de cambio de coordenadas al cambiar de referencia.

Una **función afín** o **aplicación afín** entre $\mathbb{E}$ y $\mathbb{F}$ es una función $\Lambda:\mathbb{E}\to\mathbb{F}$ del tipo $\Lambda(x)=t+L(x)$, siendo $t\in\mathbb{F}$ y $L:\mathbb{E}\to\mathbb{F}$ lineal. La traslación por t se denota por T_t y por tanto se puede definir Λ como $\Lambda=T_t\circ L$. La **parte lineal de** Λ, que es L, y la **traslación de** Λ, que es T_t (identificable con $t\in\mathbb{F}$) están unívocamente determinadas por Λ, ya que $t=\Lambda(0)$ y $L=(T_t)^{-1}\circ L=T_{-t}\circ L$. El siguiente teorema prueba propiedades de uso continuo de las funciones afines.

Teorema 68 *Las funciones afines tienen las siguientes propiedades*

1. *La composición de funciones afines es una función afín. En concreto, si*

$$\begin{cases}\Lambda_1=T_{t_1}\circ L_1:\mathbb{E}_1\to\mathbb{E}_2\\ \Lambda_2=T_{t_2}\circ L_2:\mathbb{E}_1\to\mathbb{E}_2\end{cases},\quad \textit{equivalente a}\quad \begin{cases}\Lambda_1(x_1)=t_1+L_1(x_1)\\ \Lambda_2(x_2)=t_2+L_2(x_2)\end{cases},$$

   entonces

$$\Lambda_2\circ\Lambda_1=T_{[t_2+L_2(t_1)]}\circ(L_2\circ L_1),\quad \textit{equivalente a}\quad \Lambda_2\circ\Lambda_1(x_1)=[t_2+L_2(t_1)]+(L_2\circ L_1)(x_1).$$

2. *La función afín* $\Lambda(x)=t+L(x)$ *es inyectiva, suprayectiva o biyectiva justamente si lo es la parte lineal* L.

[6] En la definición general de espacio afín $\mathbf{A}$, se distingue entre los *puntos* $P\in\mathbf{A}$ y los *vectores* $x\in\mathbb{E}$, el espacio asociado. Para quien conozca las definiciones, sabe que la P^i es *forzosamente* coordenada *afín* de P y que x^i es *forzosamente* coordenada *vectorial* de x. En este sentido, dar la definición más general de espacio afín, con su diferencia entre puntos y vectores, evita molestas confusiones. Nosotros hemos mantenido una estructura teórica más sencilla, al tratar unicamente la estructura de espacio afín asociable a cada espacio vectorial, y perdemos esta ventaja. No obstante hay que decir que al hacer problemas casi todos los autores suponen que $\mathbf{A}$ es un espacio vectorial con esta estructura naturalmente asociable, y se vuelven a encontrar con la posible confusión.

3. *La inversa de una función afín biyectiva es una función afín, en concreto, si* $\Lambda(x) = t + L(x)$,

$$\Lambda^{-1}(y) = -L^{-1}(t) + L^{-1}(y).$$

4. *La imagen por la función afín* $\Lambda(x) = t + L(x)$ *de* $\mathbb{E}$ *en* $\mathbb{F}$ *de un subespacio afín* $\mathbf{A} = s + \mathbb{A}$ *es el subespacio afín* $\Lambda(\mathbf{A}) = \Lambda(s) + L(\mathbb{A})$. *En particular la parte lineal de* Λ *lleva la dirección de* $\mathbf{A}$ *en la de* $\Lambda(\mathbf{A})$ *y si los subespacios* $\mathbf{A}_1$ *y* $\mathbf{A}_2$ *son paralelos, los subespacios afines* $\Lambda(\mathbf{A}_1)$ *y* $\Lambda(\mathbf{A}_2)$ *también lo son.*

Demostración. Para **1** verificamos que

$$\Lambda_2 \circ \Lambda_1(x_1) = t_2 + L_2(\Lambda_1(x_1)) = t_2 + L_2(t_1 + L_1(x_1)) = [t_2 + L_2(t_1)] + (L_2 \circ L_1)(x_1).$$

Para **2** empezamos observando que una composición de funciones inyectiva, suprayectiva o biyectiva, lo es también, luego $\Lambda = T_t \circ L$ tendrá esas propiedades si las tiene L. Aplicando el mismo razonamiento a $L = (T_t)^{-1} \circ \Lambda = T_{-t} \circ \Lambda$ se tiene probado **2**.

En general, hallar la inversa, si existe, de $f : X \to Y$ es resolver la ecuación $f(x) = y$ con $y \in \mathbb{Y}$ arbitrario e incógnita x. Poniendo $y = t + L(x)$, aplicamos L^{-1} (existe por **2**) y $x = L^{-1}(y - t) = -L^{-1}(t) + L^{-1}(y)$.

El apartado **4** es muy sencillo y queda para el lector. ♣

Se pueden añadir con muy poco esfuerzo más propiedades básicas de las funciones afines. Por ejemplo

1. Las funciones afines de $\mathbb{E}$ en $\mathbb{E}$ forman un grupo (llamado **grupo afín**) respecto de la operación de composición .

2. Para todo $x, y \in \mathbb{E}$ se tiene $\Lambda(x) - \Lambda(y) = L(x - y)$, y como $\Lambda(x) - L(x) = \Lambda(y) - L(y)$ quiere decirse que la función $\Lambda(x) - L(x)$ es constante y toma el valor t, la traslación de Λ.

Cuando teníamos $L : \mathbb{E} \to \mathbb{F}$ lineal y bases $\mathcal{U}$ y $\mathcal{V}$ en $\mathbb{E}$ y $\mathbb{F}$, vimos que podíamos manejar L con su matriz $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$. Si tenemos $\Lambda : \mathbb{E} \to \mathbb{F}$ afín y referencias $\mathcal{R} = (p : u_1, \ldots, u_n)$ y $\mathcal{S} = (q : v_1, \ldots, v_m)$ en $\mathbb{E}$ y $\mathbb{F}$, ¿qué se puede construir análogamente? Sea $\Lambda(x) = t + L(x)$ con $L : \mathbb{E} \to \mathbb{F}$ y $t \in \mathbb{F}$. Necesitamos la matriz $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$ y las coordenadas de la imagen del origen p de $\mathcal{R}$ en $\mathcal{S}$ (y no en $\mathcal{V}$); digamos que $\Lambda(p) - q = \sum_{i=1}^{m} \pi^i v_i$. El tal caso,

$$\Lambda(x) - q = \Lambda(x) - \Lambda(p) + \Lambda(p) - q = L(x) - L(p) + \Lambda(p) - q = L(x - p) + \Lambda(p) - q. \tag{3.12}$$

Comparamos entonces $\Lambda(x) - q = \sum_{i=1}^{m} (\Lambda(x) - q)^i v_i$ con

$$\Lambda(x) - q = L(x - p) + \Lambda(p) - q = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} a_j^i (x - p)^j v_i + \sum_{i=1}^{m} \pi^i v_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_j^i (x - p)^j + \pi^i \right) v_i,$$

y obtenemos la expresión en las referencias $\mathcal{R}$ y $\mathcal{S}$ de Λ, sea con ecuaciones o sea matricial,

$$(\Lambda(x) - q)^i = \sum_{j=1}^{n} a_j^i (x - p)^j + \pi^i,\ 1 \le i \le m,$$

$$\begin{pmatrix} (\Lambda(x) - q)^1 \\ \vdots \\ (\Lambda(x) - q)^m \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} (x - p)^1 \\ \vdots \\ (x - p)^n \end{pmatrix} + \begin{pmatrix} \pi^1 \\ \vdots \\ \pi^m \end{pmatrix}.$$

Aquí hay un punto delicado. Cuando ponemos $(\Lambda(x) - q)^i$ o $(x - p)^j$ nos estamos refiriendo a las coordenadas *vectoriales* en $\mathcal{V}$ y $\mathcal{U}$ de los vectores $\Lambda(x) - q$ o $x - p$. Pero precisamente, por definición, son las *coordenadas afines* de $\Lambda(x)$ y x en $\mathcal{S}$ y $\mathcal{R}$ de modo que, dejando claro que hay que distinguir siempre entre uno y otro tipo de coordenadas, hemos probado

Teorema 69 *Sea* $\Lambda(x) = t + L(x)$ *con* $L : \mathbb{E} \to \mathbb{F}$ *y* $t \in \mathbb{F}$. *Sean* $\mathcal{R} = (p : u_1, \ldots, u_n)$ *y* $\mathcal{S} = (q : v_1, \ldots, v_m)$ *referencias en* $\mathbb{E}$ *y* $\mathbb{F}$. *Las coordenadas* afines $\Lambda(x)^i$ *de* $\Lambda(x)$ *en* $\mathcal{S}$ *se expresan en función de las coordenadas* afines x^j *de* x *en* $\mathcal{R}$ *y* π^i *de* $\Lambda(p)$ *en* $\mathcal{S}$ *por sumatorios o ecuaciones o matriciales,*

$$\Lambda(x)^i = \sum_{j=1}^{n} a_j^i x^j + \pi^i,\ 1 \leq i \leq m, \qquad \begin{pmatrix} \Lambda(x)^1 \\ \vdots \\ \Lambda(x)^m \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} + \begin{pmatrix} \pi^1 \\ \vdots \\ \pi^m \end{pmatrix}.$$

Interesa transformar las ecuaciones matriciales en otras equivalentes más manejables, en concreto,

$$\begin{pmatrix} 1 \\ \Lambda(x)^1 \\ \vdots \\ \Lambda(x)^m \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \pi^1 & a_1^1 & \cdots & a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \pi^m & a_1^m & \cdots & a_n^m \end{pmatrix} \begin{pmatrix} 1 \\ x^1 \\ \vdots \\ x^n \end{pmatrix},$$

que se puede comprimir de modo muy útil para fórmulas teóricas en la forma

$$\mathrm{mat}^{\mathcal{S}}(\Lambda(x)) = \mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)\,\mathrm{mat}^{\mathcal{R}}(x), \quad \text{siendo} \quad \mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \pi^1 & a_1^1 & \cdots & a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \pi^m & a_1^m & \cdots & a_n^m \end{pmatrix}. \tag{3.13}$$

Se puede incluso comprimir todavía más $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)$ como

$$\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \pi^1 & a_1^1 & \cdots & a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \pi^m & a_1^m & \cdots & a_n^m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathrm{mat}^{\mathcal{V}}(\Lambda(p) - q) & \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) \end{pmatrix}. \tag{3.14}$$

Problema 151 *Consideremos* $\Lambda : \mathbb{R}^3 \to \mathbb{R}^2$ *definida por*

$$\Lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \end{pmatrix} = \begin{pmatrix} x + 2y + 2 \\ x - z - 2 \end{pmatrix}.$$

Calcular $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)$ *para* $\mathcal{R} = (p : u_1, u_2, u_3)$ *y* $\mathcal{S} = (q : v_1, v_2)$ *definidas por*

$$p = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ u_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},\ u_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},\ q = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix},\ v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \blacklozenge$$

Solución. Vamos a necesitar $\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) = \mathrm{mat}_{\mathcal{E}_2}^{\mathcal{V}}(\mathrm{id}_{\mathbb{R}^2})\,\mathrm{mat}_{\mathcal{E}_3}^{\mathcal{E}_2}(L)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}_3}(\mathrm{id}_{\mathbb{R}^3})$, que es

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{5}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Tenemos también

$$\Lambda(p) - q = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 2 \\ -2 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

Para determinar las coordenadas de $\Lambda(p)$ en $\mathcal{S}$ hay que resolver $\Lambda(p) - q = \pi^1 v_1 + \pi^2 v_2$ que da

$$\begin{pmatrix} \pi^1 \\ \pi^2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ -2 \end{pmatrix} = \begin{pmatrix} -\frac{7}{3} \\ \frac{8}{3} \end{pmatrix}.$$

Con las fórmulas teóricas generales,

$$\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{7}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{8}{3} & \frac{5}{3} & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

¡Cuidado! Si nos preguntan el valor de $x = (3,-3,-3)^\top$ *es incorrecto* multiplicar $(1,3,-3,-3)^\top$ por $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)$, que da $\left(1,-\frac{1}{3},\frac{14}{3}\right)^\top$, y responder que $\Lambda(x)$ tiene en $\mathcal{S}$ coordenadas $\left(-\frac{1}{3},\frac{14}{3}\right)^\top$. Así es porque estamos suponiendo que $(3,-3,-3)^\top$ son las coordenadas en la referencia estándar $(0:\mathcal{E}_3)$ de $\mathbb{R}^3$. Habría que calcular $\mathrm{mat}^{\mathcal{R}}(x)$, que no sería $(1,3,-3,-3)^\top$, y multiplicar ese vector por $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)$ y ahora sí tendríamos $\mathrm{mat}^{\mathcal{S}}(\Lambda(x)) = (\alpha,\beta)^\top$. Pero, ¡nueva advertencia!, α y β serían las coordenadas en $\mathcal{S}$ de $\Lambda(x)$ y no el punto de $\mathbb{R}^2$. Por supuesto, si $(3,-3,-3)^\top$ fuesen las coordenadas de x en $\mathcal{R}$, $\left(-\frac{1}{3},\frac{14}{3}\right)^\top$ serían las de $\Lambda(x)$ en $\mathcal{S}$. ♦

Cuando en los espacios vectoriales teníamos funciones lineales $L:\mathbb{E}\to\mathbb{F}$ y $M:\mathbb{F}\to\mathbb{G}$ con bases $\mathcal{U},\mathcal{V}$ y $\mathcal{W}$ en $\mathbb{E},\mathbb{F}$ y $\mathbb{G}$, probamos la fórmula fundamental $\mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(M)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(M\circ L)$. Ahora, con funciones afines $\Lambda:\mathbb{E}\to\mathbb{F}$ y $\Theta:\mathbb{F}\to\mathbb{G}$ y referencias $\mathcal{R},\mathcal{S}$ y $\mathcal{T}$ en $\mathbb{E},\mathbb{F}$ y $\mathbb{G}$, ¿hay una fórmula similar?

Teorema 70 *Consideramos funciones afines $\Lambda:\mathbb{E}\to\mathbb{F}$ y $\Theta:\mathbb{F}\to\mathbb{G}$ y sean $\mathcal{R}=(p:\mathcal{U})$, $\mathcal{S}=(q:\mathcal{V})$ y $\mathcal{T}=(t:\mathcal{W})$ referencias en $\mathbb{E},\mathbb{F}$ y $\mathbb{G}$. Entonces,*

$$\mathrm{mat}_{\mathcal{S}}^{\mathcal{T}}(\Theta)\,\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda) = \mathrm{mat}_{\mathcal{R}}^{\mathcal{T}}(\Theta\circ\Lambda).$$

Demostración. Con (3.13) comparamos $\mathrm{mat}^{\mathcal{T}}(\Theta\circ\Lambda(x)) = \mathrm{mat}_{\mathcal{R}}^{\mathcal{T}}(\Theta\circ\Lambda)\,\mathrm{mat}^{\mathcal{R}}(x)$ calculando

$$\begin{aligned}\mathrm{mat}^{\mathcal{T}}(\Theta\circ\Lambda(x)) &= \mathrm{mat}_{\mathcal{S}}^{\mathcal{T}}(\Theta)\left(\mathrm{mat}^{\mathcal{S}}\Lambda(x)\right)\\ &= \mathrm{mat}_{\mathcal{S}}^{\mathcal{T}}(\Theta)\left(\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)\,\mathrm{mat}^{\mathcal{R}}(x)\right) = \mathrm{mat}_{\mathcal{S}}^{\mathcal{T}}(\Theta)\,\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda)\,\mathrm{mat}^{\mathcal{R}}(x)\end{aligned}$$

y, al ser $x\in\mathbb{E}$ arbitrario, se obtiene la fórmula. Parece dificil pero es muy fácil. ♣

Si $\mathcal{U}$ y $\mathcal{V}$ son bases de $\mathbb{E}$, sabemos que $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$ permite, con sus columnas, expresar la base $\mathcal{V}$ por medio de la base $\mathcal{U}$ y relacionar coordenadas con $\mathrm{mat}^{\mathcal{U}}(x) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}^{\mathcal{V}}(x)$. ¿Hay algo similar con $\mathrm{mat}_{\mathcal{S}}^{\mathcal{R}}(\mathrm{id}_{\mathbb{E}})$?

Teorema 71 *Las coordenadas afines de $x\in\mathbb{E}$ en $\mathcal{R}=(p:\mathcal{U})$, $\mathcal{S}=(q:\mathcal{V})$ se relacionan por*

$$\mathrm{mat}^{\mathcal{S}}(x) = \mathrm{mat}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}}(x)) = \mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}^{\mathcal{R}}(x).$$

La matriz $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}})$ es

$$\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \pi^1 & a_1^1 & \cdots & a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \pi^m & a_1^m & \cdots & a_n^m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathrm{mat}^{\mathcal{V}}(p-q) & \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}) \end{pmatrix}.$$

Verbalmente: La primera columna de $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}})$ está formada por la matriz de $\mathbb{R}^{n+1}$ de las coordenadas del origen p de $\mathcal{R}$ (primera referencia) en $\mathcal{S}$ (segunda referencia), y las columna a_j de a está formada por las coordenadas de $u_j\in\mathcal{U}$ en $\mathcal{V}$. Aparece una situación similar a la observada al comparada en los efectos de cambio de base. La matriz $\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}})$ que da los datos de la primera referencia $\mathcal{R}$ en términos de la segunda, es la que expresa las segundas coordenadas en función de las primeras.

Demostración. Es muy sencilla aplicando el teorema 70 a $\Lambda=\mathrm{id}_{\mathbb{E}}$ pues

$$\mathrm{mat}^{\mathcal{S}}(x) = \mathrm{mat}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}}(x)) = \mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}^{\mathcal{R}}(x) \quad \text{y}$$

$$\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\mathrm{id}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \pi^1 & a_1^1 & \cdots & a_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \pi^m & a_1^m & \cdots & a_n^m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathrm{mat}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}(p)-q) & \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \end{pmatrix},$$

siendo $p-q=\sum_{i=1}^n \pi^i v_i$. ♣

Problema 152 *En $\mathbb{R}^2$ damos dos referencias*

$$\mathcal{R}=\left(\begin{pmatrix}1\\0\end{pmatrix}:\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\right),\quad \mathcal{S}=\left(\begin{pmatrix}\alpha\\\beta\end{pmatrix}:\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}2\\1\end{pmatrix}\right).$$

El punto z que en $\mathcal{R}$ tiene coordenadas $(2,2)^\top$ tiene coordenadas $(4,-4)^\top$ en $\mathcal{S}$. Determinar el origen q de $\mathcal{S}$. ♦

Solución. Calculamos primero

$$\operatorname{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id})=\operatorname{mat}_{\mathcal{E}}^{\mathcal{V}}(\mathrm{id})\operatorname{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id})=\operatorname{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id})^{-1}\operatorname{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id})=\begin{pmatrix}1&2\\2&1\end{pmatrix}^{-1}\begin{pmatrix}1&1\\1&-1\end{pmatrix}=\begin{pmatrix}\frac{1}{3}&-1\\\frac{1}{3}&1\end{pmatrix}.$$

Con el teorema 71 podemos plantear la ecuacion

$$\operatorname{mat}^{\mathcal{S}}(z)=\begin{pmatrix}1\\4\\-4\end{pmatrix}=\begin{pmatrix}1&0&0\\\sigma&\frac{1}{3}&-1\\\tau&\frac{1}{3}&1\end{pmatrix}\begin{pmatrix}1\\2\\2\end{pmatrix}=\begin{pmatrix}1\\\sigma-\frac{4}{3}\\\tau+\frac{8}{3}\end{pmatrix}=\operatorname{mat}^{\mathcal{R}}(z),$$

que da $\sigma=4+\frac{4}{3}=\frac{16}{3}$ y $\tau=-4-\frac{8}{3}=-\frac{20}{3}$. Entonces,

$$q=p+\sigma v_1+\tau v_2=\begin{pmatrix}1\\0\end{pmatrix}+\frac{16}{3}\begin{pmatrix}1\\2\end{pmatrix}+-\frac{20}{3}\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}-7\\4\end{pmatrix}.$$

Advertimos que "Determinar el origen q de $\mathcal{S}$" es un ambiguo porque podíamos dar su coordenadas en $\mathcal{U}$, $\mathcal{V}$, $\mathcal{R}$, $\mathcal{S}$ o $\mathcal{E}$; es decir, "con las coordenadas de siempre"; o sea, $q=(-7,4)^\top$. ♦

Problema 153 *En $\mathbb{R}^3$ nos dan un punto z que en dos referencias $\mathcal{R}$ y $\mathcal{S}$ tiene coordenadas $(1,0,0)$ y $(0,0,1)$. Las bases $\mathcal{U}$ y $\mathcal{V}$ de las referencias se relacionan por*

$$(v_1,v_2,v_3)=(u_1,u_2,u_3)\begin{pmatrix}1&0&1\\1&-1&0\\1&0&-1\end{pmatrix}.$$

Estudiar si hay datos suficientes para determinar las coordenadas del origen q de $\mathcal{S}$ en $\mathcal{R}$.

Un teorema esencial para construir funciones lineales $L:\mathbb{E}\to\mathbb{F}$ es tomar una base $\mathcal{U}=(u_1\ldots,u_n)$ de $\mathbb{E}$ y vectores *arbitrarios* $(w_1,\ldots,w_n)$ de $\mathbb{F}$, que pueden ser o no independientes. Se sabe que hay una y solo una función lineal $L:\mathbb{E}\to\mathbb{F}$ determinada por las condiciones $L(u_i)=w_i$, $1\le i\le n$ (teorema 51). Nos preguntamos si hay un teorema similar para funciones afines. Lo hay y es este.

Teorema 72 *Sea $\mathcal{R}=(p,u_1,\ldots,u_n)$ una referencia afin de $\mathbb{E}$ y $(w_0,w_1,\ldots,w_n)$ una sucesión de puntos arbitraria en $\mathbb{F}$. Hay una y solo una función afín Λ que cumple $\Lambda(p)=w_0$ y $\Lambda(p+u_i)=w_i$, $1\le i\le n$. Su fórmula es*

$$\Lambda(x)=w_0+\sum_{i=1}^{n}x^i(w_i-w_0).$$

Demostración. Las condiciones, si se verifican para Λ, dan

$$t+L(p)=w_0,\quad t+L(p)+L(u_i)=w_i,\quad L(u_i)=w_i-w_0.$$

Con todo esto, si $\left(x^1,\ldots,x^n\right)$ son las coordenadas afines de x en $\mathcal{R}$,

$$\Lambda(x)-\Lambda(p)=L(x-p)=L\left(\sum_{i=1}^{n}x^iu_i\right)=\sum_{i=1}^{n}x^i(w_i-w_0),$$

y obtenemos como única posible fórmula la que se anuncia. Para la existencia de Λ, definimos como L la función lineal que cumple $L(u_j)=w_j-w_0$ y $t=w_0-L(p)$. Se tiene $\Lambda(p)=t+L(p)=w_0$ y

$$\Lambda(p+u_j)=t+L(p+u_j)=t+L(p)+L(u_j)=w_0+w_j-w_0=w_j,$$

como queríamos demostrar. ♣

La definición de Λ en el teorema no requiere matrices. Sin embargo puede suceder que haya una referencia $\mathcal{S} = (q : \mathcal{V})$ en $\mathbb{F}$ y que los datos de las w_j vengan dados por sus coordenadas; es decir, conocemos $w_j - q = \sum_{i=1}^{m} w_j^i v_i$ siendo $0 \leq j \leq n$. En ese caso Λ es facilmente expresable con una matriz. En efecto, hemos visto que L, la parte lineal de Λ viene dada por las condiciones

$$L(u_j) = w_j - w_0 = (w_j - q) - (w_0 - q) = \sum_{i=1}^{m} \left(w_j^i - w_0^i\right) v_i,$$

luego $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L)$ tiene como sucesión de columnas $(w_1 - w_0, \ldots, w_n - w_0)$. Como $\left(w_0^1, \ldots, w_0^n\right)$ son, por definición, las coordenadas de $w_0 = \Lambda(p)$ en $\mathcal{S}$, obtenemos

$$\mathrm{mat}_{\mathcal{R}}^{\mathcal{S}}(\Lambda) = \begin{pmatrix} 1 & 0 \\ \left(w_0^i\right)_{i=1,\ldots,n} & \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(L) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ w_0^1 & \left(w_1^1 - w_0^1\right) & \cdots & \left(w_n^1 - w_0^1\right) \\ \vdots & \vdots & \ddots & \vdots \\ w_0^n & \left(w_1^n - w_0^n\right) & \cdots & \left(w_n^n - w_0^n\right) \end{pmatrix}.$$

Problema 154 *Dar todas las funciones afines $\Lambda : \mathbb{R}^2 \to \mathbb{R}^2$ que cumplen $\Lambda(e_1) = e_1$ y $\Lambda(e_2) = e_2$.* ♦

Solución. Tomamos en el teorema 72 como referencia $\mathcal{R}$ la estándar $\mathcal{R} = \mathcal{E} = (0 : e_1, e_2)$ y $(w_1, w_2, w_3) = (w, e_1, e_2)$ siendo w arbitrario, lo que hara que existan varios Λ. La referencia $\mathcal{S}$ será también $\mathcal{E}$. Tomemos $w = (\alpha, \beta)$. Al ser $L(e_j) = e_j - w$, vemos que

$$\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(L) = \begin{pmatrix} 1-\alpha & -\alpha \\ -\beta & 1-\beta \end{pmatrix}, \quad \mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(\Lambda) = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1-\alpha & -\alpha \\ \beta & -\beta & 1-\beta \end{pmatrix}. \blacklozenge$$

A veces es más interesante, dentro de lo que son problemas de puro cálculo, el obtener fórmulas generales, en vez de consideran casos concretos llenos de inversiones y multiplicaciones.

Problema 155 *Nos dan en $\mathbb{R}^n$ referencias $\mathcal{R} = (p : \mathcal{U})$ y $\mathcal{S} = (q : \mathcal{V})$ aparte de la referencia estándar $\mathcal{E}$. Tanto p y q como los vectores de $\mathcal{U}$ y $\mathcal{V}$ son vectores de $\mathbb{R}^n$ y yuxtapuestos estos últimos dan dos matrices $n \times n$ invertibles u y v. Probar que si Λ es la función afín que lleva $\mathcal{U}$ en $\mathcal{V}$ se tiene*

$$\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(\Lambda) = \begin{pmatrix} 1 & 0 \\ -q + vu^{-1}p & vu^{-1} \end{pmatrix}$$

Ilustrarlo con un ejemplo numérico en $\mathbb{R}^2$ donde, para $h \neq 9$,

$$\mathcal{R} = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} : \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \end{pmatrix}\right) \text{ en } \mathcal{S} = \left(\begin{pmatrix} h \\ -6 \end{pmatrix} : \begin{pmatrix} 5 \\ -2h-3 \end{pmatrix}, \begin{pmatrix} -3 \\ 2h+9 \end{pmatrix}\right)$$

Problema 156 *Sean $T(x) = x + t$ y $\Lambda(x) = s + L(x)$ una traslación y una función afín invertible. Probar que $\Lambda \circ T \circ \Lambda^{-1}$ es otra traslación $T'(x) = t' + x$. ¿Cuánto vale t' en función de t, s, L?* ♦

Solución primera. Si tenemos funciones afines $\Lambda_1, \ldots, \Lambda_k$ con partes lineales $L_1, \ldots, L_k$, el teorema 68 nos dice que la parte lineal de $\Lambda_1 \circ \cdots \circ \Lambda_k$ es $L_1 \circ \cdots \circ L_k$, Aplicando esto a $\Lambda \circ T \circ \Lambda^{-1}$ teniendo en cuenta que id es la parte lineal de una traslación, obtenemos que la parte lineal de $\Lambda \circ T \circ \Lambda^{-1}$ es $L \circ \mathrm{id} \circ L^{-1} = \mathrm{id}$, luego es una traslación. Le ofrecemos al lector que trabaje otra posible solución que tiene la ventaja de obtener que $t' = L(t)$, cosa que no dice el enunciado. ♦

Va a ser muy importante el conjunto **F** de los **puntos fijos** de Λ, también llamados **centros**, que son los que cumplen $\Lambda(c) = c$. Es sencillo ver que si este conjunto no es vacío, es un subespacio afín con dirección $\ker(\mathrm{id} - L)$; o sea que, tomando $c \in \mathcal{F}$ cualquiera, se tiene $\mathcal{F} = c + \ker(\mathrm{id} - L)$.

Problema 157 *Supongamos que $\Lambda(x) = t + L(x)$ tenga al menos un punto fijo c. Entonces,*

1. *Para* cualquier *punto fijo de Λ se puede escribir $\Lambda(x) = c + L(x - c)$.*[7]

[7] Puede haber varios puntos fijos pero el que el lado derecho parezca depender de c es solo aparente.

2. *El conjunto* $\mathbf{F}$ *de todos los puntos fijos es un subespacio afín cuya dirección es el subespacio vectorial de los vectores* v *tales que* $L(v) = v$*; es decir, los que fija* L*. Con símbolos,* $\mathbf{F} = c + \ker(\mathrm{id} - L)$

3. *Si no hay vectores fijos por* L*, hay al menos un punto fijo por* Λ*, que será único.*

Solución. Restamos las ecuaciones $c = \Lambda(c) = t + L(c)$ y $\Lambda(x) = t + L(x)$ y queda $\Lambda(x) - c = L(x) - L(c) = L(x - c)$, luego $\Lambda(x) = c + L(x - c)$. Si el lado izquierdo es independiente de c, el derecho también.

Si $x = \Lambda(x) = c + L(x - c)$, $x - c = L(x - c)$. Por tanto, x fijo equivale a $x - c \in \ker(\mathrm{id} - L)$ o $x \in c + \ker(\mathrm{id} - L)$.

Si L no fija puntos, $\ker(\mathrm{id} - L) = 0$ e $\mathrm{id} - L$ es un isomorfismo. Hay al menos un c que es solución de $t = (\mathrm{id} - L)(x)$. Operando, $t = c - L(c)$ y $c = \Lambda(c)$. Visto que hay puntos fijos para Λ, por **2**, formarán un subespacio afín de dirección $\ker(\mathrm{id} - L) = 0$, o sea, solo tendrá un punto. ♠

Hemos visto que si Λ fija c, entonces $\Lambda(x) = c + L(x - c)$. Esta forma de escribir Λ es muy útil porque indica que Λ es "esencialmente" como L, cosa que vamos a explicar. Pensemos con toda generalidad en un conjunto X y una biyección $\beta : X \to X$. Comparemos $f, g : X \to X$ relacionadas por $\beta \circ f \circ \beta^{-1} = g$. Muchas propiedades de f se trasladan a g. Por ejemplo, si $f^2 = f \circ f = f$ tenemos que g cumple lo análogo pues

$$g \circ g = \left(\beta \circ f \circ \beta^{-1}\right) \circ \left(\beta \circ f \circ \beta^{-1}\right) = \beta \circ f \circ \mathrm{id} \circ f \circ \beta^{-1} = \beta \circ f \circ \beta^{-1} = g.$$

Otra idea más difusa es que "lo que hace f en $p \in X$ lo hace g en $\beta(p) \in X$". Por ejemplo, si f fija p entonces g fija $\beta(p)$ porque $g(\beta(p)) = \left(\beta \circ f \circ \beta^{-1}\right)(\beta(p)) = \beta \circ f(p) = \beta(p)$. Volvemos a nuestra Λ inicial,

Si c es punto fijo de Λ, la ecuación $\Lambda(x) = c + L(x - c)$ equivale a $\Lambda = T_c \circ L \circ T_c^{-1}$ con $T_c(y) = c + y$. Si entendemos lo que hace L, por ejemplo, simetriza respecto a un plano a través del origen, intuimos que $\Lambda = T_c \circ L \circ T_c^{-1}$ será la simetría respecto a un plano que pasa por c, más concretamente el plano trasladado a c por T_c del plano vectorial que L fija. Esto permite presentar unas definiciones algebraicas cuyo sentido geométrico se puede comprobar más tarde. Naturalmente, tras esa comprobación, se tiene una definición que, vista solo con perspectiva algebraica, es oscura.

Sean P y S una proyección y una simetría *lineales* de $\mathbb{E}$ con el subespacio $\mathbb{F}$ como subespacio de puntos fijos. Usaremos en vez de la descomposición $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ el hecho que $P^2 = P$ y $S^2 = \mathrm{id}$ que son propiedades que definen proyecciones y simetrías. Consideremos el subespacio afín $\mathbf{F} = c + \mathbb{F}$. Definimos la proyección y simetría respecto de $\mathbf{F}$ por

$$\begin{cases} \Pi = T_c \circ P \circ T_c^{-1}, \text{ equivalente a } \Pi(x) = c + P(x - c) \\ \Sigma = T_c \circ S \circ T_c^{-1}, \text{ equivalente a } \Sigma(x) = c + S(x - c) \end{cases}$$

Por lo que acabamos de decir, Π será la proyección sobre el subespacio afín $T_c(\mathbb{F}) = c + \mathbb{F}$. Hay que asegurarse de que si $\mathbf{F}$ tiene dos expresiones, $\mathbf{F} = c_1 + \mathbb{F} = c_2 + \mathbb{F}$ las dos fórmulas Π_1 y Π_2 dan la misma función. Otro tanto hay que ver para Σ.

Problema 158 *Resolver la cuestión que acabamos de plantear.*

Trataremos esto más a fondo en los espacios afines euclidianos donde las proyecciones y simetrías serán ortogonales.

Capítulo 4

Determinantes

4.1. Permutaciones

Las permutaciones son necesarias para el desarrollo de los determinantes.[1] En la primera subsección damos lo más esencial que se necesita sin las demostraciones de los teoremas, que tienen sin embargo unos enunciados fáciles de comprender. En la segunda subsección está el desarrollo completo y puede considerarse material optativo.

4.1.1. Cuestiones básicas sobre permutaciones

Partimos de un conjunto finito C con $N \geq 1$ elementos. Una **permutación** de C es una biyección $P : C \to C$. Si C es pequeño se puede representar P con una matriz de dos filas y N columnas tal como

$$P = \begin{pmatrix} a_1 & a_2 & \cdots & a_i & \cdots & a_{N-1} & a_N \\ b_1 & b_2 & \cdots & b_i & \cdots & b_{N-1} & b_N \end{pmatrix}.$$

Esto indica que $P(a_1) = b_1, \ldots, P(a_N) = b_N$. Por ejemplo, si $C = \{1, 2, 3, 4, 5\}$ y

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix}, \tag{4.1}$$

sabemos que $P(4) = 4$ y $P(5) = 2$. Se entiende que si no aparece un $x \in C$ es porque $P(x) = x$. No es preciso que los elementos de la primera fila estén ordenados de modo creciente; de hecho C no se supone necesariamente ordenado aunque casi siempre $C \subset \mathbb{N}$ y se usa el orden natural. En los ejemplos, será $C \subset \mathbb{N}$. La función inversa de la permutación P se puede escribir invirtiendo el orden de las dos filas de la matriz como se hace en

$$P^{-1} = \begin{pmatrix} 3 & 1 & 5 & 4 & 2 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}.$$

Las permutaciones son funciones y pueden componerse con la definición general de composición de funciones $(Q \circ P)(x) = Q(P(x))$. Por ejemplo, en $C = \{1, 2, 3, 4, 5\}$

$$\begin{pmatrix} 1 & 5 & 2 \\ 2 & 1 & 5 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 1 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}.$$

Hemos definido las estructuras algebraicas de **cuerpo**, **anillo** y **espacio vectorial**. Definimos ahora la de **grupo** porque *las permutaciones de C forman un grupo*. Un grupo es un conjunto G dotado de una operación que asigna a cada par (a, b) de elementos de G otro $a \cdot b \in G$ (el **producto de** a y b) de modo que se cumplen las siguientes propiedades

1. **Asociatividad.** Para todo $a, b, c \in G$ se tiene $(a \cdot b) \cdot c = a \cdot (b \cdot c)$

[1] Hay textos, por ejemplo el de Hoffman y Kunze, que reducen su uso al mínimo desarrollando parcialmente el concepto de determinante y, con lo conseguido, probar lo que se necesita de permutaciones. A nuestro juicio es un atajo que exige demasiado trabajo.

2. **Existencia de unidad.** Existe un elemento $e \in G$, llamado la **unidad de** G, tal que para todo $a \in G$ se cumple $a \cdot e = e \cdot a = a$.

3. **Existencia de inverso**. Para todo elemento $a \in G$ existe otro elemento, denotado por a^{-1} y llamado el **inverso de** a, tal que $a \cdot a^{-1} = a^{-1} \cdot a = e$.

No se pide la **conmutatividad** $a \cdot b = b \cdot a$ como axioma de grupo, pero si se tiene, se dice que el grupo es **conmutativo** o **abeliano**. No vamos a trabajar ahora con grupos abelianos, porque el grupo de las permutaciones, que es lo que ahora nos interesa, no lo es. Sin insistir mucho diremos que si se tiene un cuerpo $\mathbb{k}$, un espacio vectorial $\mathbb{E}$, o un anillo $\mathbb{A}$ (por ejemplo, el de polinomios $\mathbb{k}[X]$), y definimos $a \cdot b = a + b$, denotando $+$ la suma de números, vectores o polinomios, se tiene un grupo abeliano. En estos ejemplos e es 0 (número, vector o polinomio 0) y $a^{-1} = -a$, el opuesto de a, sea lo que sea a.

Hasta aquí todos los ejemplos son de grupos abelianos. Hay dos tipos importantes de ejemplos de grupos no abelianos. El primero es tomar como G el conjunto de las matrices cuadradas $\mathbb{k}^{n\times n}$ *invertibles* y como producto el ordinario de matrices. En este caso $e = I_n$ (matriz unidad) y a^{-1} es la matriz inversa que ya conocemos. El otro ejemplo importante, del que ahora nos ocuparemos es el grupo de las permutaciones de C, siendo sus elementos las permutaciones de C y el producto del grupo la composición de funciones: $P \cdot Q = P \circ Q$. Los axiomas de grupo son consecuencia de las propiedades básicas de la composición de funciones. Aparece aquí de nuevo algo que ya vimos en los espacios vectoriales: los elementos del grupo son *funciones* y las operaciones algebraicas de la estructura son operaciones ya conocidas de las funciones.

En un grupo arbitrario G se define para $a \in G$ y $n \in \mathbb{N}$ el nuevo elemento $a^n = a \cdot a \cdot \cdots \cdot a$ (se multiplica a por sí mismo n veces). Se define también $a^0 = e$, la unidad de G, y $a^{-n} = \left(a^{-1}\right)^n$. Se tiene pues definido a^n para cualquier n positivo negativo o nulo y se comprueban sin problemas las reglas fundamentales $a^{m+n} = a^m \cdot a^n$ y $a^{mn} = (a^m)^n$, bien conocidas para números, que siguen siendo ciertas. Si $C = \{1,2,3,4,5\}$ y P es la permutación de (4.1) se tiene como ejemplo que P^3 y P^{-2} son

$$\begin{aligned} P^3 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix}^2 \\ &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 2 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 1 & 4 & 3 \end{pmatrix}, \end{aligned}$$

$$P^{-2} = \begin{pmatrix} 3 & 1 & 5 & 4 & 2 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}^2 = \begin{pmatrix} 3 & 1 & 5 & 4 & 2 \\ 2 & 5 & 1 & 4 & 3 \end{pmatrix}^2.$$

Por el problema 52, el número de biyecciones de un conjunto de N elementos en si mismo es $N!$, luego el grupo de las permutaciones de un conjunto finito es un grupo finito. En el caso $C = \{1, 2, \ldots, N\}$ este grupo se llama el **grupo simétrico de** N **elementos** y se suele denotar por S_N.

Hay un tipo interesante de permutaciones de C llamadas **ciclos.** Se eligen elementos *distintos* $a_1, a_2, \ldots, a_r$ en C (luego $r \leq N$) con $r \geq 2$. Al numerar estos elementos los hemos ordenado y, hecho esto, definimos la permutación

$$K = \begin{pmatrix} a_1 & a_2 & \cdots & a_i & \cdots & a_{r-1} & a_r \\ a_2 & a_3 & \cdots & a_{i+1} & \cdots & a_r & a_1 \end{pmatrix}$$

y lo llamamos el **ciclo definido por la sucesión** $(a_1, a_2, \ldots, a_r)$. Se describe fácilmente la acción de K: (a) si $x \notin \{a_1, a_2, \ldots, a_r\}$ entonces $K(x) = x$; (b) en los demás casos

$$K(a_1) = a_2,\ K(a_2) = a_3,\ K(a_3) = a_4\ \ldots, K(a_{r-1}) = a_r,\ K(a_r) = a_1.$$

Dicho más informalmente: si imaginamos $a_1, a_2, \ldots, a_r$ alineados, asignamos a cada elemento a_j el que le sigue, y al último a_r el primer puesto que ha quedado vacante, quedando fijos los otros $c \in C$ que no están en $\{a_1, a_2, \ldots, a_r\}$. Como segunda notación tendremos $(a_1, a_2, \ldots, a_r)$; o sea,

$$K = \begin{pmatrix} a_1 & a_2 & \cdots & a_i & \cdots & a_{r-1} & a_r \\ a_2 & a_3 & \cdots & a_{i+1} & \cdots & a_r & a_1 \end{pmatrix} = (a_1, a_2, \ldots, a_r).$$

Conviene admitir que (a_1) es también un ciclo cualquiera que sea a_1, pero la función es la identidad, que es entonces un ciclo por definición. Para familiarizarse, le pedimos al lector que compruebe las fórmulas

$$K^{-1} = \begin{pmatrix} a_2 & a_3 & \cdots & a_{i+1} & \cdots & a_r & a_1 \\ a_1 & a_2 & \cdots & a_i & \cdots & a_{r-1} & a_r \end{pmatrix} = (a_r, a_{r-1}, \ldots, a_2, a_1),$$

y que si $K^p = K \circ K \circ \cdots \circ K$ representa a K compuesta consigo misma p veces, se tiene

$$a_2 = K(a_1),\ a_3 = K^2(a_1),\ a_4 = K^3(a_1),\qquad a_r = K^{r-1}(a_1),\ a_1 = K^r(a_1).$$

Dado el ciclo $K = (a_1, a_2, \ldots, a_r)$, su **longitud** es r y su **órbita** el conjunto $\{a_1, a_2, \ldots, a_r\}$ (recordemos que los a_j son distintos por definición). El ejemplo más sencillo de ciclo aparte de id_C, es el que tiene longitud 2 y se llama **trasposición**. Una trasposición T es de la forma $T = (a, b)$ y lo que hace es intercambiar a y b dejando fijos los $x \neq a, b$. Concluimos las definiciones sobre ciclos definiendo que los ciclos K y L son **disjuntos** si las órbitas son disjuntas. Por ejemplo, para $C = \{1, 2, 3, 4, 5, 6\}$, los ciclos $K = (1, 4, 2)$ y $L = (3, 6)$ son disjuntos pues $\{1, 4, 2\} \cap \{3, 6\} = \emptyset$. Desde luego $K \circ L$ es una permutación pero ya no es un ciclo. De hecho,

$$K \circ L = \begin{pmatrix} 1 & 4 & 2 \\ 4 & 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 3 & 6 \\ 6 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 6 & 2 & 5 & 3 \end{pmatrix}$$

y se puede prescindir incluso de la quinta columna porque $K \circ L$ fija el 5. Admitiremos que cuando K y L son disjuntos se tiene $K \circ L = L \circ K$; es decir, *los ciclos disjuntos conmutan* (ver problema 163).

Lo que nos va a interesar principalmente del grupo de las permutaciones de C es la posibilidad de asignar a cada permutación P un número $\mathrm{sg}(P) = \pm 1$, que se llamará la **signatura de la permutación**, de modo que $\mathrm{sg}(P \circ Q) = \mathrm{sg}(P)\,\mathrm{sg}(Q)$ (la signatura del producto es el producto de las signaturas) y $\mathrm{sg}(\mathrm{id}_C) = 1$ (la signatura de la permutación identidad vale 1). La permutación es **par** o **impar** según sea $\mathrm{sg}(P) = +1$ o $\mathrm{sg}(P) = -1$. Todo lo que sigue es información para definir, calcular y manejar la signatura de una permutación lo mejor posible. En realidad hay dos definiciones de $\mathrm{sg}(P)$, en apariencia distintas, pero que son iguales. Para definirlas hay que *factorizar* P como producto de trasposiciones o de ciclos disjuntos; es decir, tener $P = T_h \circ \cdots \circ T_1$ o bien $P = K_r \circ \cdots \circ K_1$ siendo las T trasposiciones y las K ciclos disjuntos. ¿Se puede factorizar así? Sí, como se ve en la sección siguiente y aquí lo admitiremos. Sin embargo, y esto importa al definir la signatura, *estas factorizaciones no son únicas*. Aceptamos que

1. Si tenemos factorizaciones $P = T_h \circ \cdots \circ T_1 = S_k \circ \cdots \circ S_1$ con trasposiciones T y S, se verifica que $(-1)^h = (-1)^k$. Es como decir que si conseguimos una factorización de P con un número par (impar) de trasposiciones, cualquier otra tendrá asimismo un número par (impar) de trasposiciones.

2. Si $P = K_r \circ \cdots \circ K_1$ es una factorización con ciclos disjuntos, cualquier otra con ciclos disjuntos solo difiere de esta en el orden de los factores.

Damos ahora las dos posibles definiciones de signatura $\mathrm{sg}_1(P)$ y $\mathrm{sg}_2(P)$, que se puede demostrar que son iguales. La definición con trasposiciones es $\mathrm{sg}_1(P) = (-1)^h$, siendo h el número de trasposiciones que aparecen en alguna de las posibles factorizaciones $P = T_h \circ \cdots \circ T_1$ de P. La definición con ciclos disjuntos es

$$\mathrm{sg}_2(P) = (-1)^{\ell_r - 1} \cdots (-1)^{\ell_2 - 1} (-1)^{\ell_r - 1},$$

siendo las ℓ_j las longitudes de los ciclos disjuntos que aparecen en la factorización $P = K_r \circ \cdots \circ K_2 \circ K_1$. Los apartados **1** y **2** nos dicen que las definiciones son correctas (no dependen de elecciones intermedias) pero no que sean iguales, cosa que ya hemos dicho que hay que admitir, y por ello, en adelante, quitaremos el subíndice a sg y usaremos una u otra según nos convenga. Queda sin embargo la cuestión de cómo se escriben de manera efectiva las factorizaciones porque sin ellas no se puede calcular $\mathrm{sg}(P)$. Antes de ello damos como ilustración las demostraciones, definiendo sg con trasposiciones, de las fórmulas

$$\mathrm{sg}(P \circ Q) = \mathrm{sg}(P)\,\mathrm{sg}(Q),\qquad \mathrm{sg}(\mathrm{id}) = 1,\qquad \mathrm{sg}\left(P^{-1}\right) = \mathrm{sg}(P).$$

Para la primera supongamos $P = T_h \circ \cdots \circ T_1$ y $Q = R_k \circ \cdots \circ R_1$, lo que da $P \circ Q = T_h \circ \cdots \circ T_1 \circ R_k \circ \cdots \circ R_1$. Como sabemos (por teoremas indemostrados) que la signatura es la misma sea cual sea la factorización, elegimos la recién escrita y

$$\mathrm{sg}(P \circ Q) = (-1)^{h+k} = (-1)^h (-1)^k = \mathrm{sg}(P)\,\mathrm{sg}(Q).$$

Aún cuesta menos ver que $\mathrm{id}_C = T \circ T$ da $\mathrm{sg}(\mathrm{id}_C) = (-1)^2 = 1$. Por último $\mathrm{sg}\left(P^{-1}\right) = \mathrm{sg}(P)$ (la signatura de una permutación es igual a la de su inversa) porque $P \circ P^{-1} = \mathrm{id}_C$ implica

$$1 = \mathrm{sg}(\mathrm{id}_C) = \mathrm{sg}\left(P \circ P^{-1}\right) = \mathrm{sg}(P)\,\mathrm{sg}\left(P^{-1}\right),\quad \text{luego}\quad \mathrm{sg}\left(P^{-1}\right) = \mathrm{sg}(P),$$

puesto que la signatura toma solo los valores ± 1. Estas fórmulas son difíciles de probar si se toma como definición de signatura la dada por ciclos.

¿Cómo se puede calcular $\operatorname{sg}(P)$ con trasposiciones? En realidad no necesitamos en $P = T_h \circ \cdots \circ T_1$ conocer de modo exacto cómo son las T_j sino el número h. Esto se consigue con cierta facilidad si suponemos que $C = \{1, 2, \ldots, N\}$ y las trasposiciones son de la forma $(i, i+1)$; o sea, intercambian elementos adyacentes. Podemos ver

$$P = \begin{pmatrix} 1 & 2 & \cdots & i & \cdots & N-1 & N \\ a_1 & a_2 & \cdots & a_i & \cdots & a_{N-1} & a_N \end{pmatrix}.$$

como una transformación que desordena $(1, 2, \ldots, N)$ llevándolo a $(a_1, a_2, \ldots, a_N)$ y P^{-1} como la transformación que lleva la situación desordenada $(a_1, a_2, \ldots, a_N)$ a su orden natural $(1, 2, \ldots, N)$. En realidad vamos a calcular $\operatorname{sg}\left(P^{-1}\right)$ que ya sabemos que es $\operatorname{sg}(P)$. Damos el ejemplo de P en (4.1). Reordenamos $(3, 1, 5, 4, 2)$ en tres pasos como se indica en

$$(3, 1, 5, 4, 2) \xrightarrow{1} (1, 3, 5, 4, 2) \xrightarrow{3} (1, 2, 3, 5, 4) \xrightarrow{1} (1, 2, 3, 4, 5).$$

En el primer paso llevamos 1 a su sitio natural al principio, lo que supone intercambiar dos elementos adyacentes (el 3 y el 1). Luego llevamos el 2 al segundo lugar, que supone intercambiar tres veces elementos adyacentes (el 2 con el 4, luego con el 5 y luego con el 3) y en el tercer paso, intercambiar 5 y 4, los únicos que no están en orden. Se han necesitado $1 + 3 + 1 = 5$ trasposiciones (el número sobre la flecha es el número de trasposiciones requeridas), así que $\operatorname{sg}(P) = (-1)^5 = -1$. Pedimos al lector que calcule *de este modo* las signaturas de

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 2 & 3 & \cdots & N-1 & N \\ N & 1 & 2 & \cdots & N-2 & N-1 \end{pmatrix}$$

si bien adelantamos que, *como son ciclos,* con la segunda definición resultan ser $(-1)^3$ y $(-1)^{N-1}$.

Aunque $\operatorname{sg}(P)$ se puede calcular solo con trasposiciones, es bueno saber calcular con ciclos para los casos complicados. En este caso han de conocerse inevitablemente los ciclos K_j. Explicamos la idea general primero y luego damos ejemplos. Dado $x \in C$, los elementos $x, P(x), P^2(x), P^3(x), \ldots, P^m(x), \ldots$. no pueden ser todos distintos ya que C es finito, luego hay un *primer número* r tal que $P^r(x)$ es uno de los anteriores elementos $x, P(x), \ldots, P^{r-1}(x)$. Digamos que es $P^r(x) = P^q(x)$, $q < r$. Esta *primera repetición* debe ser $x = P^r(x)$ y no $P^q(x) = P^r(x)$ con $q \geq 1$. En efecto si pasara esto, aplicaríamos P^{-1} a $P^q(x) = P^r(x)$ obteniendo $P^{q-1}(x) = P^{r-1}(x)$ y la primera repetición no aparecería con $P^r(x)$ sino antes. Definimos la **órbita de x (según P),** que es $O(P, x) = \left\{x, P(x), P^2(x), \ldots, P^{r-1}(x)\right\} \subset C$. [2] Es esencial observar que P, restringida a $O(P, x)$ es el ciclo $\left(x, P(x), P^2(x), \ldots, P^{r-1}(x)\right)$. Si tomamos $y \neq x$ también tendrá órbita $O(P, y)$ pero (se prueba en la sección siguiente) se da el comportamiento extremo $O(P, x) = O(P, y)$ o bien $O(P, x) \cap O(P, y) = \emptyset$ (no puede haber un solapamiento parcial entre $O(P, x)$ y $O(P, y)$). Pues bien, una vez elegido x, sea $D = C - O(P, x)$. Elegimos y en D y hacemos con y el trabajo similar al hecho con x construyendo la órbita $\left\{y, P(y), P^2(y), , \ldots, P^m(y)\right\} = O(P, y)$ y observando que P restringida a $O(P, y)$ es el ciclo $\left(y, P(y), P^2(y), , \ldots, P^{s-1}(y)\right)$. Seguimos con $E = C - (O(P, x) \cup O(P, y))$ y elegimos z en E, calculando la órbita $\left\{z, P(z), P^2(z), \ldots, P^{t-1}(z)\right\} = O(P, z)$ y observando que P restringida a esta órbita es el ciclo $\left(z, P(z), P^2(z), \ldots, P^{t-1}(z)\right)$. Si seguimos con el procedimiento de ir quitando a C diversas órbitas llegará un momento en que nos quedará el vacío $\emptyset$, que es como decir que C será unión de las órbitas $O(P, x_1), O(P, x_2), \ldots, O(P, x_h)$ (sería $x_1 = x$, $x_2 = y$, $x_3 = z$,...) y P, restrigido a cada una se comportaría como un ciclo K_j. Se puede comprobar que $P = K_h \circ \cdots \circ K_2 \circ K_1$ aunque el orden de los K es irrelevante porque ya hemos dicho que ciclos disjuntos conmutan.

Vamos a factorizar en $C = \{1, 2, 3, 4, 5, 6\}$ como producto de ciclos la permutación

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 6 & 2 & 5 & 3 \end{pmatrix}.$$

Se tiene $O(P, 1) = \{1, 4, 2\}$, $O(P, 3) = \{3, 6\}$ y $O(P, 5)) = \{5\}$. Esto se justifica porque $1 \to 4 \to 2 \to 1$, $3 \to 6 \to 3$ y $5 \to 5$. Hay pues tres órbitas y $P = (1, 4, 2) \circ (3, 6) \circ (5)$ aunque se suele quitar

[2] La palabra órbita sugiere que al aplicar la potencias de P a x los puntos $P^n(x)$ van dando infinitas vueltas por el conjunto $O(P, x)$.

(5) pues este ciclo es la identidad. La signatura de P, al tener las tres órbitas longitudes $3,2,1$, es $\operatorname{sg}(P)=(-1)^2(-1)^1=-1$ y P es impar. Para factorizar P hemos buscado las órbitas de $1,3,5$. ¿Y si hubiéramos elegido otros elementos? Si empezamos por $x=6$ se tiene $6\to 3\to 6$; si luego se toma $4\notin O(P,6)=\{6,3\}$ se tiene $4\to 2\to 1\to 4$ y $O(P,4)=\{4,2,1\}$. Como $O(P,5)=\{5\}$ obtenemos $P=(6,3)\circ(4,2,1)\circ(5)$ que, aunque en apariencia es diferente de $P=(1,4,2)\circ(3,6)\circ(5)$, no es así ya que $(4,2,1)=(1,4,2)$ y $(3,6)=(6,3)$, conmutando los ciclos por ser disjuntos.[3]

Problema 159 *Comprobar en $C=\{1,\dots,9\}$ la siguiente factorización en ciclos disjuntos*

$$P=\begin{pmatrix}1&2&3&4&5&6&7&8&9\\3&5&7&4&9&1&6&2&8\end{pmatrix}=(1,3,7,6)\circ(2,5,9,8).$$

Dar ejemplos de otras permutaciones Q y R en C producto de tres ciclos que no sean la identidad, siendo Q par y R impar. ¿Es posible lo que se pide? ♦

Solución. Vemos que $1\to 3\to 7\to 6\to 1$, luego $(1,3,7,6)$ es el primer ciclo y $O(P,1)=\{1,3,7,6\}$. Tomamos $2\notin O(P,1)$ y vemos que $2\to 5\to 9\to 8\to 2$, luego $(2,5,9,8)$ es el segundo ciclo y $O(P,2)=(2,5,9,8)$. Fuera de la unión de $O(P,1)$ y $O(P,2)$ solo queda 4 y $4\to 4$ luego (4), que es la identidad, es el tercer ciclo, que en general no se escribe.

Para la segunda parte elegimos

$$Q=(1,2)\circ(3,4)\circ(5,6,7),\qquad R=(1,2)\circ(3,4)\circ(5,6).\ \blacklozenge$$

Problema 160 *Factorizar como producto de ciclos disjuntos*

$$P=(1,4)\circ(4,5,6,7)\circ(1,2,3),\qquad P=(a_1,b_1)\circ(a_1,\dots,a_p)\circ(b_1,\dots,b_q)$$

siendo $(a_1,\dots,a_p)$ y $(b_1,\dots,b_q)$ ciclos disjuntos. ¿Cómo sería la signatura?

Para no hacer trabajar en vano al lector, aunque a veces es instructivo, le decimos que saber la signatura es inmediato. La fórmula $\operatorname{sg}(P\circ Q)=\operatorname{sg}(P)\operatorname{sg}(Q)$, válida para todo P,Q y el hecho de que si K tiene longitud ℓ, su signatura es $(-1)^{\ell-1}$, nos da para la P general, $\operatorname{sg}(P)=(-1)^1(-1)^{p-1}(-1)^{q-1}=(-1)^{p+q-1}$. Ya que le hemos desvelado al lector la vía más fácil le pedimos que calcule $\operatorname{sg}(P)$ por la vía difícil (que empiece si quiere por el P concreto) factorizando P con ciclos disjuntos, y que constate otra vez que $\operatorname{sg}(P)=(-1)^{p+q-1}$.

Problema 161 *Factorizar como producto de ciclos disjuntos*

$$P=\begin{pmatrix}1&2&\cdots&n&n+1&n+2&\cdots&2n\\n+1&n+2&\cdots&2n&1&2&\cdots&n\end{pmatrix},$$

$$Q=\begin{pmatrix}1&2&3&\cdots&k-1&k&k+1&k+2&\cdots&n-1&n\\k&1&2&\cdots&k-2&k-1&n&k+1&\cdots&n-2&n-1\end{pmatrix}$$

y calcular sus signaturas.

4.1.2. Exposición más detallada de la signatura

Si el lector quiere solo conocer lo relativo a la signatura definida con trasposiciones basta que se limite al problema 162, el teorema 75 (con la demostración por inducción, ignorando la otra que usa ciclos), el teorema 76 y el teorema 77 (aunque no podrá comparar las dos definiciones de signatura). La lectura completa cubre todo lo relativo a ciclos e incluye problemas.

Iniciamos la sección con unos problemas que piden demostrar que ciertas igualdades de permutaciones son ciertas. Insistimos que como las permutaciones P y Q son funciones, probar que $P=Q$ exige, si no se tiene nada mejor, probar que para todo $x\in C$ se tiene $P(x)=Q(x)$. Los tres problemas que siguen se dan resueltos porque se van a usar para los teoremas principales. Lo que se pide al lector es el trabajo de verificarlos.

[3] El lector puede intentar, en este ejemplo y en los que siguen, calcular la signatura con trasposiciones. Creemos que verá que es más laborioso.

Problema 162 *Dados $a, b, z \in C$ distintos entre sí probar que $T = (a,b) = (z,b) \circ (z,a) \circ (z,b)$.* ♦

Solución. Verificamos que

$$[(z,b) \circ (z,a) \circ (z,b)](x) = \begin{cases} [(z,b) \circ (z,a)](x) = (z,b)(x) = x \text{ si } x \notin \{a,b,z\} \\ [(z,b) \circ (z,a)](a) = (z,b)(z) = b \text{ si } x = a \\ [(z,b) \circ (z,a)](z) = (z,b)(a) = a \text{ si } x = b \\ [(z,b) \circ (z,a)](b) = (z,b)(b) = z \text{ si } x = z \end{cases}$$

y esto es lo mismo que $T(x)$ para $x \notin \{a,b,z\}$, $x = a$, $x = b$ o $x = z$. ♦

Problema 163 *Probar que si K y L son ciclos disjuntos, $K \circ L = L \circ K$ y que para $C = \{1,2,3\}$ se pueden tomar ciclos tales que $K \circ L \neq L \circ K$ (por eso decíamos que los grupos de permutaciones no son conmutativos).* ♦

Solución. Sean $K = (a_1, a_2, \ldots, a_r)$ y $L = (b_1, b_2, \ldots, b_s)$ con órbitas A y B. Entonces

$$(K \circ L)(x) = \begin{cases} x \text{ si } x \notin A \cup B \\ L(x) \text{ si } x \in B \\ K(x) \text{ si } x \in A \end{cases}, \qquad (L \circ K)(x) = \begin{cases} x \text{ si } x \notin A \cup B \\ L(x) \text{ si } x \in B \\ K(x) \text{ si } x \in A \end{cases}$$

teniéndose en cuenta que $x \in A \cap B$ está descartado. Con otra notación,

$$K = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{r-1} & a_r \\ a_2 & a_3 & a_4 & \cdots & a_r & a_1 \end{pmatrix}, \qquad L = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots & b_{s-1} & b_s \\ b_2 & b_3 & b_4 & \cdots & b_s & b_1 \end{pmatrix},$$

$$K \circ L = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{r-1} & a_r & b_1 & b_2 & b_3 & \cdots & b_{s-1} & b_s \\ a_2 & a_3 & a_4 & \cdots & a_r & a_1 & b_2 & b_3 & b_4 & \cdots & b_s & b_1 \end{pmatrix} = L \circ K.$$

Si $K = (1,2,3)$ y $L = (1,2)$ se tiene

$$(1,2,3) \circ (1,2) = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \qquad (1,2) \circ (1,2,3) = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

y por tanto, $K \circ L \neq L \circ K$. ♦

Problema 164 *Probar que todo ciclo se puede expresar como composición de trasposiciones con*

$$(a_1, a_2, \ldots, a_p) = (a_1, a_p) \circ (a_1, a_{p-1}) \circ \cdots \circ (a_1, a_4) \circ (a_1, a_3) \circ (a_1, a_2). \text{ ♦} \tag{4.2}$$

Solución. Sea K el ciclo de la izquierda y L la composición de trasposiciones de la derecha. Si x no es uno de los a_j se tiene $K(x) = L(x)$. Si $x = a_j$, $j < p$, tenemos $K(a_j) = a_{j+1}$ y $K(a_p) = a_1$. Probemos con L,

$$L(a_1) = [(a_1, a_p) \circ (a_1, a_{p-1}) \circ \cdots \circ (a_1, a_4) \circ (a_1, a_3) \circ (a_1, a_2)](a_2) = a_2$$

ya que a_2 no aparece en las trasposiciones $(a_1, a_p), (a_1, a_{p-1}), \ldots, (a_1, a_3)$. De modo similar,

$$\begin{aligned} L(a_2) &= [(a_1, a_p) \circ (a_1, a_{p-1}) \circ \cdots \circ (a_1, a_4) \circ (a_1, a_3)](a_1) \\ &= [(a_1, a_p) \circ (a_1, a_{p-1}) \circ \cdots \circ (a_1, a_4)](a_3) = a_3 \end{aligned}$$

ya que a_3 no aparece en las trasposiciones $(a_1, a_p), (a_1, a_{p-1}), \ldots, (a_1, a_4)$. Con generalidad, para $j < p$,

$$\begin{aligned} L(a_j) &= [(a_1, a_p) \circ \cdots \circ (a_1, a_j) \circ \cdots \circ (a_1, a_2)](a_j) \overset{1}{=} [(a_1, a_p) \circ \cdots \circ (a_1, a_{j+1}) \circ (a_1, a_j)](a_j) \\ &= [(a_1, a_p) \circ \cdots \circ (a_1, a_{j+2}) \circ (a_1, a_{j+1})](a_1) = [(a_1, a_p) \circ \cdots \circ (a_1, a_{j+2})](a_{j+1}) \overset{2}{=} a_{j+1}. \end{aligned}$$

Se usa en $\overset{1}{=}$ que a_j no aparece en $(a_1, a_{j-1}), (a_1, a_{j-2}), \ldots, (a_1, a_2)$ y en $\overset{2}{=}$ que a_{j+1} no aparece en $(a_1, a_p), \ldots, (a_1, a_{j+2})$. Y usando en $\overset{1}{=}$ que a_p no está en las trasposiciones $(a_1, a_{p-1}), \ldots, (a_1, a_2)$,

$$L(a_p) = [(a_1, a_p) \circ (a_1, a_{p-1}) \circ \cdots \circ (a_1, a_3) \circ (a_1, a_2)](a_p) \overset{1}{=} (a_1, a_p)(a_p) = a_1.$$

Esto prueba que $K = L$; o sea (4.2). ♦

Teorema 73 *Sea $P \neq \mathrm{id}$ una permutación. Entonces, las órbitas $O(P,x)$ y $O(P,y)$ son disjuntas o iguales. Por tanto C se puede expresar como unión disjunta*[4] *de todas las órbitas, incluidas las que solo tienen un punto.*

Demostración. Sean $O(P,x) = \left\{x, P(x), \ldots, P^{r-1}(x)\right\}$ y $O(P,y) = \left\{y, P(y), \ldots, P^{s-1}(y)\right\}$ recordando que $P^r(x) = x$ y $P^s(y) = y$. Supongamos que $O(P,x) \cap O(P,y) \neq \emptyset$. Para ver que $O(P,x) = O(P,y)$ basta mostrar que $x \in O(P,y)$ lo que llevará a $O(P,x) \subset O(P,y)$. De modo simétrico se obtendrá $O(P,y) \subset O(P,x)$ y finalmente $O(P,x) = O(P,y)$. Al ser $O(P,x) \cap O(P,y) \neq \emptyset$ se pueden tomar m y n tales que $P^m(x) = P^n(y)$ con $m < r$ y $n < s$. Aplicamos P^{r-m} a $P^m(x) = P^n(y)$ y queda

$$P^{r-m}\left(P^n(y)\right) = P^{r-m}\left(P^m(x)\right) = P^r(x) = x$$

y $x = P^{r-m+n}(y) \in O(P,y)$ como queríamos. Hemos visto que si las órbitas no son disjuntas es porque son la misma. Como siempre $x \in O(P,x)$ tenemos que C es unión disjunta de las órbitas. ♣

Teorema 74 *Cada permutación $P \neq \mathrm{id}$ se puede factorizar como composición de ciclos disjuntos $P = K_r \circ \cdots \circ K_1$ que conmutan entre sí, siendo esta factorización única salvo reordenación de los factores.*

Demostración. Vamos a hacer de modo un poco más formal la construcción de la sección anterior para factorizar una permutación como producto de ciclos. Tomamos x_1 tal que $P(x_1) \neq x_1$, consideramos la órbita y el ciclo

$$O(P,x_1) = \left\{x_1, P(x_1), P^2(x_1), \ldots, P^{r_1-1}(x_1)\right\}, \qquad K_1 = \left(x_1, P(x_1), P^2(x_1), \ldots, P^{r_1-1}(x_1)\right).$$

Elegimos, si existe, $x_2 \notin O(P,x_1)$ y consideramos también la órbita y ciclo

$$O(P,x_2) = \left\{x_2, P(x_2), P^2(x_2), \ldots, P^{r_2-1}(x_2)\right\}, \qquad K_1 = \left(x_2, P(x_2), P^2(x_2), \ldots, P^{r_2-1}(x_2)\right).$$

Por el teorema 73 precedente, $O(P,x_1) \cap O(P,x_2) = \emptyset$ porque si no sería $O(P,x_1) = O(P,x_2)$ y $x_2 \in O(P,x_1)$ contra la hipótesis. Elegimos, si existe, $x_3 \notin O(P,x_1) \cup O(P,x_2)$ y consideramos la órbita y ciclo $O(P,x_3)$ y K_3. Hay que observar que las tres órbitas son disjuntas pues si dos se cortaran serían iguales. Así vamos construyendo las orbitas disjuntas de $x_1, x_2, x_3, \ldots, x_k$ y, al ser C finito llega un momento en que no hay puntos fuera de $O(P,x_1) \cup \ldots \cup O(P,x_r)$ para cierto r y C es unión disjunta de estas órbitas. Afirmamos que $P = K_r \circ \cdots \circ K_1$ cosa no muy difícil de probar. Sea $y \in C$. Deberá estar en una y solo una de las órbitas. Digamos $y = P^n(x_j) \in O(P,x_j)$. Obviamente, $P(y) = P^{n+1}(x_j)$. Por otra parte, como los ciclos, al ser disjuntos conmutan,

$$(K_r \circ \cdots \circ K_1)(y) = K_j \circ (K_r \circ \cdots \circ K_{j+1} \circ K_{j-1} \circ \cdots \circ K_1)(y) \overset{1}{=} K_j(y)$$

y $\overset{1}{=}$ está justificado porque y no está en las órbitas distintas de la de K_j y por tanto los K_h con $h \neq j$ no mueven y. Como K_j es un ciclo e $y = P^n(x_j)$ se tiene $K_j(y) = P^{n+1}(x_j) = P(y)$ y acaba la primera parte de la demostración.

Si tenemos otra factorización $P = L_s \circ \cdots \circ L_1$ teniendo los ciclos órbitas disjuntas, la orbita de cada L_i es una de las órbitas de P luego esta factorización es la anterior con otro orden de factores. ♣

Con este teorema se define la signatura $\mathrm{sg}_2(P)$ de la sección precedente factorizando $P = K_r \circ \cdots \circ K_1$ con ciclos disjuntos siendo ℓ_j la longitud de K_j y $\mathrm{sg}_2(P) = (-1)^{(\ell_r - 1)} \cdot (-1)^{(\ell_{r-1} - 1)} \cdot \cdots \cdot (-1)^{(\ell_1 - 1)}$. Si definimos la **paridad** de $n \in \mathbb{Z}$ como $(-1)^n$, se puede decir que la signatura de un ciclo de longitud ℓ es la paridad de $\ell - 1$ y que si P se factoriza como producto de ciclos disjuntos, la signatura de P es el producto de las paridades de los ciclos. Como otra factorización solo se diferencia de la inicial por permutación de ciclos, la definición es correcta. Para la definición $\mathrm{sg}_1(P)$ necesitamos saber que toda permutación es factorizable como composición de trasposiciones. Hay demostraciones directas, pero lo más rápido es aprovechar lo ya disponible.

Teorema 75 *Cada permutación se puede factorizar como composición de trasposiciones $P = T_h \circ \cdots \circ T_1$, si bien la factorización no es única.*

[4] Se dice que A es **unión disjunta** de los subconjuntos $B_1, B_2, \ldots, B_n$ si $A = B_1 \cup B_2 \cup \ldots \cup B_n$ y para todo par (i,j) con $i \neq j$ se tiene que $B_i \cap B_j = \emptyset$.

Demostración. Se puede utilizar (4.2), pues al mostrar que cada ciclo se puede factorizar como producto de trasposiciones y cada permutación como producto de ciclos (teorema 74) resulta al final que toda permutación es producto de trasposiciones. ♣

Si T es una trasposición cualquiera, al ser $T^2 = T \circ T = \mathrm{id}$, se puede poner $\mathrm{id} = T^2 = T^4 = T^2 \circ U^2$ con otra trasposición U. Vemos que id se puede factorizar de muchas maneras como producto de trasposiciones pero, al menos en los ejemplos, el número de trasposiciones es par. Esto siempre sucede así.

Teorema 76 *Si se factoriza* $\mathrm{id} = T_h \circ T_{h-1} \circ \cdots \circ T_2 \circ T_1$ *como producto de trasposiciones (que se pueden repetir, siendo* $T_i = T_j$ *aunque sea* $i \neq j$*) siempre se tiene que* h*, el número de trasposiciones, es par.*

Demostración. Fijamos $z \in C$. Si $T_j = (a_j, b_j)$ con $z = a_j$ o $z = b_j$ la dejamos como está. Si no, escribimos, utilizando el problema 162 anterior, $(a_j, b_j) = (z, b_j) \circ (z, a_j) \circ (z, b_j)$. Tenemos una nueva factorización $\mathrm{id} = U_1 \circ \cdots \circ U_k$, donde hemos añadido un número par de factores, dos por cada T_j con $z \notin \{a_j, b_j\}$. Por tanto h y k tienen la misma paridad y para probar que h es par bastará mostrar que k lo es. Lo que importa ahora es que todas las trasposiciones U son de la forma $U_j = (z, x_j)$ con $x_j \neq z$. Sea X el conjunto de las x_j que aparecen en las U_j. Es muy posible que una x de $X \subset C$ aparezca en varias U_j, pero debe hacerlo en un número par de ellas, porque si no sería $\mathrm{id}(x_j) = z \neq x_j$. El número k se puede expresar por tanto como una suma de números pares y es par. ♣

Si $P = T_1 \circ \cdots \circ T_r = S_1 \circ \cdots \circ S_t$ son dos factorizaciones de P, usamos que $S_j^{-1} = S_j$ luego

$$\mathrm{id} = (T_1 \circ \cdots \circ T_r)(S_1 \circ \cdots \circ S_t)^{-1} = T_1 \circ \cdots \circ T_r \circ S_t \circ \cdots \circ S_1$$

y el teorema precedente afirma que $r + t = 2h$, un número par. Si $r - t$ fuera impar, $r - t = 2k + 1$, se tendría sumando que $2r = 2(h + k) + 1$, que es imposible. Queda pues que $r - t$ es par $(-1)^r = (-1)^t$ y definir $\mathrm{sg}_1(P)$ como la paridad del número de de trasposiciones de una factorización de P es una definición correcta. Ya probamos en la sección precedente que

$$\mathrm{sg}_1(P \circ Q) = \mathrm{sg}_1(P)\,\mathrm{sg}_1(Q), \qquad \mathrm{sg}_1(\mathrm{id}) = 1, \qquad \mathrm{sg}_1\left(P^{-1}\right) = \mathrm{sg}_1(P).$$

Solo queda probar que $\mathrm{sg}_1(P) = \mathrm{sg}_2(P)$, que está en el teorema siguiente.

Teorema 77 *Las dos definiciones de signatura coinciden; es decir* $\mathrm{sg}_1(P) = \mathrm{sg}_2(P)$.

Demostración. Si P es un ciclo $K = (a_1, a_2, \ldots, a_p)$ lo factorizamos como producto de trasposiciones con la ecuación (4.2) y como intervienen $p - 1$ trasposiciones y p es la longitud ℓ de K, se tiene que $\mathrm{sg}_1(K) = (-1)^{\ell - 1}$ luego para ciclos al menos es $\mathrm{sg}_1(K) = \mathrm{sg}_2(K)$. Para P arbitrario, con el teorema 74 factorizamos $P = K_r \circ \cdots \circ K_1$ como producto de ciclos disjuntos siendo ℓ_j la longitud de K_j. Entonces,

$$\mathrm{sg}_1(P) = \mathrm{sg}_1(K_r \circ \cdots \circ K_1) \overset{1}{=} \mathrm{sg}_1(K_r) \cdot \cdots \cdot \mathrm{sg}_1(K_1) \overset{2}{=} (-1)^{\ell_r - 1} \cdots (-1)^{\ell_1 - 1} \overset{3}{=} \mathrm{sg}_2(P).$$

Se ha usado en $\overset{1}{=}$ que $\mathrm{sg}_1(P \circ Q) = \mathrm{sg}_1(P)\,\mathrm{sg}_1(Q)$, en $\overset{2}{=}$ que si K tiene longitud ℓ, $\mathrm{sg}_1(K) = (-1)^{\ell - 1}$, y en $\overset{3}{=}$ la propia definición de $\mathrm{sg}_2(P)$. ♣

Hay mucho más que se puede decir sobre permutaciones y que aquí no se trata. Cerramos la sección unos comentarios sobre cómo contar el número de números *enteros* en un intervalo de $\mathbb{Z}$ y con un problema. Frecuentemente tenemos enteros $a < b$ y el conjunto $S = \{x \in \mathbb{Z} \,|\, a \leq x \leq b\}$. ¿Cuántos elementos tiene S? La respuesta es $\#S = (b - a) + 1$. Como ejemplos,

$$\#\{1, 2, \ldots, n\} = (n - 1) + 1 = n, \qquad \#\{7\} = (7 - 7) + 1 = 1,$$

$$\#\{-2, -1, \ldots, n - 1, n\} = (n - (-2)) + 1 = n + 3, \qquad \#\{x \in \mathbb{Z} \mid -k \leq x \leq k\} = 2k + 1.$$

Sean $a < b$, con $a, b \in \mathbb{Z}$ y los conjuntos

$$T = \{x \in \mathbb{Z} \,|\, a < x < b\}, \quad U = \{x \in \mathbb{Z} \,|\, a < x \leq b\}, \quad V = \{x \in \mathbb{Z} \,|\, a \leq x < b\},$$

El número de sus elementos es respectivamente

$$\#S = (b - a) + 1, \qquad \#T = (b - a) - 1, \qquad \#U = b - a = \#V.$$

El lector puede si quiere darlas por evidentes y, si es exigente, probarlas por inducción.

Problema 165 *Sea* $P : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ *una permutación que no es la identidad. Probar que hay al menos un* i *tal que* $P(i) < i$*. Indicación: usar inducción.*

4.2. Determinantes

Es casi seguro que el lector conoce la definición de determinante de una matriz 2×2 o 3×3 y que le han dado un procedimiento para calcularlo en el caso 4×4, que es generalizable a $n\times n$, pero casi con seguridad no le han dado la definición de determinante de modo general. Esto se explica porque es una definición compleja, poco manejable, y a la que no se le ve inicialmente su posible utilidad. Decimos esto para prevenirle al lector de que le espera un periodo de dificultades técnicas y que, aunque hay variantes en el modo de explicar el concepto, ninguna hace que sea un concepto intuitivo como bastantes que hemos visto en los capítulos anteriores. Pedimos al lector que acepte que los determinantes son importantes y que hay que saber calcularlos, optando por dar directamente la definición.

Sea $\Bbbk$ un cuerpo[5] y $a\in\Bbbk^{n\times n}$; o sea, una matriz cuadrada. El **determinante** de a, se denota de varios modos[6]

$$\det(a),\quad \det\begin{pmatrix} a_1^1 & \cdots & a_n^1\\ \vdots & \ddots & \vdots\\ a_1^n & \cdots & a_n^n\end{pmatrix},\quad |a|,\quad \begin{vmatrix} a_1^1 & \cdots & a_n^1\\ \vdots & \ddots & \vdots\\ a_1^n & \cdots & a_n^n\end{vmatrix}$$

y es un elemento de $\Bbbk$. Viene definido por

$$\det(a)=\sum_{P\in S_n}\operatorname{sg}\begin{pmatrix}1 & 2 & \cdots & n-1 & n\\ P(1) & P(2) & \cdots & P(n-1) & P(n)\end{pmatrix}a_1^{P(1)}a_2^{P(2)}\cdots a_{n-1}^{P(n-1)}a_n^{P(n)}. \tag{4.3}$$

Hay que examinar la fórmula con detenimiento. Recordemos que S_n es el grupo de las permutaciones de $\{1,2,\dots,n\}$, luego la expresión tiene tantos sumandos como permutaciones. Dado que $\#S_n=n!$ vemos que hay $2!=2$ y $3!=9$ para $n=2,3$, pero el número de sumandos aumenta a gran velocidad si crece n porque $4!=24$ y $5!=120$. Ya se empieza a ver porqué la fórmula, sin un conocimiento importante de sus propiedades, es inmanejable. Para cada permutación aparece un sumando con su signatura ±1 en primer lugar y luego un producto de un elemento de la primera columna por otro de la segunda... por otro de la última, según marca la permutación P. Como ejemplo con $n=4$,

$$P=\begin{pmatrix}1&2&3&4\\3&4&1&2\end{pmatrix}=(1,3)(2,4),\quad \operatorname{sg}(P)=(-1)^2=1,\quad \operatorname{sg}(P)a_1^{P(1)}a_2^{P(2)}a_3^{P(3)}a_4^{P(4)}=1\cdot a_1^3a_2^4a_3^1a_4^2.$$

Dicho con palabras: para $a_1^{P(1)}a_2^{P(2)}a_3^{P(3)}a_4^{P(4)}$ se toma de la columna 1 el elemento 3, de la columna 2 el elemento 4, de la columna 3 el elemento 1, y de la columna 4 el elemento 2; se multiplican y se pone el signo de $\operatorname{sg}(P)$. Por supuesto, si se quiere escribir (4.3) hay que hacer esto 23 veces más para las 23 permutaciones que faltan. La fórmula en los casos $n=2,3$ es (el lector debe verificarlo para familiarizarse con (4.3))

$$\begin{vmatrix} a_1^1 & a_2^1\\ a_1^2 & a_2^2\end{vmatrix}=\operatorname{sg}\begin{pmatrix}1&2\\1&2\end{pmatrix}a_1^1a_2^2+\operatorname{sg}\begin{pmatrix}1&2\\2&1\end{pmatrix}a_1^2a_2^1=a_1^1a_2^2-a_1^2a_2^1,$$

$$\begin{aligned}\begin{vmatrix} a_1^1 & a_2^1 & a_3^1\\ a_1^2 & a_2^2 & a_3^2\\ a_1^3 & a_2^3 & a_3^3\end{vmatrix}&=\operatorname{sg}\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}a_1^1a_2^2a_3^3+\operatorname{sg}\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}a_1^3a_2^1a_3^2+\operatorname{sg}\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}a_1^2a_2^3a_3^1\\&+\operatorname{sg}\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}a_1^1a_2^3a_3^2+\operatorname{sg}\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}a_1^3a_2^2a_3^1+\operatorname{sg}\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}a_1^2a_2^1a_3^3\\&=a_1^1a_2^2a_3^3+a_1^3a_2^1a_3^2+a_1^2a_2^3a_3^1-a_1^1a_2^3a_3^2-a_1^3a_2^2a_3^1-a_1^2a_2^1a_3^3.\end{aligned}$$

Aunque el lector podrá recordar esta última fórmula con la **regla de Sarrús** pensamos que cada vez verá más claro que el manejo directo de (4.3) es muy poco operativo y solo valdrá para probar propiedades teóricas. Un último ejemplo:

$$\begin{vmatrix}1&2\\3&4\end{vmatrix}=-2,\quad \begin{vmatrix}1&2&3\\4&5&6\\7&8&9\end{vmatrix}=0,\quad \begin{vmatrix}1&2&3&4\\5&6&7&8\\9&10&11&12\\13&14&15&16\end{vmatrix}=0$$

[5] Buena parte de la teoría vale para un anillo conmutativo con unidad (todos los cuerpos lo son) pero para no sobrecargar al lector con nuevas ideas le permitimos que se limite a cuerpos.

[6] La notación $|a|$ es algo peligrosa porque si $n=1$ y $\Bbbk=\mathbb{R},\mathbb{C}$ se podría confundir con el módulo o valor absoluto del número a. En la práctica, esta confusión es poco probable.

y siguen siendo cero los determinantes de las matrices generalizadas a dimensiones $n \times n$. Intentar verificarlo directamente es un enorme trabajo, pero luego se verá que es fácil hacerlo con las propiedades que iremos mostrando del determinante. El siguiente teorema da una gran simplificación.

Teorema 78 *Sea a una matriz triangular inferior; o sea $a_j^i = 0$ si es $i < j$. Entonces,* $\det(a) = a_1^1 a_2^2 \cdots a_n^n$, *el producto de los términos de la diagonal principal.*

Demostración. Si $P \neq \mathrm{id}$ vimos en el problema 165 que hay un j tal que $P(j) < j$. En el correspondiente sumando de (165) se tiene con 0 en el lugar de $a_j^{P(j)}$ que

$$a_1^{P(1)} \cdot \cdots \cdot a_j^{P(j)} \cdot \cdots \cdot a_n^{P(n)} = a_1^{P(1)} \cdot \cdots \cdot 0 \cdot \cdots \cdot a_n^{P(n)} = 0.$$

Solo es $\neq 0$ el sumando correspondiente a $P = \mathrm{id}$ con lo que $\det(a) = a_1^1 a_2^2 \cdots a_n^n$. ♣

Imaginamos otro teorema análogo si a es triangular superior. Lo hay.

Teorema 79 *Una matriz y su traspuesta tienen el mismo determinante.*

Demostración. Sea $b = a^\top$, luego $b_j^i = a_i^j$. Estudiemos

$$\det(a) = \sum_{P \in S_n} \mathrm{sg}(P)\, a_1^{P(1)} \cdots a_n^{P(n)}, \quad \det(b) = \sum_{Q \in S_n} \mathrm{sg}(Q)\, b_1^{Q(1)} \cdots b_n^{Q(n)} = \sum_{Q \in S_n} \mathrm{sg}(Q)\, a_{Q(1)}^1 \cdots a_{Q(n)}^n.$$

La clave de la demostración está en observar que $\det(a)$ y $\det\left(a^\top\right) = \det(b)$ tienen los mismos sumandos (¡no lo parece!) y por tanto los dos determinantes son iguales. ¿Cómo se corresponden los sumandos? Elegimos una permutación P y mostramos que el sumando que corresponde a P en $\det(a)$ es el mismo que corresponde a $Q = P^{-1}$ en $\det\left(a^\top\right)$; o sea,

$$\mathrm{sg}(P)\, a_1^{P(1)} \cdots a_n^{P(n)} = \mathrm{sg}(Q)\, a_{Q(1)}^1 \cdots a_{Q(n)}^n \text{ siendo } Q = P^{-1}.$$

Como $Q = P^{-1}$ se tiene $\mathrm{sg}(P) = \mathrm{sg}(Q)$ y basta ver que $a_1^{P(1)} \cdots a_n^{P(n)} = a_{Q(1)}^1 \cdots a_{Q(n)}^n$ cuando $Q = P^{-1}$. ¡Cuidado! *No se va a probar* que sea $a_1^{P(1)} = a_{Q(1)}^1, \ldots, a_n^{P(n)} = a_{Q(n)}^n$ sino que $a_1^{P(1)} \cdots a_n^{P(n)} = a_{Q(1)}^1 \cdots a_{Q(n)}^n$ porque aparecen los mismos números *solo que en distinto orden.* Esto es fácil de comprobar (quizás no de intuir) ya que, como $Q \circ P(i) = i$,

$$a_i^{P(i)} = a_{Q \circ P(i)}^{P(i)} = a_{Q(P(i))}^{P(i)}.$$

Dicho con palabras: el $i-$ésimo factor en $a_1^{P(1)} \cdots a_n^{P(n)}$ es el $P(i)-$ésimo factor en $a_{Q(1)}^1 \cdots a_{Q(n)}^n$. Esto acaba la demostración. ♣

Se puede ilustrar el punto esencial de la demostración para $n = 5$ con

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 3 & 4 \end{pmatrix}, \qquad Q = P^{-1} = \begin{pmatrix} 2 & 1 & 5 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix},$$

$$a_1^{P(1)} a_2^{P(2)} a_3^{P(3)} a_4^{P(4)} a_5^{P(5)} = a_1^2 a_2^1 a_3^5 a_4^3 a_5^4, \qquad a_{Q(1)}^1 a_{Q(2)}^2 a_{Q(3)}^3 a_{Q(4)}^4 a_{Q(5)}^5 = a_2^1 a_1^2 a_4^3 a_5^4 a_3^5.$$

El factor a_3^5, que a la izquierda está en tercer lugar por el subíndice 3, está a la derecha en quinto lugar pues $P(3) = 5$.

Como la traspuesta de una matriz triangular inferior es triangular superior, sale con los teoremas 79 y 78 que el determinante de una matriz triangular (superior o inferior) es el producto de los términos de su diagonal. La demostración de $\det(a) = \det\left(a^\top\right)$ nos da como información adicional (véase el principio de la demostración del teorema 79) que

$$\det(a) = \sum_{P \in S_n} \mathrm{sg}(P)\, a_1^{P(1)} \cdots a_n^{P(n)} = \sum_{Q \in S_n} \mathrm{sg}(Q)\, a_{Q(1)}^1 \cdots a_{Q(n)}^n.$$

Por tanto, al definir el determinante con un sumatorio, da lo mismo que se elija que los productos de las a_j^i vayan por orden creciente de índices o de superíndices. *Hay dos definiciones de determinante con sumatorios pero son iguales.*

4.3. El determinante como función de los vectores columna

Interesa cambiar ligeramente el punto de vista e identificar $a \in \mathbb{k}^{n\times n}$ con la sucesión de sus columnas $(a_1,\ldots,a_n)$. Tenemos entonces una función D identificable con det y definida por

$$D : \mathbb{k}^n \times \mathbb{k}^n \times \ldots \times \mathbb{k}^n \longrightarrow \mathbb{k}, \qquad D(a_1,\ldots,a_n) = \det(a).$$

Esto facilita el enunciado del siguiente teorema fundamental.[7] Previamente introducimos una notación que ahorra espacio. La sucesión $(a_1,\ldots,\widehat{a_j},\ldots,a_n)$ representa $(a_1,\ldots,a_{j-1},a_{j+1},\ldots,a_n)$; es decir, el tejadillo indica que se prescinde de ese elemento. Por otra parte, $\left(a_1,\ldots,\overset{(j)}{x},\ldots,a_n\right)$ indica que x está en el lugar j.

Teorema 80 *El determinante, como función de las columnas; o sea, la función D, verifica*

1. *Para toda elección de $(a_1,\ldots,\widehat{a_j},\ldots,a_n)$, la función de $\mathbb{k}^n$ en $\mathbb{k}$ dada por*

$$f(x) = D(a_1,\ldots,a_{j-1},x,a_{j+1},\ldots,a_n) = D\left(a_1,\ldots,\overset{(j)}{x},\ldots,a_n\right)$$

es una función lineal; lo que significa que

$$D\left(a_1,\ldots,\overset{(j)}{x+y},\ldots,a_n\right) = D\left(a_1,\ldots,\overset{(j)}{x},\ldots,a_n\right) + D\left(a_1,\ldots,\overset{(j)}{y},\ldots,a_n\right),$$

$$D\left(a_1,\ldots,\overset{(j)}{\lambda x},\ldots,a_n\right) = \lambda D\left(a_1,\ldots,\overset{(j)}{x},\ldots,a_n\right).$$

2. *Si a tiene dos columnas iguales, $D(a_1,\ldots,a_n) = 0$.*
3. *El determinante de la matriz unidad I_n vale 1. De modo equivalente, como I_n se identifica con $(e_1,\ldots,e_n)$, siendo $(e_1,\ldots,e_n)$ la base estándar, tenemos que $D(e_1,\ldots,e_n) = 1$*

Demostración. Comprobamos de **1** la parte de la suma solamente. Al ser $(x+y)^k = x^k + y^k$,

$$f(x+y) = \sum_{P\in S_n} \operatorname{sg}(P)\, a_1^{P(1)} \cdots \overset{(j)}{(x+y)^{P(j)}} \cdots a_n^{P(n)} = \sum_{P\in S_n} \operatorname{sg}(P)\, a_1^{P(1)} \cdots \overset{(j)}{\left(x^{P(j)} + y^{P(j)}\right)} \cdots a_n^{P(n)}$$

$$= \sum_{P\in S_n} \operatorname{sg}(P)\, a_1^{P(1)} \cdots \overset{(j)}{x^{P(j)}} \cdots a_n^{P(n)} + \sum_{P\in S_n} \operatorname{sg}(P)\, a_1^{P(1)} \cdots \overset{(j)}{y^{P(j)}} \cdots a_n^{P(n)} = f(x) + f(y).$$

Para **2** tomamos $r < s$ y suponemos $a_r = a_s$. La clave para probar que $D(a_1,\ldots,a_n) = 0$ es mostrar que si T es la trasposición (r,s) se puede emparejar cada permutación par P con la permutación impar $Q = P\circ T$ y entonces $a_1^{P(1)}\cdots a_n^{P(n)} = a_1^{Q(1)}\cdots a_n^{Q(n)}$. Si lo aceptamos momentáneamente y abreviamos llamando H a estos productos, en la fórmula del determinante aparecerán $\operatorname{sg}(P)H + \operatorname{sg}(Q)H = \operatorname{sg}(P)(H-H) = 0$, luego $D(a_1,\ldots,a_n) = 0$ porque es suma de $n!/2$ ceros. Probemos entonces que $a_1^{P(1)}\cdots a_n^{P(n)} = a_1^{Q(1)}\cdots a_n^{Q(n)}$. Hay dos observaciones **(a)** $a_r = a_s$ implica que a_r^h es sustituible por a_s^h cualquiera que sea $h = 1,\ldots,n$ y podemos en tanto que subíndices intercambiar r y s; y **(b)** P y $Q = P\circ T$ se relacionan por $Q(j) = j$ si $j \neq r,s$, $Q(r) = P(s)$ y $Q(s) = P(r)$. Por **(b),** en los productos concretamos solo los factores en lugar r y s porque los otros no varían. Con esto, poniendo un factor debajo del otro para ver mejor las sustituciones,

$$\begin{aligned} &\cdots a_r^{P(r)} \cdots a_s^{P(s)} \cdots \\ \overset{1}{=} &\cdots a_r^{Q(s)} \cdots a_s^{Q(r)} \cdots \\ \overset{2}{=} &\cdots a_s^{Q(s)} \cdots a_r^{Q(r)} \cdots \\ \overset{3}{=} &\cdots a_r^{Q(r)} \cdots a_s^{Q(s)} \cdots \end{aligned}$$

[7] De momento marcamos la diferencia entre D y det, pero según tomemos familiaridad no haremos distinciones, escribiendo por ejemplo $\det(a_1,\ldots,a_n)$.

como queríamos demostrar. En $\overset{1}{=}$ se ha usado $Q(r) = P(s)$ y $Q(s) = P(r)$ de **(b)**; en $\overset{2}{=}$ se ha usado **(a)**; y en $\overset{3}{=}$ la conmutatividad del producto en $\mathbb{k}$ se pasa el factor $a_s^{Q(s)}$ del lugar r al lugar s y el factor $a_r^{Q(r)}$. Dicho con palabras: el probar que $\cdots a_r^{P(r)} \cdots a_s^{P(s)} \cdots = \cdots a_r^{Q(r)} \cdots a_s^{Q(s)} \cdots$ es como probar que en la expresión inicial de $a_1^{P(1)} \cdots a_n^{P(n)}$ se pueden sustituir en los factores r y s los superíndices $P(r)$ y $P(s)$ por $Q(r)$ y $Q(s)$ sin que varíe la expresión. Obviamente, se necesita $a_r = a_s$ para esto.

El punto **3** es trivial. ♣

Damos al lector para la demostración de **2** un ejemplo concreto con $n = 4$, $r = 2$ y $s = 3$,

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix},\ Q = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix},$$

$$a = \begin{pmatrix} 1 & p & p & 5 \\ 2 & q & q & 6 \\ 3 & r & r & 7 \\ 4 & s & s & 8 \end{pmatrix}, \qquad \begin{cases} a_1^{P(1)} a_2^{P(2)} a_3^{P(3)} a_4^{P(4)} = a_1^2 a_2^1 a_3^4 a_4^3 = 2ps7 \\ a_1^{Q(1)} a_2^{Q(2)} a_3^{Q(3)} a_4^{Q(4)} = a_1^2 a_2^4 a_3^1 a_4^3 = 2sp7 \end{cases},$$

y el cálculo abstracto se concreta en

$$a_1^{P(1)} a_2^{P(2)} a_3^{P(3)} a_4^{P(4)} = a_1^2 a_2^1 a_3^4 a_4^3 \overset{1}{=} a_1^2 a_2^1 a_3^4 a_4^3 \overset{2}{=} a_1^2 a_3^1 a_2^4 a_4^3 \overset{3}{=} a_1^2 a_2^4 a_3^1 a_4^3 = a_1^{Q(1)} a_2^{Q(2)} a_3^{Q(3)} a_4^{Q(4)}.$$

Teorema 81 *Si $Q \in S_n$ se tiene la relación $D\left(a_{Q(1)}, \ldots, a_{Q(n)}\right) = \operatorname{sg}(Q) D(a_1, \ldots, a_n)$. Verbalmente: si permutamos las columnas de a, el determinante de la nueva matriz es el mismo si la permutación es par y es el opuesto en signo si la permutación es impar.*

Demostración. Empezamos con el caso de una trasposición $Q = T = (r, s)$ con $r < s$. Definimos

$$f : \mathbb{k}^n \times \mathbb{k}^n \longrightarrow \mathbb{k}, \quad f(x, y) = D\left(a_1, \ldots, \overset{(r)}{x}, \ldots, \overset{(s)}{y}, \ldots, a_n\right);$$

es decir, poniendo los vectores columna x e y como nuevas columnas r y s de a. Hacemos $x = y = a_r + a_s$ y por **1** y **2** en el teorema anterior,

$$0 = f(a_r + a_s, a_r + a_s) = f(a_r, a_r) + f(a_r, a_s) + f(a_s, a_r) + f(a_s, a_s) = f(a_r, a_s) + f(a_s, a_r).$$

Esto nos lleva a

$$\begin{aligned} D\left(a_{T(1)}, \ldots, \overset{(r)}{a_{T(r)}}, \ldots, \overset{(s)}{a_{T(s)}}, \ldots, a_n\right) &= D\left(a_1, \ldots, \overset{(r)}{a_s}, \ldots, \overset{(s)}{a_r}, \ldots, a_n\right) = f(a_s, a_r) \\ &= -f(a_r, a_s) = -D\left(a_1, \ldots, \overset{(r)}{a_r}, \ldots, \overset{(s)}{a_s}, \ldots, a_n\right) \\ &= \operatorname{sg}(T) D\left(a_1, \ldots, \overset{(r)}{a_r}, \ldots, \overset{(s)}{a_s}, \ldots, a_n\right). \end{aligned}$$

La demostración para Q arbitraria se hace por inducción sobre el número de trasposiciones h con las que se puede factorizar Q. El teorema es cierto como acabamos de ver si $h = 1$. Si lo suponemos cierto para $h-1$, escribimos $Q = T_h \circ T_{h-1} \circ \cdots \circ T_1 = T_h \circ L$ con $L = T_{h-1} \circ \cdots T_1$. Entonces, si $b = \left(a_{L(1)}, \ldots, a_{L(n)}\right)$,

$$\begin{aligned} D\left(a_{Q(1)}, \ldots, a_{Q(n)}\right) &= D\left(b_{T_h(1)}, \ldots, b_{T_h(n)}\right) \overset{1}{=} \operatorname{sg}(T_h) D(b_1, \ldots, b_n) \\ &\overset{2}{=} \operatorname{sg}(T_h) \operatorname{sg}(L) D(a_1, \ldots, a_n) \overset{3}{=} \operatorname{sg}(Q) D(a_1, \ldots, a_n). \end{aligned}$$

Se usa en $\overset{1}{=}$ el resultado ya probado para una sola trasposición (aplicado a b; no a a); en $\overset{2}{=}$ la hipótesis inductiva, y en $\overset{3}{=}$ que la signatura de la composición es el producto de las signaturas.

Si se quiere hacer la demostración de modo más intuitivo,

$$\begin{aligned} D\left(a_{Q(1)}, \ldots, a_{Q(n)}\right) &= D\left(a_{T_h \circ T_{h-1} \circ \cdots \circ T_1(1)}, \ldots, a_{T_h \circ T_{h-1} \circ \cdots \circ T_1(n)}\right) \\ &= (-1) D\left(a_{T_{h-1} \circ \cdots \circ T_1(1)}, \ldots, a_{T_{h-1} \circ \cdots \circ T_1(n)}\right) \\ &= (-1)^2 D\left(a_{T_{h-2} \circ \cdots \circ T_1(1)}, \ldots, a_{T_{h-2} \circ \cdots \circ T_1(n)}\right) \\ &\vdots \\ &= (-1)^{h-1} D\left(a_{T_1(1)}, \ldots, a_{T_1(n)}\right) = (-1)^h D(a_1, \ldots, a_n). \end{aligned}$$

Como $(-1)^h = \mathrm{sg}(Q)$, el teorema está probado. ♣

Los dos teoremas precedentes tiene unas versiones análogas sustituyendo "columna(s)" por "fila(s)" en el enunciado. Si identificamos a con $\left(a^1, \ldots, a^n\right)$, la sucesión de sus filas, podemos definir

$$D' : \mathbb{k}^{1\times n} \times \ldots \times \mathbb{k}^{1\times n} \longrightarrow \mathbb{k}, \qquad D'\left(a^1, \ldots, a^n\right) = \det(a).$$

Las segundas versiones resultan utilizando el teorema 79 que dice que una matriz y su traspuesta tienen el mismo determinante. Si se quiere probar, por ejemplo, que $\det(a) = 0$ si a tiene dos filas iguales, se hace $b = a^\top$, con lo que b tiene dos columnas iguales y $\det(a) = \det(b) = 0$. Es rutinario pero muy pesado dar la "versión filas" de todos los teoremas, pero hay que saberlas y saber manejarlas. Nosotros daremos tan solo la "versión columnas".

Teorema 82 *Si en a se añade a la columna j una combinación lineal de las restantes, el determinante de la nueva matriz es el mismo. Si una columna de b es combinación lineal de las restantes,* $\det(b) = 0$

Demostración. Sea c la nueva matriz. Se tiene con **1** en el teorema 80,

$$\begin{aligned}\det(c) &= D\left(a_1, \ldots, a_{j-1}, \lambda^1 a_1 + \ldots \lambda^{j-1} a_{j-1} + a_j + \lambda^{j+1} a_{j+1} \ldots + \lambda^n a_n, a_{j+1}, \ldots . a_n\right)\\ &= D(a_1, \ldots, a_{j-1}, a_j, a_{j+1}, \ldots . a_n) + \sum_{k\neq j} \lambda^k D(a_1, \ldots, a_{j-1}, a_k, a_{j+1}, \ldots . a_n).\end{aligned}$$

Obsérvese que como $k \neq j$, en $(a_1, \ldots, a_{j-1}, a_k, a_{j+1}, \ldots . a_n)$ hay dos columnas a_k. Por tanto, según **2** en el teorema 81, todos estos determinantes son cero y queda $\det(c) = \det(a)$.

La segunda afirmación tiene demostración muy similar. Supongamos para no complicar demasiado la notación que sea $b_1 = \sum_{k=2}^n \lambda^k b_k$. Entonces

$$\det(b) = D\left(\sum_{k=2}^n \lambda^k b_k, b_2, \ldots, b_n\right) = \sum_{k=2}^n \lambda^k D(b_k, b_2, \ldots, b_n) = 0$$

pues al ser $k \geq 2$, en cada $(b_k, b_2, \ldots, b_n)$ está repetido b_k. ♣

Una estrategia general para calcular determinantes es sustituir la matriz a por otra b a la que se ha llegado por operaciones columna, como tantas veces se ha hecho en los capítulos anteriores. Si b se obtiene multiplicando por λ una columna de a, $\det(b) = \lambda \det(a)$. Si b se obtiene permutando las columnas de a según la permutación P, se tiene $\det(b) = \mathrm{sg}(P)\det(a)$. Si b se obtiene sumando a una columna a_j de a múltiplos de las restantes columnas, $\det(b) = \det(a)$. Sabemos que eligiendo adecuadamente estas operaciones podemos convertir la matriz a en una matriz triangular superior o inferior b y, ya dijimos en el teorema 78 que las matrices triangulares tienen como fácil determinante el producto de elementos de la diagonal principal. El cálculo requiere práctica pero es efectivo. Daremos luego más procedimientos para calcular determinantes.

Problema 166 *Calcular los determinantes de las matrices*

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 1\\ 0 & 2 & h\\ -1 & h & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 2 & 2\\ 0 & 0 & 3 & 3 & 3\\ 0 & 4 & 4 & 4 & 4\\ 5 & 5 & 5 & 5 & 5 \end{pmatrix}. \blacklozenge$$

Solución. En el primer caso[8] se sustituye en las columnas 3 y 4 en la forma

$$a = (a_1, a_2, a_3, a_4, a_5) \to b = (a_1, a_2, a_3 - a_4, a_4 - a_5, a_5) = \begin{pmatrix} 1 & 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

queda b triangular y es obvio que $\det(b) = 1^5 = 1$.

[8] Cuando veamos el cálculo por cajas un poco más adelante, se puede calcular también con este procedimiento.

El segundo determinante se calcula con

$$\begin{vmatrix} 1 & 3 & 1 \\ 0 & 2 & h \\ -1 & h & 2 \end{vmatrix} \overset{1}{\to} \begin{vmatrix} 1 & 3 & 0 \\ 0 & 2 & h \\ -1 & h & 3 \end{vmatrix} \overset{2}{\to} \begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & h \\ -1 & h+3 & 3 \end{vmatrix} \overset{3}{\to} \begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ -1 & h+3 & 3-(h+3)\frac{h}{2} \end{vmatrix} = 6-(h+3)h.$$

En $\overset{1}{\to}$ se resta a la columna 3 la columna 1; en $\overset{2}{\to}$ se resta a la columna 2 el triple de la primera; y en $\overset{3}{\to}$ se resta a la columna 3 la columna 2 por $h/2$. Se puede hacer con otras operaciones columna o con operaciones fila.

Es mejor abordar el tercer caso en abstracto. Si b se obtiene de a como $b=(a_n,a_{n-1},\ldots,a_2,a_1)$; o sea, poniendo las columnas en orden inverso, hemos aplicado una permutación

$$P=\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix};\ \text{en nuestro caso } P=\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix}=(1,5)\circ(2,4)$$

con signatura $(-1)^2=1$. Por tanto, en nuestro caso,

$$\begin{vmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 3 & 3 & 3 \\ 0 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 \end{vmatrix} = 1\cdot \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 4 & 0 \\ 5 & 5 & 5 & 5 & 5 \end{vmatrix} = 5!\ .$$

En el caso general, que no se ha pedido pero queda muy a tiro con cierta habilidad técnica. Vemos que según sea $n=2k$ o $n=2k+1$ se puede factorizar P como

$$P=\begin{pmatrix} 1 & 2 & \cdots & k & k+1 & \cdots & 2k-1 & 2k \\ 2k & 2k-1 & \cdots & k+1 & k & \cdots & 2 & 1 \end{pmatrix}=(1,2k)\circ(2,2k-1)\circ\cdots\circ(k,k+1),$$

$$P=\begin{pmatrix} 1 & 2 & \cdots & k & k+1 & k+2 & \cdots & 2k & 2k+1 \\ 2k+1 & 2k & \cdots & k+2 & k+1 & k-1 & \cdots & 2 & 1 \end{pmatrix}=(1,2k+1)\circ(2,2k)\circ\cdots\circ(k,k+2).$$

En ambos casos $\operatorname{sg}(P)=(-1)^k$, luego $D(a_n,a_{n-1},\ldots,a_2,a_1)=(-1)^k D(a_1,a_2,\ldots,a_{n-1},a_n)$ y $n=2k$ o $2k+1$. ♦

Insistimos en que muchos determinantes se calculan de manera mucho más eficaz viéndolos como funciones de las filas o columnas de la matriz y buscando simplificaciones en determinantes intermedios nulos porque tienen dos columnas iguales. Un ejemplo

Problema 167 *Calcular los determinantes*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},\quad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix},\quad \begin{pmatrix} 1 & 2 & \cdots & n \\ 1+n & 2+n & \cdots & n+n \\ 1+2n & 2+2n & \cdots & n+2n \\ \vdots & \vdots & \ddots & \vdots \\ 1+(n-1)n & 2+(n-1)n & \cdots & n+(n-1)n \end{pmatrix}.\ ♦$$

Solución. Todos valen cero. Si u es el vector fila $(1,2,\ldots,n)$ y v el vector fila $(n,n,\ldots,n)$ se tiene

$$\det(a)=D'(u,u+v,u+2v,\ldots,u+(n-1)v).$$

Al desarrollar $D'(u,u+v,u+2v,\ldots,u+(n-1)v)$ salen muchos sumandos pero en todos ellos o bien u está en al menos dos columnas o bien múltiplos de v están en al menos dos columnas. Por tanto todos estos sumandos son cero. Si se quiere ver esto en el caso de la matriz 3×3,

$$\begin{aligned}\det(a)&=D'(u,u+v,u+2v)=D'(u,u,u+2v)+D'(u,v,u+2v)=D'(u,v,u+2v)\\&=D'(u,v,u)+D'(u,v,2v)=0.\end{aligned}$$

Si se quiere hacer de modo riguroso puede probarse $D\left(u,u+\lambda^2v,u+\lambda^3v,\ldots,u+\lambda^{n-1}v\right)=0$ por inducción para $n\geq 3$ cualquiera que sean las λ. ♦

Problema 168 *Calcular para* $\Bbbk = \mathbb{R}$ *los determinantes con operaciones fila o columna*

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 3 & 6 & 8 & 11 \\ 7 & 13 & 20 & 26 \\ 31 & 23 & 55 & 46 \end{vmatrix}, \qquad \begin{vmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{vmatrix}.$$

Puede irse todavía más rápido combinando operaciones fila y columna o con técnicas que luego veremos (desarrollo por una fila o columna, desarrollo por cajas).

Teorema 83 *Dadas* $a, b \in \Bbbk^{n\times n}$ *se tiene* $\det(ab) = \det(a)\det(b)$ *(el determinante del producto es el producto de los determinantes de los factores).*

Demostración. Sea $c = ab$. Las columnas c_j pueden calcularse como combinación de las columnas de a. Más concretamente,

$$c_j^i = \sum a_k^i b_j^k \text{ para } 1 \le i, j \le n \text{ da } n \text{ ecuaciones vectoriales } c_j = \sum_{k=1}^{n} b_j^k a_k, \text{ para } 1 \le j \le n.$$

Tenemos entonces

$$\det(c) = D(c_1, \ldots, c_n) = D\left(\sum_{k_1=1}^{n} b_1^{k_1} a_{k_1}, c_2, \ldots, c_n\right) = \sum_{k_1=1}^{n} b_1^{k_1} D(a_{k_1}, c_2, \ldots, c_n).$$

Si sustituimos c_2 como combinación lineal de las a_{k_2},

$$\det(c) = \sum_{k_1=1}^{n} b_1^{k_1}\left(\sum_{k_2=1}^{n} b_2^{k_2} D(a_{k_1}, a_{k_2}, c_3, \ldots, c_n)\right) = \sum_{k_1,k_2=1}^{n} b_1^{k_1} b_2^{k_2} D(a_{k_1}, a_{k_2}, c_3, \ldots, c_n)$$

Así, sustituyendo sucesivamente $c_3, \ldots$ hasta c_n como combinación lineal de las columnas de a se llega a

$$\det(c) = \det(ab) = \sum_{k_1,k_2,\ldots,k_n=1}^{n} b_1^{k_1} b_2^{k_2} \ldots b_n^{k_n} D(a_{k_1}, a_{k_2}, \ldots, a_{k_n}).$$

Hay una enorme cantidad de sumandos (de hecho n^n) pues se elige cualquier sucesión $(k_1, \ldots, k_n)$ de los números $1, 2, \ldots, n$. Esta sucesión $(k_1, \ldots, k_n)$ puede ser con términos iguales. Sin embargo muchos sumandos son nulos; precisamente aquellos en los que hay repeticiones en $(k_1, \ldots, k_n)$ pues por **2** en el teorema 80 se obtiene que $D(a_{k_1}, a_{k_2}, \ldots, a_{k_n}) = 0$. Solo hay que considerar los sumandos con $(a_{k_1}, a_{k_2}, \ldots, a_{k_n})$ sin repeticiones. Si P es la permutación

$$P = \begin{pmatrix} 1 & 2 & \cdots & n \\ k_1 & k_2 & \cdots & k_n \end{pmatrix}$$

podemos cambiar la notación y esto nos lleva a algo más familiar

$$\begin{aligned} \det(c) &= \sum_{P\in S_n} b_1^{P(1)} b_2^{P(2)} \ldots b_n^{P(n)} D\left(a_{P(1)}, a_{P(2)}, \ldots, a_{P(n)}\right) \\ &= \sum_{P\in S_n} b_1^{P(1)} b_2^{P(2)} \ldots b_n^{P(n)} \operatorname{sg}(P) D(a_1, a_2, \ldots, a_n) = \det(b)\det(a), \end{aligned}$$

usándose el teorema 84 en el salto de línea. ♣

Teorema 84 *Una matriz* a *es invertible si y solo si* $\det(a) \neq 0$, *en cuyo caso,* $\det\left(a^{-1}\right) = 1/\det(a)$.

Demostración. Si existe a^{-1} es $aa^{-1} = I$, luego $\det(a)\det\left(a^{-1}\right) = \det(I) = 1$. Esto da sucesivamente que no puede ser $\det(a) = 0$ y que $\det\left(a^{-1}\right) = 1/\det(a)$.

Recíprocamente, supongamos que $\det(a) \neq 0$ pero a no es invertible. Su rango r es $r < n$ y $r = \dim \lg(a_1, \ldots, a_n)$, el subespacio de $\Bbbk^n$ generado por sus columnas. Debe haber una columna, digamos a_j, que sea combinación de las demás y, según el teorema 82 esto implica $\det(a) = 0$. Contradicción. ♣

Este teorema no da una fórmula explícita para conocer a^{-1}, pero para muchas cuestiones teóricas el teorema 84 es fundamental. Sabemos del capítulo anterior que a invertible equivale a $\operatorname{rg}(a) = n$ y a la independencia de sus filas o columnas, luego el teorema 84 indica que $\det(a) \neq 0$ equivale a estas circunstancias.

El teorema que cierra la sección dice resumidamente que si $\Delta : \Bbbk^n \times \ldots \times \Bbbk^n \to \Bbbk$ verifica las tres operaciones **1**,**2** y **3** en el teorema 80 es porque es el determinante; es decir, $D = \Delta$. Esto es más que una curiosidad porque permite probar ciertas fórmulas con poco esfuerzo como luego veremos.

Teorema 85 *Sea $\Delta : \Bbbk^n \times \ldots \times \Bbbk^n \to \Bbbk$ una función que cumple las mismas propiedades **1,2,3** que cumple D en el teorema 80; es decir,*

1. *Para toda elección de $(a_1, \ldots, \widehat{a_j}, \ldots, a_n)$, la función f de $\Bbbk^n$ en $\Bbbk$ dada por*

$$f(x) = \Delta(a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n) = \Delta\left(a_1, \ldots, \overset{(j)}{x}, \ldots, a_n\right)$$

   es lineal.

2. *Si en $(a_1, \ldots, a_n)$ hay dos columnas iguales, $\Delta(a_1, \ldots, a_n) = 0$.*

3. *Para la base estándar, $\Delta(e_1, \ldots, e_n) = 1$.*

*Entonces, $\Delta(a_1, \ldots, a_n) = D(a_1, \ldots, a_n)$; o sea, Δ es el determinante como función de columnas. Si Δ solo cumple **1 y 2** se tiene que $\Delta(a_1, \ldots, a_n) = \Delta(e_1, \ldots, e_n)\det(a)$; o sea, Δ es un múltiplo del determinante con factor de proporcionalidad $\Delta(e_1, \ldots, e_n)$.*

Demostración. Observamos primero que en la demostración del teorema 81 se han usado solamente las propiedades **1** y **2**, luego Δ verifica también $\Delta\left(a_{Q(1)}, \ldots, a_{Q(n)}\right) = \operatorname{sg}(Q)\Delta(a_1, \ldots, a_n)$ para toda permutación Q. Con esto podemos llegar a una fórmula para Δ realizando un cálculo similar al hecho en el teorema 83. Con la base estándar escribimos

$$a_1 = \sum_{i_1=1}^{n} a_1^{i_1} e_{i_1}, \quad a_2 = \sum_{i_2=1}^{n} a_2^{i_2} e_{i_2}, \quad \ldots \quad , a_n = \sum_{i_3=1}^{n} a_n^{i_n} e_{i_n}$$

y tan solo con las propiedades **1** y **2**, junto con $\Delta\left(a_{Q(1)}, \ldots, a_{Q(n)}\right) = \operatorname{sg}(Q)\Delta(a_1, \ldots, a_n)$ que se sigue de ellas,

$$\begin{aligned}
\Delta(a_1, \ldots, a_n) &= \sum_{i_1=1}^{n} a_1^{i_1} \Delta\left(e_{i_1}, \sum_{i_2=1}^{n} a_2^{i_2} e_{i_2}, \sum_{i_3=1}^{n} a_3^{i_3} e_{i_3}, \ldots, \sum_{i_3=1}^{n} a_n^{i_n} e_{i_n}\right) \\
&= \sum_{i_1=1}^{n} a_1^{i_1} \left(\sum_{i_2=1}^{n} a_2^{i_2} \Delta\left(e_{i_1}, e_{i_2}, \sum_{i_3=1}^{n} a_3^{i_3} e_{i_3}, \ldots, \sum_{i_3=1}^{n} a_n^{i_n} e_{i_n}\right)\right) \\
&= \sum_{i_1, i_2=1}^{n} a_1^{i_1} a_2^{i_2} \Delta\left(e_{i_1}, e_{i_2}, \sum_{i_3=1}^{n} a_3^{i_3} e_{i_3}, \ldots, \sum_{i_3=1}^{n} a_n^{i_n} e_{i_n}\right) \\
&\vdots \\
&= \sum_{i_1, i_2, \ldots i_n=1}^{n} a_1^{i_1} a_2^{i_2} \ldots a_n^{i_n} \Delta(e_{i_1}, e_{i_2}, e_{i_3}, \ldots, e_{i_n}).
\end{aligned}$$

Los sumandos en donde la sucesión $(i_1, i_2, \ldots, i_n)$ tiene índices repetidos son nulos por **2**. Solo quedan aquellas con índices distintos. A cada una de estas $(i_1, i_2, \ldots, i_n)$ le asignamos la permutación

$$P = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}$$

y con este cambio de notación entramos en un terreno más familiar,

$$\begin{aligned}
\Delta(a_1, \ldots, a_n) &= \sum_{P \in S_n} a_1^{P(1)} a_2^{P(2)} \ldots a_n^{P(n)} \Delta\left(e_{P(1)}, e_{P(2)}, \ldots, e_{P(n)}\right) \\
&= \sum_{P \in S_n} a_1^{P(1)} a_2^{P(2)} \ldots a_n^{P(n)} \operatorname{sg}(P) \Delta(e_1, e_2, \ldots, e_n) = \det(a) \Delta(e_1, e_2, \ldots, e_n).
\end{aligned}$$

La última fórmula del teorema está probada usando nada más que las propiedades **1** y **2.** Si se dispone de **3**, obtenemos $\Delta(a_1,\ldots,a_n)=\det(a)$. ♣

Si se llama determinante a una función Δ como en el teorema, se tiene que el determinante es único. Muchos autores desarrollan el concepto de determinante del siguiente modo: *postulan* que existe una función $\Delta:\mathbb{k}^n\times\ldots\times\mathbb{k}^n\to\mathbb{k}$ cumpliendo **1,2,3** en el teorema 85 y obtienen los teoremas de esta sección, *sin que sea necesaria la fórmula del sumatorio* (4.3). Esta carencia no es grave porque esa fórmula es poco útil. Más tarde, si son exigentes, prueban que Δ existe y tiene que venir dada por (4.3). Como se ve, si se quiere contar toda la historia, se trabaja lo mismo, estando la diferencia en el momento en que toca ver los teoremas más pesados o difíciles. Por ejemplo, si se quiere volver a probar que $\det(ab)=\det(a)\det(b)$ con este segundo enfoque, para a fija se define

$$\Delta:\mathbb{k}^n\times\ldots\times\mathbb{k}^n=\mathbb{k}^{n\times n}\longrightarrow\mathbb{k},\qquad \Delta(b_1,\ldots,b_n)=\det(ab)=D(ab_1,\ldots,ab_n).$$

y se comprueba con poco trabajo que Δ cumple **1** y **2** en el teorema 85. Entonces, con este teorema,

$$\det(ab)=\Delta(b_1,\ldots,b_n)=\Delta(e_1,\ldots,e_m)\,D(b_1,\ldots,b_n)=\det(aI_n)\det(b)=\det(a)\det(b).$$

En todo caso el lector que quiera probar $\det(ab)=\det(a)\det(b)$ tendrá que hacer cálculos con sumatorios muy parecidos, bien sean los del teorema 83 o los del 85.

4.4. Sobre cálculo de determinantes y sus aplicaciones

4.4.1. Cálculo de determinantes por cajas

Hay otra aplicación ingeniosa del teorema 85 que permite nuevos modos de calcular determinantes: es el **cálculo por cajas**. Primero una definiciones. Tenemos una matriz $a\in\mathbb{k}^{n\times n}$ con $n=p+q$. Dividimos a en cuatro **bloques** o **cajas,** que es el dividir la tabla de la matriz en cuatro partes

$$a=\begin{pmatrix} u & v\\ z & w\end{pmatrix}=\begin{pmatrix} u_1^1 & \cdots & u_p^1 & v_1^1 & \cdots & v_q^1\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ u_1^p & \cdots & u_p^p & v_1^p & \cdots & v_q^p\\ z_1^1 & \cdots & z_p^1 & w_1^1 & \cdots & w_q^1\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ z_1^q & \cdots & z_p^q & w_1^q & \cdots & w_q^q\end{pmatrix}$$

siendo u,v,w,z, en vez de números, matrices con $u\in\mathbb{k}^{p\times p}$ y $w\in\mathbb{k}^{q\times q}$; o sea, *cuadradas.*

Teorema 86 *Si la matriz z tiene todos sus coeficientes 0, se tiene que* $\det(a)=\det(u)\det(w)$.

Como se ve, esto supone una simplificación importante pues, valgan lo que valgan α y β,

$$\begin{vmatrix} 1 & 2 & 3 & \alpha & \beta\\ 1 & 0 & 1 & 0 & 1\\ 0 & 1 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 & -1\\ 0 & 0 & 0 & 6 & 7\end{vmatrix}=\begin{vmatrix} 1 & 2 & 3\\ 1 & 0 & 1\\ 0 & 1 & 0\end{vmatrix}\begin{vmatrix} 1 & -1\\ 6 & 7\end{vmatrix}=2\cdot 13=26.$$

Demostración. Fijemos $v\in\mathbb{k}^{p\times q}$ y sea

$$\Delta_v:\mathbb{k}^{p\times p}=\mathbb{k}^p\times\times\mathbb{k}^p\to\mathbb{k},\qquad \Delta_v(u_1,\ldots,u_p)=\det\begin{pmatrix} u & v\\ 0 & I_q\end{pmatrix}=\begin{vmatrix} u & v\\ 0 & I_q\end{vmatrix}.$$

Es muy fácil mostrar con las propiedades del determinante que tenemos una función cumpliendo **1** y **2** en el teorema 85, y entonces la fórmula general $\Delta(a_1,\ldots,a_n)=\Delta(e_1,\ldots,e_n)\det(a)$ da aquí

$$\det\begin{pmatrix} u & v\\ 0 & I_q\end{pmatrix}=\Delta_v(u_1,\ldots,u_p)=\Delta_v(e_1,\ldots,e_p)\,D(u_1,\ldots,u_p)=\det\begin{pmatrix} I_p & v\\ 0 & I_q\end{pmatrix}\det(u)=\det(u).$$

Pronto usaremos esta fórmula. Ahora, fijamos u y v y definimos

$$\Delta_{u,v} : \mathbb{k}^{q\times q} = \mathbb{k}^{1\times q} \times \ldots \times \mathbb{k}^{1\times q} \longrightarrow \mathbb{k}, \qquad \Delta_{u,v}\left(w^1, \ldots, w^q\right) = \det \begin{pmatrix} u & v \\ 0 & w \end{pmatrix},$$

que es función de las *filas* de w. También $\Delta_{u,v}$ cumple **1** y **2** en el teorema 85, lo que implica que

$$\begin{vmatrix} u & v \\ 0 & w \end{vmatrix} = \Delta_{u,v}\left(w^1, \ldots, w^q\right) = \Delta_{u,v}\left(e_1, \ldots, e_q\right) D\left(w^1, \ldots, w^q\right) = \begin{vmatrix} u & v \\ 0 & I_q \end{vmatrix} \det(w) = \det(u)\det(w)$$

y se ha usado en el último paso el cálculo del párrafo anterior. ♣

Sea $n = p_1 + \cdots + p_r$ y la matriz $a \in \mathbb{k}^{n\times n}$ con cajas $b(j) \in \mathbb{k}^{p_j \times p_j}$ a lo largo de la diagonal y cajas cero bajo ellas; o sea,

$$a = \begin{pmatrix} b(1) & \cdots & * & * & * \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & b(j) & \cdots & * \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & b(r) \end{pmatrix}$$

Se puede probar por inducción sobre r que $\det(a) = \det(b(1)) \cdot \cdots \cdot \det(b(r))$. Por ejemplo,

$$\begin{vmatrix} 1 & 2 & 6 & 6 & 7 & 7 \\ 3 & 4 & 4 & 5 & 9 & 9 \\ 0 & 0 & 5 & 6 & 2 & 2 \\ 0 & 0 & 7 & 8 & 1 & 1 \\ 0 & 0 & 0 & 0 & 9 & 10 \\ 0 & 0 & 0 & 0 & 11 & 12 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} \cdot \begin{vmatrix} 5 & 6 \\ 7 & 8 \end{vmatrix} \cdot \begin{vmatrix} 9 & 10 \\ 11 & 12 \end{vmatrix} = (-2)^3 = -8.$$

4.4.2. Cálculo de la inversa y desarrollo por filas o columnas

Dada la matriz $a \in \mathbb{k}^{n\times n}$, el **menor** M_j^i es el determinante de la matiz $(n-1)\times(n-1)$ que se obtiene al tachar en a la fila i y y la columna j. Por tanto a y M_j^i son respectivamente

$$\begin{pmatrix} a_1^1 & \cdots & a_{j-1}^1 & \mathbf{a}_j^1 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^{i-1} & \cdots & a_{j-1}^{i-1} & \mathbf{a}_j^{i-1} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ \mathbf{a}_1^i & \cdots & \mathbf{a}_{j-1}^i & \mathbf{a}_j^i & \mathbf{a}_{j+1}^i & \cdots & \mathbf{a}_n^i \\ a_1^{i+1} & \cdots & a_{j-1}^{i+1} & \mathbf{a}_j^{j+1} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^n & \cdots & a_{j-1}^n & \mathbf{a}_j^n & a_{j+1}^n & \cdots & a_n^n \end{pmatrix}, \qquad \begin{vmatrix} a_1^1 & \cdots & a_{j-1}^1 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^{i-1} & \cdots & a_{j-1}^{i-1} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ a_1^{i+1} & \cdots & a_{j-1}^{i+1} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_1^n & \cdots & a_{j-1}^n & a_{j+1}^n & \cdots & a_n^n \end{vmatrix}.$$

Es esencial observar que, *salvo un signo*, el menor, que se define como *determinante* de una matriz $(n-1)\times(n-1)$, es también el determinante de una matriz $n\times n$. Esta matriz, puesta como sucesión de columnas o con detalle es

$$\left(a_1, \ldots, a_{j-1}, \overset{(j)}{e_i}, a_{j+1}, \ldots, a_n\right) = \begin{pmatrix} a_1^1 & \cdots & a_{j-1}^1 & \mathbf{0} & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^{i-1} & \cdots & a_{j-1}^{i-1} & \mathbf{0} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ \mathbf{a}_1^i & \cdots & \mathbf{a}_{j-1}^i & \mathbf{1} & \mathbf{a}_{j+1}^i & \cdots & \mathbf{a}_n^i \\ a_1^{i+1} & \cdots & a_{j-1}^{i+1} & \mathbf{0} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^n & \cdots & a_{j-1}^n & \mathbf{0} & a_{j+1}^n & \cdots & a_n^n \end{pmatrix}$$

Es como si en a, la cruz que forman la fila i y la columna j se rellenara con ceros, excepto el punto de corte de los brazos, donde va un uno. La fórmula a probar es[9]

$$(-1)^{i+j} M_j^i = \det\left(a_1, \ldots, a_{j-1}, \overset{(j)}{e_i}, a_{j+1}, \ldots, a_n\right). \tag{4.4}$$

[9] Obsérvese que en la ecuación (4.4) la posición de los índices i, j se ha invertido, pasando el superior a posición inferior.

Esto es fácil si se ponen las matrices con detalle porque $\det\left(a_1,\ldots,a_{j-1},\overset{(j)}{e_i},a_{j+1},\ldots,a_n\right)$ es

$$\begin{vmatrix} a_1^1 & \cdots & a_{j-1}^1 & 0 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^{i-1} & \cdots & a_{j-1}^{i-1} & 0 & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ a_1^i & \cdots & a_{j-1}^i & 1 & a_{j+1}^i & \cdots & a_n^i \\ a_1^{i+1} & \cdots & a_{j-1}^{i+1} & 0 & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_1^n & \cdots & a_{j-1}^n & 0 & a_{j+1}^n & \cdots & a_n^n \end{vmatrix} = (-1)^{j-1} \begin{vmatrix} 0 & a_1^1 & \cdots & a_{j-1}^1 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{i-1} & \cdots & a_{j-1}^{i-1} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ 1 & a_1^i & \cdots & a_{j-1}^i & a_{j+1}^i & \cdots & a_n^i \\ 0 & a_1^{i+1} & \cdots & a_{j-1}^{i+1} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \cdots & a_{j-1}^n & a_{j+1}^n & \cdots & a_n^n \end{vmatrix}$$

tras pasar la columna j al primer lugar. A continuación, pasando la fila i al primer lugar con otra permutación cíclica,

$$\begin{vmatrix} 0 & a_1^1 & \cdots & a_{j-1}^1 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{i-1} & \cdots & a_{j-1}^{i-1} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ 1 & a_1^i & \cdots & a_{j-1}^i & a_{j+1}^i & \cdots & a_n^i \\ 0 & a_1^{i+1} & \cdots & a_{j-1}^{i+1} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \cdots & a_{j-1}^n & a_{j+1}^n & \cdots & a_n^n \end{vmatrix} = (-1)^{i-1} \begin{vmatrix} 1 & a_1^i & \cdots & a_{j-1}^i & a_{j+1}^i & \cdots & a_n^i \\ 0 & a_1^1 & \cdots & a_{j-1}^1 & a_{j+1}^1 & \cdots & a_n^1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^{i-1} & \cdots & a_{j-1}^{i-1} & a_{j+1}^{i-1} & \cdots & a_n^{i-1} \\ 0 & a_1^{i+1} & \cdots & a_{j-1}^{i+1} & a_{j+1}^{i+1} & \cdots & a_n^{i+1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_1^n & \cdots & a_{j-1}^n & a_{j+1}^n & \cdots & a_n^n \end{vmatrix}.$$

Esto último es $(-1)^{i-1} M_j^i$ (calcular por cajas) y (4.4) es inmediato.

Se llama el **cofactor** de a_j^i al elemento

$$\widetilde{a}_j^i = (-1)^{i+j} M_j^i = \det\left(a_1,\ldots,a_{j-1},\overset{(j)}{e_i},a_{j+1},\ldots,a_n\right).$$

(Insistimos, ¡ojo a la posición de los índices!). Los cofactores son menores con signo.

Teorema 87 *Se tiene que* $\left(\widetilde{a}\right)^\top a = \det(a)\, I$. *Con palabras: la* traspuesta *de la matriz de los cofactores multiplicada por* a *da* $\det(a)\, I$, *múltiplo escalar de la matriz unidad.*

Demostración. La demostración es muy sencilla con (4.4). En efecto,

$$\left(\left(\widetilde{a}\right)^\top a\right)_k^i = \sum_{j=1}^n \left(\left(\widetilde{a}\right)^\top\right)_j^i a_k^j = \sum_{j=1}^n \widetilde{a}_i^j a_k^j = \sum_{j=1}^n \det\left(a_1,\ldots,a_{i-1},\overset{(i)}{e_j},a_{i+1},\ldots,a_n\right) a_k^j$$

$$= \det\left(a_1,\ldots,a_{i-1},\sum_{j=1}^n \overset{(i)}{a_k^j e_j},a_{i+1},\ldots,a_n\right) = \det\left(a_1,\ldots,a_{i-1},a_k,a_{i+1}\ldots,a_n\right).$$

Si $k = i$ resulta $\left(\left(\widetilde{a}\right)^\top a\right)_i^i = \det\left(a_1,\ldots,a_{i-1},a_i,a_{i+1}\ldots,a_n\right) = \det(a)$. Si $k \neq i$, la columna a_k está repetida dos veces en $\det\left(a_1,\ldots,a_{i-1},a_k,a_{i+1}\ldots,a_n\right)$ y ese determinante es nulo. Resumiendo, $\left(\left(\widetilde{a}\right)^\top a\right)_k^i = \det(a)\,\delta_k^i$, que es el teorema.[10] ♣

Este teorema tiene como corolario un procedimiento para calcular el determinante desarrollando por una columna.

Corolario 3 *Sumando los productos de cada elemento de la columna* i *por su correspondiente cofactor se obtiene el determinante. Con símbolos*

$$\det(a) = \sum_{j=1}^n \widetilde{a}_i^j a_i^j = \sum_{j=1}^n (-1)^{i+j} M_i^j a_i^j.$$

[10] El lector habrá observado que, de las dos definiciones de cofactor, se usa la que tiene forma de determinante $n \times n$ para este teorema teórico, pero en los cálculos prácticos es mejor que sea el determinante obtenido al tachar una fila y columna con un signo.

Demostración. Inmediata porque $\det(a) = \left(\left(\widetilde{a}\right)^\top a\right)_i^i = \sum_{j=1}^n \widetilde{a}_i^j a_i^j$. ♣

Corolario 4 *Si* $\det(a) \neq 0$ *la inversa de* a *es*

$$\frac{1}{\det(a)}\left(\widetilde{a}\right)^\top = a^{-1}. \tag{4.5}$$

Demostración. El teorema 84 nos dice que a^{-1} existe y el teorema 87 da

$$\left[\frac{1}{\det(a)}\left(\widetilde{a}\right)^\top\right] a = I.$$

Multiplicando a la derecha por a^{-1} sale la fórmula. ♣

Problema 169 *Probar para* $n = 2$ *la fórmula*

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix}^{-1} = \frac{1}{ps - qr}\begin{pmatrix} s & -q \\ -r & p \end{pmatrix}$$

supuesto $ps - qr \neq 0$.

Problema 170 *Probar que si*

$$a = \begin{pmatrix} 1 & p & q \\ 0 & 1 & r \\ 0 & 0 & 1 \end{pmatrix} \text{ entonces } a^{-1} = \begin{pmatrix} 1 & p & q \\ 0 & 1 & r \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -p & pr - q \\ 0 & 1 & -r \\ 0 & 0 & 1 \end{pmatrix}$$

con (4.5) y con las técnicas del primer capítulo. Nota: el objetivo es convencer al lector que las técnicas del primer capítulo son más cómodas que (4.5) aun con una matriz sencilla.

Problema 171 *Probar que la suma de los elementos de una fila por sus cofactores es el determinante.*

4.4.3. Otras cuestiones sobre determinantes

Es bien conocida, quizás con otra presentación, la **regla de Cramer** para resolver un sistema lineal.

Teorema 88 *Sea* $ax = y$ *un sistema lineal con* $a \in \Bbbk^{n\times n}$ *y* $\det(a) \neq 0$. *La solución es*

$$x^k = \frac{\det\left(a_1, \ldots, a_{k-1}, \overset{(k)}{y}, a_{k+1}, \ldots, a_n\right)}{\det(a_1, \ldots, a_k, \ldots, a_n)}, \qquad 1 \leq k \leq n.$$

Demostración. Como a es invertible, hay solución única $x = a^{-1}y$. Escribimos $x = \sum_{j=1}^n x^j a_j$ y

$$\begin{aligned}\det\left(a_1, \ldots, a_{k-1}, \overset{(k)}{y}, a_{k+1}, \ldots, a_n\right) &= \det\left(a_1, \ldots, a_{k-1}, \sum_{j=1}^n x^j a_j, a_{k+1}, \ldots, a_n\right) \\ &= \sum_{j=1}^n x^j \det(a_1, \ldots, a_{k-1}, a_j, a_{k+1}, \ldots, a_n) \\ &= x^k \det(a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_n)\end{aligned}$$

porque si $j \neq k$ está repetida a_j y el sumando vale cero. Se acaba dividiendo por $\det(a_1, \ldots, a_n)$. ♣

El problema que sigue es una fórmula de Cramer "con disfraz".

Problema 172 *Sea* $\mathcal{U} = (u_1, \ldots, u_n)$ *una base de* $\Bbbk^n$. *Las coordenadas* $\left(\xi^1, \ldots, \xi^n\right)$ *de* x *en* $\mathcal{U}$ *(¡no en la base estándar!), vienen dadas por*

$$\xi^k = \frac{\det\left(u_1, \ldots, \overset{(k)}{x}, \ldots, u_n\right)}{\det(u_1, \ldots, u_k, \ldots, u_n)}, \qquad 1 \leq k \leq n.$$

Elijamos para $1 \leq r \leq m$, $1 \leq s \leq n$ dos sucesiones $1 \leq i_1 < i_2 < \ldots < i_r \leq m$, $1 \leq j_1 < j_2 < \ldots < j_s \leq n$ que llamaremos para abreviar I y J. Dada una matriz $a \in \Bbbk^{m\times n}$ (puede no ser cuadrada) llamamos **submatriz (correspondiente a I, J)** a la matriz b de dimensiones $r \times s$ que se obtiene quedándonos con los coeficientes donde se cortan las filas $i_1, i_2, \ldots, i_r$ y las columnas $j_1, j_2, \ldots j_s$. Antes de dar una definición con símbolos ponemos un ejemplo con $m = 6$, $n = 5$, $r = s = 3$, $I = (1, 2, 5)$, $J = (2, 3, 5)$, donde $c \in \Bbbk^{5\times 3}$ es el paso intermedio tachando las columnas $1, 4$,

$$a = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{5} & 2 & \mathbf{2} \\ 0 & \mathbf{0} & \mathbf{1} & 2 & \mathbf{8} \\ 0 & 3 & 2 & 1 & 0 \\ 1 & 1 & 2 & 2 & 2 \\ 3 & \mathbf{4} & \mathbf{0} & 6 & \mathbf{1} \\ 2 & 3 & 2 & 9 & 0 \end{pmatrix}, \qquad c = \begin{pmatrix} 0 & 5 & 2 \\ 0 & 1 & 8 \\ 3 & 2 & 0 \\ 1 & 2 & 2 \\ 4 & 0 & 1 \\ 3 & 2 & 0 \end{pmatrix}, \qquad b = \begin{pmatrix} 0 & 5 & 2 \\ 0 & 1 & 8 \\ 4 & 0 & 1 \end{pmatrix},$$

y donde los coeficientes que pasan a b aparecen en negrita en a. La definición formal de b es $b_q^p = a_{j_q}^{i_p}$. El siguiente problema nos permite calcular el rango de una matriz $a \in \Bbbk^{m\times n}$ (puede no ser cuadrada) buscando el máximo r tal que haya submatrices $b \in \Bbbk^{r\times r}$ invertibles. Por supuesto, se puede ver si b es invertible con $\det(b) \neq 0$.

Problema 173 *Probar que el rango de a es el máximo s tal que hay una submatriz $b \in \Bbbk^{s\times s}$ invertible (equivalente a $\det(b) \neq 0$). En ese caso, todas las submatrices $c \in \Bbbk^{t\times t}$ con $t > s$ tienen $\det(c) = 0$.*

Se toma una matriz a cuyos coeficientes son 1 o 0 puestos como en un tablero de ajedrez. Por ejemplo,

$$a = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Problema 174 *Probar que hay un número $\ell \in \mathbb{N}$ tal que* cualquier *submatriz $h \times h$ de a con $h \geq \ell$ tiene determinante nulo. ¿Cómo se generaliza el problema a matrices $m \times m$ (aquí $m = 5$)?*

Tomamos $\Bbbk = \mathbb{C}$, los números complejos. Para i, la unidad imaginaria, observamos que $\left(i, i^2, i^3, i^4\right) = (i, -1, -i, 1)$ y

$$\left(i, i^2, i^3, i^4, \ldots, i^{4k+1}, i^{4k+2}, i^{4k+3}, i^{4k4}, \ldots\right) = (i, -1, -i, 1, \ldots, i, -1, -i, 1, \ldots).$$

Problema 175 *Plantear y resolver un problema análogo al anterior para*

$$a = \begin{pmatrix} i & i^2 & i^3 & i^4 \\ i^2 & i^3 & i^4 & i \\ i^3 & i^4 & i & i^2 \\ i^4 & i & i^2 & i^3 \end{pmatrix}.$$

Un comentario general sobre las fórmulas y procedimientos que requieren calcular determinantes. Tienen el atractivo de dar una fórmula completa y concreta y no un procedimiento de cálculo. Buen ejemplo son las fórmulas de la inversa de una matriz 2×2 del problema 169. Pero en general, estos cálculos requieren un número de operaciones mucho mayor con determinantes que con la aplicación de operaciones elementales. En los ejemplos de los libros, las matrices son pequeñas y se nota poco la diferencia, pero si $n = 10$ debemos saber que $10! = 3{,}628.800$, que son los sumandos que tendrá la expresión de una matriz 10×10 si no conocemos un atajo o hay alguna estrategia simplificadora. Contra lo que pueda indicar una primera intuición, los determinantes aparecen mucho más en las partes teóricas. Los cálculos puramente numéricos se efectúan combinando procesos como "aproximar" la matriz a una forma triangular, desarrollar por cajas, filas o columnas, y hacer aparecer muchos ceros.

Dicho esto, tiene cierto atractivo el cálculo de determinantes en matrices con los coeficientes repartidos de una forma "regular", aunque es imposible definir con precisión lo que esto significa. Los problemas 178 y 172 dan en abstracto ideas que funcionan en problemas concretos. Es más fácil verificar lo abstracto que descubrir cómo encaja en ello lo concreto. La idea simplificadora está en pensar en las n columnas o filas de la matriz y no en sus n^2 coeficientes; *en ver el determinante como función de filas o columnas.*

Problema 176 *Elegimos w en $\Bbbk^n$ y λ en $\Bbbk$. Sea s la suma de las coordenadas de w. Probar que*

$$\Delta = D\left(w+\lambda e_1, w+\lambda e_2, \ldots, w+\lambda e_n\right) = \lambda^n + \lambda^{n-1}s.$$

El interés del siguiente problema no es el de *comprobar* lo que se dice tras cierto trabajo, sino el hacerlo con arte y gracia. La fuerza bruta luce poco aquí. El lector imaginará que usa el problema 176.

Problema 177 *Probar que para $a, b, c \in \Bbbk$ se tiene*

$$\begin{vmatrix} a-b-c & 2a & 2a \\ 2b & -a+b-c & 2b \\ 2c & 2c & -a-b+c \end{vmatrix} = (a+b+c)^3.$$

El problema que sigue es también muy útil para simplificar cálculos.

Problema 178 *Sustituimos la matriz $a = (a_1, \ldots, a_n)$ por $b = (b_1, \ldots, b_n)$, siendo $b_j = a_j + c_j$ y c_j una combinación lineal de $a_1, a_2, \ldots, a_{j-1}$. Probar que* $\det(a) = \det(b)$.

Damos como ejemplo el famoso **determinante de Vandermonde**.

Problema 179 *Sean $x_1, x_2, \ldots, x_n \in \Bbbk$. Probar que, para $n \geq 2$,*

$$\Delta_n = \begin{vmatrix} 1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-2} & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-2} & (x_2)^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & x_{n-1} & (x_{n-1})^2 & \cdots & (x_{n-1})^{n-2} & (x_{n-1})^{n-1} \\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-2} & (x_n)^{n-1} \end{vmatrix} = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Nota: En $(x_i)^p$, p es un exponente. Como notación general, si H es un conjunto de índices, $\prod_{h\in H}\lambda_h$ es el producto de todos los λ_h cuando h recorre H. ♦

Solución. Ilustramos primero el caso $n = 3$ con a, b, c en vez de x_1, x_2, x_3. Se empieza restando a la columna 2 la primera multiplicada por a y a la tercera la segunda multiplicada por a,

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} \to \begin{vmatrix} 1 & 0 & 0 \\ 1 & b-a & b^2-ba \\ 1 & c-a & c^2-ca \end{vmatrix} = \begin{vmatrix} b-a & b^2-ba \\ c-a & c^2-ca \end{vmatrix} = \begin{vmatrix} b-a & b(b-a) \\ c-a & c(c-a) \end{vmatrix}$$

$$= (b-a)(c-a)\begin{vmatrix} 1 & b \\ 1 & c \end{vmatrix} = (b-a)(c-a)(b-c).$$

En general, se prueba por inducción. Es cierta la fórmula para $n = 2$. Supongámosla cierta para $n-1$. Queremos decir con esto que para *toda sucesión* $(x_1, \ldots, x_n)$ de n elementos de $\Bbbk$ se tiene la fórmula. Entonces,

$$\Delta_n = \begin{vmatrix} 1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-2} & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-2} & (x_2)^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & x_{n-1} & (x_{n-1})^2 & \cdots & (x_{n-1})^{n-2} & (x_{n-1})^{n-1} \\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-2} & (x_n)^{n-1} \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots \\ 1 & x_2-x_1 & (x_2)^2 - x_2x_1 & \cdots & (x_2)^k - (x_2)^{k-1}x_1 & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots \\ 1 & x_{n-1}-x_1 & (x_{n-1})^2 - x_{n-1}x_1 & \cdots & (x_{n-1})^k - (x_{n-1})^{k-1}x_1 & \cdots \\ 1 & x_n-x_1 & (x_n)^2 - x_nx_1 & \cdots & (x_n)^k - (x_n)^{k-1}x_1 & \cdots \end{vmatrix}$$

Lo que hemos hecho es restar a la columna k la columna $k-1$ multiplicada por x_1 para $k=2,3,\ldots,n$. Desarrollando por la primera fila vemos que Δ_n es expresable con determinantes $(n-1)\times(n-1)$ (matrices más pequeñas)

$$\Delta_n = \begin{vmatrix} x_2-x_1 & (x_2)^2-x_2x_1 & \cdots & (x_2)^k-(x_2)^{k-1}x_1 & \cdots & (x_2)^{n-1}-(x_2)^{n-2}x_1 \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{n-1}-x_1 & (x_{n-1})^2-x_{n-1}x_1 & \cdots & (x_{n-1})^k-(x_{n-1})^{k-1}x_1 & \cdots & (x_{n-1})^{n-1}-(x_{n-1})^{n-2}x_1 \\ x_n-x_1 & (x_n)^2-x_nx_1 & \cdots & (x_n)^k-(x_n)^{k-1}x_1 & \cdots & (x_n)^{n-1}-(x_n)^{n-2}x_1 \end{vmatrix}$$

$$= \begin{vmatrix} x_2-x_1 & x_2(x_2-x_1) & \cdots & (x_2)^{k-1}(x_2-x_1) & \cdots & (x_2)^{n-2}(x_2-x_1) \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{n-1}-x_1 & x_{n-1}(x_{n-1}-x_1) & \cdots & (x_{n-1})^{k-1}(x_{n-1}-x_1) & \cdots & (x_{n-1})^{n-2}(x_{n-1}-x_1) \\ x_n-x_1 & x_n(x_n-x_1) & \cdots & (x_n)^{k-1}(x_n-x_1) & \cdots & (x_n)^{n-2}(x_n-x_1) \end{vmatrix}.$$

En la primera fila se saca factor común x_2-x_1, en la segunda x_3-x_1,... y en la $n-1$ se saca factor común x_n-x_1. Queda entonces

$$\Delta_n = (x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)\begin{vmatrix} 1 & x_2 & \cdots & (x_2)^{k-1} & \cdots & (x_2)^{n-2} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{n-1} & \cdots & (x_{n-1})^{k-1} & \cdots & (x_{n-1})^{n-2} \\ 1 & x_n & \cdots & (x_n)^{k-1} & \cdots & (x_n)^{n-2} \end{vmatrix}$$

$$= (x_2-x_1)(x_3-x_1)\cdots(x_n-x_1)\,\Delta'_{n-1},$$

siendo Δ'_{n-1} el determinante de Vandermonde para la sucesión $(x_2,\ldots,x_n)$. Por la hipótesis inductiva,

$$\Delta_m = \left[\prod_{j=2}^{n}(x_j-x_1)\right]\Delta'_{n-1} = \left[\prod_{j=2}^{n}(x_j-x_1)\right]\left[\prod_{2\le i<j\le n}(x_j-x_i)\right] = \prod_{1\le i<j\le n}(x_j-x_i).\ \blacklozenge$$

Problema 180 *Calcular*

$$\Delta_1 = \begin{vmatrix} bcd & 1 & a & a^2 \\ acd & 1 & b & b^2 \\ abd & 1 & c & c^2 \\ abc & 1 & d & d^2 \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} h+k & k & \cdots & k & k \\ k & h+k & \cdots & k & k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k & k & \cdots & h+k & k \\ k & k & \cdots & k & h+k \end{vmatrix}, \quad \Delta_3 = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{vmatrix}$$

Nota: Hay que relacionar Δ_1 con el determinante de Vandermonde.

Problema 181 *Sea $a\in\mathbb{k}^{n\times n}$. Dar la relación entre $D(a_1,a_2,\ldots,a_{n-1},a_n)$ y $D(a_n,a_{n-1},\ldots,a_2,a_1)$? (Se ha invertido el orden de las columnas.) Ahora más complicado: en una matriz 3×3 hacemos las operaciones de "invertir el orden de las columnas", a continuación, "invertir el orden de las filas" y, finalmente "trasponer la matriz". En concreto,*

$$\begin{pmatrix} h & p & x \\ i & q & y \\ j & r & z \end{pmatrix} \to \begin{pmatrix} x & p & h \\ y & q & i \\ z & r & j \end{pmatrix} \to \begin{pmatrix} z & r & j \\ y & q & i \\ x & p & h \end{pmatrix} \to \begin{pmatrix} z & y & x \\ r & q & p \\ j & i & h \end{pmatrix}.$$

¿Cuál es la relación entre los determinantes de la matriz inicial y la última?

Problema 182 *Calcular el siguiente determinante de una matriz $a(n)\in\mathbb{k}^{n\times n}$ definida por*

$$a(1)=(1), \quad a(2)=\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad a(n)=\begin{vmatrix} 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \end{vmatrix}$$

con 1 en la diagonal principal y las dos adyacentes y 0 en los otros lugares. $\blacklozenge$

Solución. Vamos a obtener una fórmula de $\Delta_n = \det a(n)$ en función de otros Δ_k con $k < n$. Operamos y

$$\begin{vmatrix} 1 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} \overset{1}{=} \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} \overset{2}{=} \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}.$$

En $\overset{1}{=}$ se resta a la segunda fila la primera y en $\overset{2}{=}$ se resta a la cuarta fila la segunda. Queda una matriz (ver más abajo) con una caja 3×3 en la esquina superior izquierda, $a(n-3)$ en la inferior derecha, una matriz b cuya forma precisa no importa y la matriz $0_{(n-3)\times 3}$ de $n-3$ filas y tres columnas con todo ceros. Por tanto, multiplicando por cajas,

$$\det a(n) = \det \begin{pmatrix} 1 & 1 & 0 & \\ 0 & 0 & 1 & b \\ 0 & 1 & 1 & \\ 0_{(n-3)\times 3} & & & a(n-3) \end{pmatrix} = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} \det a(n-3) = -\det a(n-3).$$

Calculamos directamente $\Delta_1 = 1$, $\Delta_2 = 0$ y $\Delta_3 = -1$ y, como $\Delta_n = -\Delta_{n-3}$,

1. Si $n = 3q$ es divisible por 3 nos queda
$$\Delta_n = (-1)^1 \Delta_{n-3} = (-1)^2 \Delta_{n-6} = \ldots = (-1)^r \Delta_{n-3r} = \ldots = (-1)^{q-1} \Delta_3 = (-1)^q.$$
2. Si $n = 3q+1$ nos queda
$$\Delta_n = (-1)^1 \Delta_{n-3} = (-1)^2 \Delta_{n-6} = \ldots = (-1)^r \Delta_{n-3r} = \ldots = (-1)^q \Delta_1 = (-1)^q.$$
3. Si $m = 3q+2$ nos queda
$$\Delta_n = (-1)^1 \Delta_{n-3} = (-1)^2 \Delta_{n-6} = \ldots = (-1)^r \Delta_{m-3r} = \ldots = (-1)^q \Delta_2 = 0.$$

Por ejemplo, $\Delta_{12} = (-1)^4 = 1$, $\Delta_{10} = (-1)^3 = -1$ y $\Delta_5 = 0$. Los valores $0, \pm 1$ dependen del resto que queda al dividir n por 3. ♦

Damos ahora unos determinantes, primero de matrices 5×5 para ayudar a la comprensión, pero la idea es calcular la matriz generalizada a $n \times n$. Se trata de ver que si Δ_n es su determinante, hay una fórmula recursiva que permite expresar Δ_n en función de los anteriores Δ_j. Como Δ_1 y Δ_2 son inmediatos, se calcula Δ_n.

Problema 183 *Calcular los determinantes de las matrices*

$$a(5) = \begin{vmatrix} 1 & h & 0 & 0 & 0 \\ 1 & 1 & h & 0 & 0 \\ 1 & 1 & 1 & h & 0 \\ 1 & 1 & 1 & 1 & h \\ 1 & 1 & 1 & 1 & 1 \end{vmatrix}, \quad b(5) = \begin{vmatrix} 1 & q & 0 & 0 & 0 \\ 0 & 1 & q & 0 & 0 \\ 0 & 0 & 1 & q & 0 \\ 0 & 0 & 0 & 1 & q \\ q & q & q & q & 1 \end{vmatrix}, \quad c(5) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}.$$

Generalizar a matrices $n \times n$ calculando los determinantes por una fórmula recursiva.

El problema que sigue tiene dificultad algo superior a la media. El lector decidirá si quiere o no aceptar el desafío.

Problema 184 *Consideremos los números* $17 \cdot 41 = 697$, $17 \cdot 94 = 1598$, $17 \cdot 139 = 2363$ *y* $17 \cdot 365 = 6205$. *Con ellos se forma el determinante* 4×4

$$\Delta = \begin{vmatrix} 0 & 6 & 9 & 7 \\ 1 & 5 & 9 & 8 \\ 2 & 3 & 6 & 3 \\ 6 & 2 & 0 & 5 \end{vmatrix}.$$

Probar que Δ es divisible por 17. Ya imaginará el lector que no se trata de calcular Δ directamente (sale $255 = 17 \cdot 15$) sino de justificar que hay una razón profunda por la que se puede dar un teorema general.

Si se prefiere al teorema general y abstracto, es este. Se consideran números

$$\left\{\begin{array}{l} A^0 = a_0^0 \cdot 10^0 + a_1^0 \cdot 10^1 + a_2^0 10^2 + \ldots + a_n^0 10^n \\ A^1 = a_0^1 \cdot 10^0 + a_1^1 \cdot 10^1 + a_2^1 10^2 + \ldots + a_n^1 10^n \\ \qquad\qquad \vdots \\ A^n = a_0^n \cdot 10^0 + a_1^n \cdot 10^1 + a_2^n 10^2 + \ldots + a_n^n 10^n \end{array}\right. .$$

El número A^j tiene pues expresión decimal $a_n^j a_{n-1}^j \cdots a_1^j a_0^j$. Supongamos que todos los números A^j son divisibles por otro entero B. Probar que la matriz $(n+1) \times (n+1)$ construida con sus cifras

$$a = \begin{pmatrix} a_n^0 & a_{n-1}^0 & \cdots & a_1^0 & a_0^0 \\ a_n^1 & a_{n-1}^1 & \cdots & a_1^1 & a_0^1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_n^{n-1} & a_{n-1}^{n-1} & \cdots & a_1^{n-1} & a_0^{n-1} \\ a_n^n & a_{n-1}^n & \cdots & a_1^n & a_0^n \end{pmatrix}$$

tiene un determinante Δ que es también divisible por B.

Hacemos una observación que tendrá su importancia al hablar más adelante del polinomio característico. Hemos definido $\det(a)$ cuando los coeficientes de a están en un cuerpo $\Bbbk$. Sin embargo, si $\mathbb{A}$ es un anillo conmutativo y unitario (piénsese en $\mathbb{Z}$ o $\Bbbk[X]$) se pueden definir matrices con coeficientes en $\mathbb{A}$, sumarlas y multiplicarlas con las propiedades usuales, definir matriz invertible, y el determinante de a. Es notable lo poco que se usa la existencia de inversos multiplicativos en $\Bbbk$ al desarrollar los determinantes. En un ejercicio muy pesado el comprobar que todo funciona con tal de sustituir en algunos teoremas la condición $\det(a) \neq 0$ por "$\det(a)$ es un elemento invertible de $\mathbb{A}$". Por ejemplo, una matriz con coeficientes en $\mathbb{Z}$ es invertible (existe b con coeficientes en $\mathbb{Z}$, no vale que sean en $\mathbb{Q}$, tal que $ab = ba = I$) si y solo si $\det(a) = \pm 1$. En $\Bbbk[X]$ el teorema análogo es que $\det(a)$ sea un polinomio de grado 0. No vamos a profundizar en este asunto limitándonos a decir que cuando al tratar el polinomio característico aparezcan matrices con coeficientes en $\Bbbk[X]$ podemos manejarlas por lo que a determinantes se refiere de modo natural.

4.5. El determinante de un endomorfismo

Desde el comienzo del capítulo no se habla de espacios vectoriales abstractos. Podemos plantearnos si igual que existe la función $D : \Bbbk^n \times \ldots \times \Bbbk^n \longrightarrow \Bbbk$ (n factores $\Bbbk^n$) que cumple **1**,**2** y **3** en el teorema 80, no se podría construir otra función análoga $D_{\mathbb{E}} : \mathbb{E} \times \cdots \times \mathbb{E} \longrightarrow \Bbbk$ que cumpla al menos **1** y **2**, ya que **3** exige una base "distinguida" $\mathcal{U}$ y en $\mathbb{E}$ abstracto no la hay. La respuesta es que sí, pero que hay un obstáculo insalvable, y es que hay que elegir a la fuerza una base $\mathcal{U}$ y convertirla en "distinguida". En estas circunstancias sí que hay una función $D_{\mathcal{U}} : \mathbb{E} \times \cdots \times \mathbb{E} \longrightarrow \Bbbk$ que cumple **1**,**2** y **3,** convirtiéndose **3** en $D_{\mathcal{U}}(u_1, \ldots, u_n) = 1$. Esto generaliza el determinante que hasta ahora se ha tratado, porque si $\mathbb{E} = \Bbbk^n$ y $\mathcal{U} = \mathcal{E}$, la $D_{\mathcal{U}}$ construida con generalidad se convierte en D, el determinante de las columnas. Damos estos comentarios sin demostraciones simplemente para resaltar que si se quiere que el determinante actúe sobre sucesiones de vectores $(a_1, \ldots, a_n)$ de $\mathbb{E}$ necesitamos elegir una base $\mathcal{U}$ y que esto quita bastante interés al concepto porque para otra base $\mathcal{V}$ se puede tener (de hecho se tiene) $D_{\mathcal{V}} \neq D_{\mathcal{U}}$ y hay que relacionar las dos funciones.

Es sin embargo muy notable que, sin tenernos que ligar a una base concreta, sí que se puede definir el determinante para una función lineal $L : \mathbb{E} \to \mathbb{E}$. Obsérvese que no decimos que sea $L : \mathbb{E} \to \mathbb{F}$ admitiendo incluso $\dim(\mathbb{E}) = \dim(\mathbb{F})$, sino que L debe llevar $\mathbb{E}$ en sí mismo; o sea, ser un endomorfismo. La clave está en elegir primero una base $\mathcal{U}$, escribir $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$ su matriz en esta base, y definir

$$\det(L) = \det(a) = \det\left(\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)\right).$$

Naturalmente, hay que aclarar que la definición no depende de la base, pues si no debería ponerse $\det_{\mathcal{U}}(L)$ y estaríamos atados a una base.[11] Vamos a hacerlo. La clave es la relación

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) \cdot \mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})^{-1}$$

[11] El procedimiento es similar al del capítulo anterior para definir la traza de L.

del teorema 65 que relaciona las matrices de una función lineal referidas a bases $\mathcal{U}$ y $\mathcal{V}$ diferentes. Aligeramos esta notación tan pesada. Si $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$ y $b = \mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L)$ tenemos una relación $b = cac^{-1}$ donde c es una matriz de cambio de base (no importa en qué sentido). Con los teoremas 83 y 84 se obtiene

$$\det(b) = \det\left(cac^{-1}\right) = \det(c)\det(a)\frac{1}{\det(c)} = \det(a).$$

Verbalmente: si se toma como matriz de L una cualquiera, pero con la misma base inicial y final, aunque la matriz varía con la base, no lo hace su determinante, que se convierte por definición en el **determinante del endomorfismo** L. Esto implica en particular que se puede elegir la base como queramos para calcular $\det(L)$, aunque procuraremos que la matriz a sea lo más sencilla posible, para calcular $\det(a)$.

Problema 185 *Recordamos que si $\mathbb{E}$ es descomponible en suma directa $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ tenemos la simetría S respecto a $\mathbb{F}$ dada por $S(y+z) = y - z$ suponiendo que $x \in \mathbb{E}$ se escribe $x = y + z$ con $y \in \mathbb{F}$ y $z \in \mathbb{G}$. Calcular* $\det(S)$. *(Depende de las dimensiones.) Ídem para la proyección $P(x) = P(y+z) = z$.*

Teorema 89 *El determinante de un endomorfismo verifica*

1. *El endomorfismo $L : \mathbb{E} \to \mathbb{E}$ es invertible si y solo si* $\det(L) \neq 0$.
2. *Si $M : \mathbb{E} \to \mathbb{E}$ es otro endomorfismo,* $\det(L \circ M) = \det(L)\det(M)$.
3. $\det(\mathrm{id}_{\mathbb{E}}) \neq 0$.

Problema 186 *Probar el teorema precedente.*

Puede pensarse visto el inicio de la sección que al estar en espacios abstractos solo habrá que considerar determinantes de endomorfismos. En realidad no es así pues muchas veces lo que cuenta es el que el determinante sea o no sea cero, y no su valor concreto. Por ejemplo, si tenemos $L : \mathbb{E} \to \mathbb{F}$ entre espacios *distintos pero de la misma dimensión* y a es la matriz de L para bases cualesquiera, tenemos que L es un isomorfismo si y solo si a es invertible, y esto equivale a $\det(a) \neq 0$. Pueden cambiar, al cambiar las bases, matrices y determinantes, pero el que los determinantes sean (todos) no nulos, solo depende de L.

Problema 187 *Sea $\mathbb{E} = \mathbb{R}_n[X]$ el espacio de los polinomios de grado $\leq n$ y sea $\mathbb{F} = \mathbb{R}^{n+1}$. Denotaremos por $\mathcal{E}'$ y $\mathcal{E}''$ a sus respectivas bases estándar. Se eligen $t_1, t_2, \ldots t_{n+1}$ en $\mathbb{R}$ que pueden ser iguales o distintos entre sí. Se define*

$$L : \mathbb{R}_n[X] \to \mathbb{R}^{n+1}, \qquad L(P(X)) = (P(t_1), P(t_2), \ldots, P(t_{n+1}))^{\top}.$$

Determinar el rango de L.

Capítulo 5

Autovalores y autovectores

El objetivo de este capítulo es estudiar los conceptos de autovalor y autovector y la posibilidad de que un endomorfismo admita una base en la que su matriz sea diagonal. Explicaremos las ideas generales más adelante pero es necesario tener unos conocimientos básicos sobre polinomios, y con esto empezamos.

5.1. Una breve visita a los polinomios

Suponemos que el lector tiene un mínimo de familiaridad con los polinomios, en el sentido de que sabe operar con ellos. Al escribir $P(X) = a_0 + a_1X + a_2X^2 + \ldots + a_nX^n$ se supone, salvo aviso contrario, que $a_n \neq 0$ y el grado[1] de $P(X)$ es $n = \deg P(X)$. Si este a_n es 1, el polinomio de llama **mónico.** El polinomio con todos los $a_i = 0$ no tiene grado definido, si bien muchos autores le atribuyen el grado $-\infty$, cosa útil para simplificar ciertas afirmaciones. Por ejemplo, se suele decir "el grado del producto es la suma de los grados". Si deg(0) no está definido habría que añadir "supuestos $P(X)$ y $Q(X)$ no nulos". Si se admite $\deg(0) = -\infty$, es siempre cierta

$$\deg P(X)\,Q(X) = \deg P(X) + \deg Q(X)$$

contando con que $-\infty + n = -\infty$. Preferimos el primer convenio. Los polinomios de grado 1; o sea, $P(X) = a_0 + a_1X$ se llaman **lineales** (que es equívoco, aunque todos lo decimos, porque no son funciones lineales de $\mathbb{R}$ en $\mathbb{R}$); los de grado 2 son **cuadráticos** y los de grado 3 **cúbicos**. Admitiremos sin demostración el siguiente teorema esencial.

Teorema 90 *Dados polinomios $A(X)$ y $B(X)$ con $B(X) \neq 0$, existen polinomios únicos $Q(X)$ y $R(X)$, llamados el **cociente** y **resto** de la división, que cumplen*

$$\textbf{(a)} \quad A(X) = B(X)\,Q(X) + R(X), \qquad \textbf{(b)} \quad \deg B(X) > \deg R(X) \ \text{ o bien } R(X) = 0.$$

Por definición, si el resto es cero, se dice que $A(X)$ es **múltiplo** de $B(X)$ o que $B(X)$ es **divisor** de $A(X)$. Un caso particular muy importante es aquel donde $B(X) = X - c$ con $c \in \Bbbk$. Partimos de

$$A(X) = \sum_{i=0}^{m} a_iX^i = (X-c)\,Q(X) = (X-c)\sum_{j=0}^{m-1} q_jX^j + r$$

siendo r el polinomio resto $R(X)$ que, como es nulo o de grado 0, está en $\Bbbk$. Naturalmente, pretendemos encontrar las q_j y r del mejor modo posible. Tenemos

$$\begin{aligned}(X-c)\,Q(X) &= \sum_{j=0}^{m-1} q_jX^{j+1} - \sum_{j=0}^{m-1} cq_jX^j + r \overset{1}{=} \sum_{i=1}^{m} q_{i-1}X^i - \sum_{j=0}^{m-1} cq_jX^j + r \\ &= q_{m-1}X^m + \sum_{i=1}^{m-1} (q_{i-1} - cq_i)\,X^i + (r - cq_0)\,.\end{aligned}$$

[1] Se usa deg por la influencia del inglés y francés (degree y degré).

En $\overset{1}{=}$ hacemos un cambio de variable *en el índice*, con $j+1=i$ y $j=i-1$, moviéndose de este modo i en $1,2,\ldots,m$ ya que j lo hacía en $0,1,\ldots,m-1$. Igualando coeficientes co $A(X)=\sum_{i=0}^{m}a_iX^i$ quedan las ecuaciones

$$\left\{\begin{array}{l} a_m=q_{m-1}\\ a_i=q_{i-1}-cq_i\\ a_0=r-cq_0\end{array}\right. \text{ que equivalen a } \left\{\begin{array}{l} q_{m-1}\\ q_{i-1}=a_i+cq_i\\ r=a_0+cq_0\end{array}\right. , \text{ siendo } i=m-1,m-2,\ldots,1,$$

y que permiten calcular las q_j en orden *decreciente* de índices. La manera práctica de disponer los cálculos es escribir la tabla

$$\begin{array}{cccccccccc} & a_m & a_{m-1} & a_{m-2} & \cdots & a_i & a_i & \cdots & a_1 & a_0\\ c & \bigstar & cq_{m-1} & cq_{m-2} & \cdots & cq_i & cq_{i-1} & \cdots & cq_1 & cq_0\\ & q_{m-1}=a_m & q_{m-2} & q_{m-3} & \cdots & q_{i-1} & q_{i-2} & \cdots & q_0 & r\end{array}$$

que ha de irse completando en este orden: (a) se escriben los a en la primera fila; (b) se escribe $q_{m-1}=a_m$ en el rincón inferior izquierdo; (c) se pone cq_{m-1} en el lugar indicado y, en la tercera fila, el resultado q_{m-2} de $a_{m-2}+cq_{m-1}$, con lo que se completa la segunda columna; (d) si se ha llegado hasta la columna de a_i, que tiene q_{i-1} al pie, se completa la columna siguiente con a_{i-1}, cq_{i-1} y $q_{i-2}=a_{i-1}+cq_{i-2}$; (e) se completa hasta la penúltima columna que acaba en q_0 y la última que da r. Siempre el elemento al pie de la columna la suma de los dos que tiene encima. Este algoritmo de cálculo al dividir por polinomios $X-c$ se conoce por **algoritmo de Ruffini** o **de Hörner.**

Damos como ejemplo $A=X^5+X+1$ y $c=2$. Se tiene

$$\begin{array}{cccccccc} & 1 & 0 & 0 & 0 & 1 & 1 & \\ 2 & \bigstar & 2 & 4 & 8 & 16 & 34 & ,\\ & 1 & 2 & 4 & 8 & 17 & 35 & \end{array} \qquad Q=X^4+2X^3+4X^2+8X+17, \quad r=35,$$

$$X^5+X+1=(X-2)\left(X^4+2X^3+4X^2+8X+17\right)+35.$$

Problema 188 *Usar el algoritmo obteniendo Q y R para*

$$A=X^3+2X^2+3X+4, \qquad B=X+2.$$

Si quiere más práctica se puede poner algún caso más o proponerlo a otros.

Los polinomios se han supuesto con coeficientes en un cuerpo $\Bbbk$, aunque casi siempre será $\Bbbk=\mathbb{R},\mathbb{C}$. Una **raíz** de $P(X)$ es un elemento $c\in\Bbbk$ tal que $0=P(c)=a_0+a_1c+a_2c^2+\ldots+a_nc^n$.

Dado $P(X)=a_0+a_1X+a_2X^2+\ldots+a_nX^n$, al escribirlo en la forma

$$P(X)=a_0+a_1X+a_2X^2+\ldots+a_nX^n=(X-c)\,Q(X)+r, \qquad c,r\in\Bbbk.$$

se deduce inmediatamente que $r=P(c)$ y que c es raíz de $P(X)$ si y solo si $X-c$ divide a $P(X)$. Si c es raíz de $P(X)$ hay un *máximo* $m\in\mathbb{N}$ tal que $(X-c)^m$ divida a $P(X)$, o equivalentemente, que se pueda escribir $P(X)=(X-c)^m\,Q(X)$ sin que c sea raíz del polinomio $Q(X)$. Se llama a m la **multiplicidad de la raíz** c, con el adjetivo **simple**, **doble, triple**, etc. para $m=1,2,3,\ldots$. En general, si $m\geq 2$ se dice que la raíz es **múltiple**. Si se tiene la raíz c_1 de $P(X)$ de multiplicidad m_1, se puede buscar una raíz c_2 de multiplicidad m_2 de $Q(X)=Q_1(X)$, y escribir

$$Q_1(X)=(X-c)^{m_2}Q_2(X), \qquad P(X)=(X-c_1)^{m_1}Q_1(X)=(X-c_1)^{m_1}(X-c_2)^{m_2}Q_2(X).$$

Esto se puede reiterar buscando una raíz c_3 de $Q_2(X)$ de multiplicidad m_3 y factorizar $Q_2(X)=(X-c_3)^{m_3}Q_3(X)$, etcétera, y llegar finalmente a

$$P(X)=(X-c_1)^{m_1}(X-c_2)^{m_2}\cdots(X-c_k)^{m_k}Q_k(X)$$

de forma que $Q_k(X)$ no tenga raíces. Obsérvese respecto a los grados que

$$\deg P(X)=m_1+\deg Q_1(X)=m_1+m_2+\deg Q_2(X)=\ldots=m_1+\ldots+m_k+\deg Q_k(X)$$

de manera que el proceso de llegar a un polinomio sin raíces, que puede ser un polinomio de grado 0 (una constante), se acaba en un número finito de pasos.

Advertimos que es posible que el proceso anterior no pueda iniciarse porque $P(X)$ no tenga siquiera una raíz. Los polinomios X^2+1 y X^2-2 para $\Bbbk=\mathbb{R},\mathbb{Q}$ respectivamente, son ejemplos de ello. También puede pasar que la suma de las multiplicidades de las raíces, que siempre cumple

$$m_1+\ldots+m_k\leq \deg P(X)$$

presente desigualdad estricta. Esto equivale a que el último $Q_k(X)$ no sea constante. Es decir, puede pasar que el número de raíces, contadas con su multiplicidad sea menor estrictamente que el grado del polinomio. Por ejemplo

$$P(X)=(X-1)\left(X^2+1\right),\qquad P(X)=(X-1)\left(X^2-2\right),\qquad \text{siendo } \Bbbk \text{ respectivamente } \mathbb{R},\mathbb{Q}.$$

Nada de esto sucede si $\Bbbk=\mathbb{C}$ por el **teorema fundamental del Álgebra**, que recordamos

Teorema 91 *Todo polinomio no constante en $\mathbb{C}[X]$ tiene al menos una raíz.*

Como consecuencia, todo polinomio complejo $P(X)\in\mathbb{C}[X]$ de grado ≥ 1 se puede factorizar

$$P(X)=\gamma(X-c_1)^{m_1}(X-c_2)^{m_2}\cdots(X-c_k)^{m_k},\quad \gamma\in\mathbb{C},\quad m_1+\ldots+m_k=\deg P(X). \tag{5.1}$$

Una de las razones principales por las que se estudian los espacios complejos y no solo los reales, es porque el teorema fundamental del Álgebra permite clasificaciones más completas y satisfactorias que el caso real y la diagonalización, objeto de este capítulo es un buen ejemplo. Se requieren menos condiciones para diagonalizar una matriz o endomorfismo si $\Bbbk=\mathbb{C}$ que si $\Bbbk=\mathbb{R}$.

El teorema fundamental del Álgebra no es constructivo y para calcular las raíces se necesita en general mucho trabajo de Análisis Numérico y, con todo, se obtienen tan solo valores aproximados. Al final de la sección veremos lo que se puede hacer en la práctica.

Hay que ver con cuidado un punto de cierta sutileza. Como $\mathbb{R}\subset\mathbb{C}$ se puede considerar un polinomio $P(X)\in\mathbb{R}[X]$ como polinomio complejo pero esto lleva a tener que distinguir al hablar de las raíces r de $P(X)$ si son **raíces como polinomio real** o **como polinomio complejo**. Por ejemplo, $P(X)=X^2+1\in\mathbb{R}[X]$ no tiene raíces como polinomio real pues para todo $r\in\mathbb{R}$ es $r^2\geq 0$ y $P(r)\geq 1>0$. Sin embargo, $r=i$ y $r=-i$ son raíces de X^2+1 como polinomio complejo. Igualmente $P(X)=(X-1)\left(X^2+1\right)$ tiene como polinomio real solo la raíz 1 pero como polinomio complejo tiene tres, que son $1,i,-i$. Por supuesto, como sucede a las expresiones largas en Matemáticas, hay siempre tendencia a acortarlas y decir que "i es raíz de X^2+1" pero se debe tener presente que automáticamente implicamos que $X^2+1\in\mathbb{C}[X]$. Como se ve, en la definición de raíz r de $P(X)\in\Bbbk[X]$, hay que tener muy en cuenta que, desde que se escribe $P(X)\in\Bbbk[X]$, se necesita que r esté en $\Bbbk$ (aparte de $P(r)=0$).

Paramos un momento la discusión para decir que hay otra cuestión parecida pero distinta y más sencilla. Si tenemos $P(X)\in\mathbb{C}[X]$ podemos hablar de **las raíces reales del polinomio complejo** $P(X)$, refiriéndonos a aquellas raíces $r\in\mathbb{C}$ de $P(X)$ con parte imaginaria nula: o sea, con $r\in\mathbb{C}$. Por ejemplo, X^2+1 no tiene raíces reales puesto que $i,-i\notin\mathbb{R}$ y el polinomio $(X-1)\left(X^2+1\right)$, de las tres raíces que tiene, solo una es real. Naturalmente, siempre pensamos en X^2+1 y $(X-1)\left(X^2+1\right)$ como polinomios complejos.

Volviendo al tema principal, vamos a ver que si $P(X)$ es un polinomio real, las raíces que tiene como polinomio complejo que no sean reales cumplen una propiedad importante: vienen en pares $\left(\lambda,\bar{\lambda}\right)$, luego si λ es raíz, su conjugado $\bar{\lambda}$ lo es también y además con la misma multiplicidad.

Teorema 92 *Sea $P(X)\in\mathbb{R}[X]$ y $\lambda\in\mathbb{C}$ una raíz no real de $P(X)$ como polinomio complejo. Entonces, $\bar{\lambda}$ es también raíz de $P(X)$ como polinomio complejo. Además, el polinomio $A(X)=(X-\lambda)\left(X-\bar{\lambda}\right)=X^2+\left(\lambda+\bar{\lambda}\right)X+|\lambda|^2$ es un polinomio real que divide a $P(X)$ y hay un máximo $m\in\mathbb{N}$ que cumple*

$$P(X)=[A(X)]^m Q(X),$$

siendo $Q(X)$ un polinomio real que no tiene como polinomio complejo las raíces λ o $\bar{\lambda}$.

Demostración. Aplicamos la conjugación a $P(\lambda)=0$ y, como para todo j es $\bar{a}_j=a_j$,

$$0=\bar{0}=\overline{P(\lambda)}=\overline{a_0}+\overline{a_1\lambda}+\overline{a_2\lambda^2}+\ldots+\overline{a_n\lambda^n})=a_0+a_1\bar{\lambda}+a_2\bar{\lambda}^2+\ldots+a_n\bar{\lambda}^n=P\left(\bar{\lambda}\right),$$

luego $\bar{\lambda}$ es una raíz. Podría suceder que λ y $\bar{\lambda}$ fueran raíces pero con *distinta* multiplicidad. Sin embargo esto es imposible. Si por ejemplo fuese $m_2 - m_1 = k > 0$ para las multiplicidades m_1 y m_2 de λ y $\bar{\lambda}$, podríamos factorizar

$$P(X) = \left[\gamma (X-\lambda)^{m_1} \left(X-\bar{\lambda}\right)^{m_1}\right] \left(X-\bar{\lambda}\right)^{k} Q(X).$$

de modo que $Q(X)$ no tuviera como raíces a λ ni a $\bar{\lambda}$. El polinomio $P_1(X) = \left(X-\bar{\lambda}\right)^{k} Q(X)$, cociente de $P(X)$ al dividirlo por $\gamma (X-\lambda)^{m_1} \left(X-\bar{\lambda}\right)^{m_1}$, ha de ser un polinomio real (teorema 90) y tiene la raíz $\bar{\lambda}$, luego por la primera parte, debería ser también $P_1(\lambda) = 0$ y $0 = \left(\lambda-\bar{\lambda}\right)^{k} Q(\lambda)$, que es imposible porque $Q(\lambda) \neq 0$ y $\lambda - \bar{\lambda} \neq 0$ (aquí se usa que $\lambda \notin \mathbb{R}$). Esta contradicción nos dice que $m_1 = m_2 = m$. Podemos por tanto factorizar

$$P(X) = \gamma (X-\lambda)^{m} \left(X-\bar{\lambda}\right)^{m} Q(X) = \gamma \left[(X-\lambda)\left(X-\bar{\lambda}\right)\right]^{m} Q(X) = A(X) = \gamma A(X)^{m} Q(X)$$

sin que $Q(X)$ tenga a λ o $\bar{\lambda}$ como raíces. Es trivial que $A(X)$ es un polinomio real y también $Q(X)$ ya que es cociente de polinomios reales (teorema 90). ♣

Teorema 93 *Todo polinomio real $P(X)$ se puede factorizar en la forma*

$$P(X) = \gamma (X-c_1)^{m_1} \cdots (X-c_k)^{m_k} \left[X^2+u_1X+v_1\right]^{n_1} \cdots \left[X^2+u_\ell X+v_\ell\right]^{n_\ell}, \tag{5.2}$$

siendo $c_1, \ldots, c_k$ las raíces reales de $P(X)$ con multiplicidades respectivas $m_1, \ldots, m_k$ y los polinomios $X^2+u_iX+v_i$ con coeficientes reales y con dos raíces complejas no reales, que son conjugadas una de otra, ambas con multiplicidad n_i. La relación entre grados y multiplicidades de raíces es

$$\deg(P(X)) = m_1 + \ldots + m_k + 2n_1 + \ldots + 2n_\ell.$$

Demostración. Por los teoremas 91 y 92 sabemos que, contadas con su multiplicidad, hay tantas raíces complejas como $\deg(P(X))$ con las complejas no reales en pares $\left(\lambda_1, \bar{\lambda}_1\right), \ldots, \left(\lambda_\ell, \bar{\lambda}_\ell\right)$ y λ_j y $\bar{\lambda}_j$ con la misma multiplicidad n_j. Sean $c_1, \ldots, c_k$ las raíces reales. La factorización (5.1) nos da

$$\begin{aligned} P(X) &= \gamma (X-c_1)^{m_1} \cdots (X-c_k)^{m_k} \left[(X-\lambda_1)\left(X-\bar{\lambda}_1\right)\right]^{n_1} \cdots \left[(X-\lambda_\ell)\left(X-\bar{\lambda}_\ell\right)\right]^{n_\ell} \\ &= \gamma (X-c_1)^{m_1} \cdots (X-c_k)^{m_k} \left[X^2+u_1X+v_1\right]^{n_1} \cdots \left[X^2+u_\ell X+v_\ell\right]^{n_\ell} \end{aligned}$$

con $\deg(P(X)) = m_1 + \ldots + m_k + 2n_1 + \ldots + 2n_\ell$. ♣

Si $\deg P(X)$ es impar tiene que ser algún $m_i > 0$, lo que significa que hay raíces reales y se obtiene el importante corolario: *los polinomios reales de grado impar tienen al menos una raíz real.* Admitiremos que las factorizaciones descritas de los polinomios reales y complejos son únicas salvo por el orden de los factores.[2]

Problema 189 *Probar que los polinomios* mónicos *reales sin raíces reales de grado 2 son los que se pueden escribir en la forma $P(X) = (X-p)^2 + q^2$ con $q > 0$. Las raíces complejas son entonces $p \pm qi$.*

Con este problema vemos que los factores de segundo grado en (5.2) pueden describirse de dos maneras (a) como de la forma X^2+uX+v precisando que no hay raíces reales (equivalente a $u^2 - 4v < 0$); o bien (b) como de la forma $(X-p)^2 + q^2$ con $q > 0$.

¿Cuántas posibilidades hay de escribir un polinomio *mónico* de segundo grado real en la forma (5.2) según sean sus raíces múltiples o no o sean reales o no reales. Con un poco de método es fácil la respuesta. Si todas las raíces son reales, puede haber dos o una doble; es decir, $c_1 \neq c_2$ o $c_1 = c_2 = c$. Si las hay complejas no reales, solo se tiene la posibilidad de raíces $\lambda, \bar{\lambda}$ distintas. En resumen, el polinomio es

$$(X-c_1)(X-c_2), \qquad (X-c)^2, \qquad (X-\lambda)\left(X+\bar{\lambda}\right) = (X-p)^2 + q^2.$$

Problema 190 *Hacer el trabajo similar para un polinomio real mónico de grado* 3.

Queda una cuestión pendiente. ¿Cómo se calculan las raíces? La respuesta es evidente si nos dan factorizado $P(X)$ en la forma (5.1) o (5.2), pero lo normal es que nos den el polinomio sin factorizar y que haya que calcularlas, precisamente para factorizarlo. Si el polinomio es de segundo grado, tenemos

[2] Tras definir en $\mathbb{k}[X]$ arbitrario **polinomio irreducible**, se puede probar que la factorización de un polinomio con sus factores irreducibles es única salvo por el orden de factores.

las fórmulas bien conocidas; si es de tercero o cuarto grado hay fórmulas para las raíces descubiertas por matemáticos italianos del Renacimiento, aunque son muy complicadas. Un hito de las Matemáticas fue probar que no hay fórmula "general y análoga a las anteriores" (ponemos las comillas para reconocer una imprecisión) si el grado es mayor o igual a 5. En resumen, no hay procedimiento general y solo se puede decir que hay algoritmos que con un ordenador calculan por aproximación pero nunca con exactitud estas raíces. Pronto veremos que en muchos casos son números irracionales y una cifra con un número finito de decimales es forzosamente imprecisa. Los problemas que pueden tener una solución completa si $\deg P(X) \geq 3$ se basan en poder "adivinar" una o varias raíces de $P(X)$, dividir por los monomios correspondientes y conseguir un cociente de grado ≤ 2 donde podemos usar la fórmula tradicional. Para "adivinar raíces" disponemos del siguiente teorema

Teorema 94 *Sea $P(X) = a_0 + a_1X + a_2X^2 + \cdots a_mX^m$ un polinomio real o complejo con coeficientes en $\mathbb{Z}$. Si $\xi = r/s$ es raíz* racional *de $P(X)$ con $\xi = r/s$ fracción irreducible (o sea, el máximo común divisor de r y s es 1) se tiene que r debe dividir a a_0 y s debe dividir a a_m. En particular, si el coeficiente de X^m es ± 1, las raíces racionales (si existen) tienen que ser números enteros r que dividan a a_0.*

Por ejemplo, si $P(X) = X^3 + X^2 + 4X + 4$, sus raíces enteras, si existen, estarán en $\{\pm 1, \pm 2, \pm 4\}$. Probamos si vale $c = 4$ con la regla de Ruffini.[3] Los cálculos son

$$\begin{array}{cccccc} & 1 & 1 & 4 & 4 & \\ 4 & \bigstar & 4 & 20 & 96 & \text{que da} \\ & 1 & 5 & 24 & 100 & \end{array} \left\{ \begin{array}{l} P(X) = \left(X^2 + 5X + 24\right)(X-4) + 100 \\ P(4) = 100 \neq 0 \end{array} \right.$$

y 4 no es raíz. El lector puede verificar que -1 es raíz porque

$$\begin{array}{cccccc} & 1 & 1 & 4 & 4 & \\ -1 & \bigstar & -1 & 0 & -4 & \text{que da} \\ & 1 & 0 & 4 & 0 & \end{array} \left\{ \begin{array}{l} P(X) = \left(X^2 + 4\right)(X+1) \\ P(-1) = 0 \end{array} \right.$$

Las raíces de $X^2 + 4$ son $\pm 2i$ luego $-1, \pm 2i$ son las tres raíces de $P(X) = X^3 + X^2 + 4X + 4$.

Subrayamos que el teorema no dice como encontrar las raíces de P sino solo nos da un conjunto manejable donde podemos encontrar si hay suerte una raíz fácil, y fácil significa en esta situación que es un número entero (si $a_m = \pm 1$) o racional. Ilustramos el uso del teorema preparando un problema. Definimos

$$P(X) = (X - (1+i))(X - (1-i))(X-1)(X+1) = X^4 - 2X^3 + X^2 + 2X - 2.$$

Problema 191 *Calcular las raíces de $P(X) = X^4 - 2X^3 + X^2 + 2X - 2$ y factorizarlo como polinomio real y como polinomio complejo* ♦

Solución. Naturalmente, hemos hecho trampa, y se sabe que la factorización compleja es la de arriba y las raíces complejas son ± 1 y $1 \pm i$. Pero se supone que no lo sabemos. El polinomio es mónico, luego, *si tiene raíces enteras,* deben estar en el conjunto $\{\pm 1, \pm 2\}$ de divisores enteros de -2. Probamos con Ruffini como más arriba y

$$\begin{array}{cccccc} & 1 & -2 & 1 & 2 & -2 \\ 1 & \bigstar & 1 & -1 & 0 & 2 \\ & 1 & -1 & 0 & 2 & 0 \end{array}, \qquad \begin{array}{cccc} 1 & -1 & 0 & 2 \\ \bigstar & -1 & 2 & -2 \\ 1 & -2 & 2 & 0 \end{array}$$

luego $P(X) = (X-1)\left(X^3 - X^2 + 2\right) = (X-1)(X+1)\left(X^2 - 2X + 2\right)$. El polinomio $X^2 - 2X + 2$ tiene raíces complejas $1 + i, 1 - i$. Como polinomio real, $P(X)$ tiene dos raíces ± 1 y la factorización $(X-1)(X+1)\left(X^2 - 2X + 2\right)$. Como polinomio complejo, $P(X)$ tiene cuatro raíces simples $\{\pm 1, 1 \pm i\}$ y la factorización es la de antes del enunciado. ♦

Problema 192 *Calcular las raíces y factorizar $P_1(X) = X^3 - 2X + 1$ y $P_2(X) = X^4 - 1$ como polinomios reales o complejos.*

[3] Lo usual es empezar con ± 1 pero elegimos a posta $c = 4$ para que vea el lector que esta regla facilita el cálculo de $P(c)$ y si $P(c) \neq 0$, c no es raíz.

Problema 193 *En cualquier* $\Bbbk$ *se tiene que* $P(X) = X^n - 1$ *tiene la raíz* 1. *Factorizar* $X^n - 1 = (X-1)\,Q(X)$ *explicitando* $Q(X)$.

Problema 194 *Probar que la raíz enésima de un número natural, si no es otro número natural, es un número que no es racional.*[4]

El problema generaliza muchísimo la afirmación de que $\sqrt{2}$ es irracional. Como un número irracional tiene un desarrollo decimal ilimitado, se ve que el cálculo *exacto* de raíces (de polinomios o de números) es casi siempre imposible, aunque se puedan conseguir buenas aproximaciones.

5.2. Autovalores y autovectores

En todo el capítulo $L : \mathbb{E} \to \mathbb{E}$ será un endomorfismo del espacio vectorial $\mathbb{E}$ sobre un cuerpo $\Bbbk$ arbitrario. Para algunas cuestiones no se requiere dimensión finita, pero sí para la mayoría, y n será $\dim(\mathbb{E})$.

Un **autovector** o **vector propio** (también llamado **vector característico**)[5] de L es un $x \in \mathbb{E}$ *no nulo* tal que $L(x) = \lambda x$ para cierto $\lambda \in \Bbbk$. A λ se le llama el **autovalor** o **valor propio** (también **valor característico**) correspondiente a x. Obsérvese que x determina unívocamente a λ, pues $L(x) = \lambda_1 x = \lambda_2 x$ da $(\lambda_1 - \lambda_2)\,x = 0$ y, al ser $x \neq 0$, obtenemos $\lambda_1 - \lambda_2 = 0$. Es fácil probar que si $y = \mu x$, con $\mu \neq 0$, también y es vector propio con el mismo valor propio λ. Por la propia definición, si λ es valor propio, $\ker(L - \lambda\,\mathrm{id}_{\mathbb{E}}) \neq 0$. Se llama al subespacio $\ker(L - \lambda\,\mathrm{id}_{\mathbb{E}})$ el **subespacio propio de** λ.[6] Está formado por todos los vectores propios del valor propio λ junto con 0 y se denota también por $\mathbb{E}(\lambda)$, sobrentendiendo L. En el caso en el que sea $\dim(\mathbb{E}) < \infty$, el subespacio $\ker(L - \lambda\,\mathrm{id}_{\mathbb{E}})$ tendrá obviamente dimensión finita, llamada la **multiplicidad geométrica del valor propio** λ. El caso que más va a aparecer es aquel en que se tiene una matriz *cuadrada* $a \in \Bbbk^{n\times n}$ y $L : \Bbbk^n \to \Bbbk^n$ es $L(x) = ax$ (o sea, L está asociado a la matriz a). Se habla entonces de **autovectores, autovalores, subespacios propios** de la *matriz* a.[7]

Antes de tratar ejemplos vamos a discutir algo informalmente la importancia de estos conceptos. Hay muchos más vectores propios de los que pudiera parecer y es frecuente, *aunque no siempre se consigue*, que haya incluso una base $\mathcal{U} = (u_1, \ldots, u_n)$ formada por vectores propios (con diferentes valores propios, quizás repetidos). Al ser $L(u_j)$ proporcional a u_j, es inmediato que la matriz de L en $\mathcal{U}$ es diagonal, lo que tiene enormes ventajas para estudiar L. Hay que decir antes de seguir que podemos encontrar casos en donde L no tiene vectores propios (recordemos que deben ser no nulos). Por ejemplo, en $\mathbb{R}^2$,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} -x^2 \\ x^1 \end{pmatrix},\ \text{y si es proporcional a } \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \text{ obtenemos } \left(x^1\right)^2 + \left(x^2\right)^2 = 0,$$

luego $x = 0$. Sin embargo, cambiando un signo en la matriz,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} x^2 \\ x^1 \end{pmatrix},\ \text{y es proporcional a } \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \text{ cuando } x^1 = x^2.$$

Como ejemplos de máxima sencillez tenemos los casos $L(x) = \kappa x$, $\kappa \in \Bbbk$ en donde todos los vectores son propios. Este ejemplo incluye al endomorfismo 0 y a $\mathrm{id}_{\mathbb{E}}$. Si $\mathbb{E} = \mathbb{R}[X]$, los polinomios de grado arbitrario, $0 \in \mathbb{R}$ es un valor propio de D, cuyo espacio propio son los polinomios constantes. Si $\mathbb{E}$ es el espacio de las funciones continuas y reales en $I = [0,1]$ y

$$L(f(t)) = \left(\int_0^1 f(\tau)\,d\tau\right) t$$

es inmediato que $f(t) = t$ es vector propio con valor propio 1.[8]

[4] Debe usarse que si p primo divide a un producto de enteros debe dividir al menos a uno de ellos. Por cierto, si se usa este resultado sobre números primos se demuestra enseguida que $\mathbb{Z}_p$ con p primo es un cuerpo, aunque el teorema de Bezout es el que nos da el procedimiento efectivo de cálculo del inverso.

[5] En inglés, por influencia del alemán, se usa mucho *eigenvector*.

[6] Identificaremos $\lambda \in \Bbbk$ con $\lambda\,\mathrm{id}_{\mathbb{E}}$ o, en el caso de matrices, con λI. Por tanto, el subespacio propio será $\ker(L - \lambda)$.

[7] El autor usa con mayor frecuencia "...propio" pero hay un pequeño riesgo porque un subconjunto propio de X es otro Y tal que $Y \neq X$. En ese sentido, un subespacio propio de $\mathbb{E}$ podría ser dos cosas diferentes, pero la confusión es improbable.

[8] Suena raro llamar "vector" a una función pero, desde el momento en el que decimos que un cierto espacio de funciones es un espacio vectorial, está justificado llamar a las funciones vectores.

Problema 195 *Sea $\mathbb{E}$ el espacio de las funciones continuas y reales en $I=[0,1]$ y $L(f(t))=\int_0^t f(\tau)\,d\tau$. Probar que no tiene vectores propios.*[9]

5.3. El polinomio característico

Obtener los valores propios en dimensión finita necesita otro concepto tan esencial como los anteriores: el **polinomio característico** (de L). Se va a definir primero el **polinomio característico de la matriz** $a\in\Bbbk^{n\times n}$. Se introduce X como variable polinómica y se define este polinomio como

$$C(X)=\begin{vmatrix} a_1^1-X & a_2^1 & \cdots & a_{n-1}^1 & a_n^1\\ a_1^2 & a_2^2-X & \cdots & a_{n-1}^2 & a_n^2\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ a_1^{n-1} & a_2^{n-1} & \cdots & a_{n-1}^{n-1}-X & a_n^{n-1}\\ a_1^n & a_2^n & \cdots & a_{n-1}^n & a_n^n \end{vmatrix}$$
$$=\det(a-XI_n)=\det(a_1-Xe_1,a_2-Xe_2,\ldots,a_n-Xe_n).$$

La segunda notación viene abreviada viendo al determinante como función de las columnas. Antes de saber para qué vale $C(X)$ vamos ir viendo cómo es. Por ejemplo,

$$\text{si } a=\begin{pmatrix} p & -q\\ q & p\end{pmatrix},\ \text{entonces } C(X)=\det\begin{pmatrix} p-X & -q\\ q & p-X\end{pmatrix}=X^2-2pX+p^2+q^2=(p-X)^2+q^2.$$

Debe recordarse siempre lo fácil que es el cálculo de $C(X)$ para matrices triangulares porque

$$\text{si } a=\begin{pmatrix} \alpha_1 & * & \cdots & *\\ 0 & \alpha_2 & \cdots & *\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \alpha_m\end{pmatrix},\ \text{entonces } C(X)=(\alpha_1-X)(\alpha_2-X)\cdots(\alpha_m-X);$$

(se supone que bajo la diagonal todo son ceros).Un último ejemplo es

$$a=\begin{pmatrix} 1 & 1 & 1 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 1 & 1 & 1 & 1\end{pmatrix}, \tag{5.3}$$

y entonces, desarrollando por la primera columna,

$$C(X)=\begin{vmatrix} 1-X & 1 & 1 & 1\\ 0 & 1-X & 0 & 0\\ 0 & 0 & 1-X & 0\\ 1 & 1 & 1 & 1-X\end{vmatrix}=(1-X)\begin{vmatrix} 1-X & 0 & 0\\ 0 & 1-X & 0\\ 1 & 1 & 1-X\end{vmatrix}-\begin{vmatrix} 1 & 1 & 1\\ 1-X & 0 & 0\\ 0 & 1-X & 0\end{vmatrix}$$
$$=(1-X)^2\begin{vmatrix} 1-X & 0\\ 0 & 1-X\end{vmatrix}-\begin{vmatrix} 1-X & 0\\ 0 & 1-X\end{vmatrix}=(1-X)^4-(1-X)^2=(1-X)^2X(X-2).$$

Es posible que el lector haya observado que en los tres ejemplos hemos dado $C(X)$ *factorizado*; o sea, salvo signo, es producto de factores tipo $(X-\lambda)^m$. No siempre se puede conseguir, pero es muy conveniente para el uso que vamos a dar a $C(X)$.

¿Cómo es $C(X)$? Necesitamos la fórmula $C(X)=\det(a_1-Xe_1,a_2-Xe_2,\ldots,a_n-Xe_n)$. Al desarrollar hay 2^n sumandos, dos de los cuales son

$$\det(a_1,a_2,\ldots,a_n)=\det(a)\ \text{ y }\ \det(-Xe_1,-Xe_2,\ldots,-Xe_n)=(-1)^nX^n\det(e_1,e_2,\ldots,e_n)=(-1)^nX^n.$$

Los otros son de la forma $\det(\ldots,Xe_i,\ldots,a_j,\ldots)$, donde los Xe_i aparecen en k huecos y no aparecen en los $n-k$ huecos restantes. Por ejemplo en el caso $n=3$,

$$\begin{aligned}C(X)=&\det(a_1,a_2,a_3)+\det(-Xe_1,a_2,a_3)+\det(a_1,-Xe_2,a_3)+\det(a_1,a_2,-Xe_3)\\&+\det(a_1,-Xe_2,-Xe_3)+\det(-Xe_1,a_2,-Xe_3)+\det(-Xe_1,-Xe_2,a_3)\\&+\det(-Xe_1,-Xe_2,-Xe_3).\end{aligned}$$

[9] El problema es difícil pues se intuye que hay que usar algo de Análisis pero no se sabe qué. No está mal el que haya alguna vez un problema de ese tipo que, guardado en memoria, nos aparezca un feliz día como sencillo.

El primer sumando es el término constante de $C(X)$ y vale $\det(a)$; sacando X en los tres restantes de la fila 1 queda el monomio de grado 1; sacando X en los tres de la fila 2 queda el monomio de grado 2, y finalmente $-X^3$ que es el monomio de grado 3. Damos un ejemplo numérico donde

$$a=\begin{pmatrix}1&0&1\\0&1&0\\1&0&-1\end{pmatrix}\ \text{ y }\ C(X)=\begin{vmatrix}1-X&0&1\\0&1-X&0\\1&0&-1-X\end{vmatrix}.$$

Los sumandos 5 y 8 de la fórmula anterior son

$$\det(-Xe_1,a_2,a_3)=\begin{vmatrix}-X&0&1\\0&1&0\\0&0&-1\end{vmatrix},\quad \det(a_1,-Xe_2,-Xe_3)=\begin{vmatrix}1&0&0\\0&-X&0\\1&0&-X\end{vmatrix}.$$

Para calcular $C(X)$ no se suele usar la descomposición en 2^n sumandos, pero sí es útil para saber cómo van a ser los coeficientes de X^k, al menos para ciertos valores de k.

Teorema 95 *El polinomio característico es de grado n. Los coeficientes de* X^0, X^{n-1} *y* X^n *son respectivamente* $\det(a)$, $(-1)^{n-1}\operatorname{tr}(a)$ *y* $(-1)^n$

Demostración. Hacemos solo lo relativo X^{n-1} pues el resto es sencillo. Se tiene que el monomio de grado $n-1$ es, quitando el signo $(-1)^{n-1}$,

$$\begin{aligned}\sum_{i=1}^{n}\det(Xe_1,\ldots,Xe_{i-1},a_i,Xe_{i+1}\ldots,Xe_n)&=\sum_{i=1}^{n}X^{n-1}\det(e_1,\ldots,e_{i-1},a_i,e_{i+1},\ldots,e_n)\\&=X^{n-1}\sum_{i=1}^{n}\det\left(e_1,\ldots,e_{i-1},\sum_{j=1}^{n}a_i^je_j,e_{i+1},\ldots,e_n\right)\\&=X^{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}a_i^j\det(e_1,\ldots,e_{i-1},e_j,e_{i+1},\ldots,e_n)\\&=X^{n-1}\sum_{i=1}^{n}a_i^i\det(e_1,\ldots,e_{i-1},e_i,e_{i+1},\ldots,e_n)\end{aligned}$$

Se usa al final que si $j\neq i$ es $\det(e_1,\ldots,e_{i-1},e_j,e_{i+1},\ldots,e_n)=0$ por repetición de variables. Solo cuentan por tanto los sumandos con $j=i$ que llevan a $X^{n-1}\operatorname{tr}(a)$ como queríamos probar. ♣

Este teorema basta para $n=2$ pues dice que $C(X)=X^2-\operatorname{tr}(a)X+\det(a)$. Para $n=3$,

$$C(X)=-X^3+\operatorname{tr}(a)X^2-SX+\det(a)$$

siendo S la suma de los tres menores de elementos de la diagonal de a. Por ejemplo, si

$$a=\begin{pmatrix}1&0&1\\0&2&0\\1&0&3\end{pmatrix}$$

se tiene que $C(X)=-X^3+6X^2+10X+4$ ya que

$$C(X)=-X^3+(1+2+3)X^2-\left(\begin{vmatrix}2&0\\0&3\end{vmatrix}+\begin{vmatrix}1&1\\1&3\end{vmatrix}+\begin{vmatrix}1&0\\0&2\end{vmatrix}\right)X+\begin{vmatrix}1&0&1\\0&2&0\\1&0&3\end{vmatrix}.$$

Recordamos que a y b en $\mathbb{k}^{n\times n}$ son semejantes (o conjugadas) si existe $c\in\mathbb{k}^{n\times n}$ invertible tal que $a=c^{-1}bc$. Los polinomios característicos de matrices semejantes son iguales. Esto se debe a que

$$C_a(X)=\det(a-XI)\overset{1}{=}\det\left(c^{-1}(a-XI)c\right)=\det\left(c^{-1}ac-Xc^{-1}Ic\right)=\det(b-XI)=C_b(X).$$

Se ha usado en $\overset{1}{=}$ que los determinantes de matrices conjugadas son iguales. Esto nos permite, en dimensión finita, definir el **polinomio característico de un endomorfismo**. Se elige una base $\mathcal{U}$ y

se dispone de $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$. La definición es $C_L(X) = C_a(X)$. La definición no depende de la base elegida porque la matriz b de L en otra base $\mathcal{V}$ cumple

$$a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}_{\mathcal{V}}^{\mathcal{V}}(L)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})\,;\ \text{o sea, } a = c^{-1}bc,$$

y $C_L(X) = C_a(X) = C_b(X)$. Vemos ahora la utilidad del polinomio característico.

Teorema 96 *Son equivalentes los siguientes asertos para $\lambda \in \Bbbk$ y dimensión finita:*

1. *λ es un valor propio de L; o sea,* $\ker(L - \lambda) \neq 0$.
2. *El endomorfismo $(L - \lambda)$ no es invertible.*
3. *λ es raíz del polinomio característico $C(X)$.*

Demostración. En general, un endomorfismo es invertible si y solo si su núcleo no es 0. Con esto, **1** y **2** son equivalentes. Por otra parte, en cualquier base $\mathcal{U}$,

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L - \lambda) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) - \lambda I, \quad \det \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L - \lambda) = \det\left(\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) - \lambda I\right) = C(\lambda).$$

Por tanto, el que $L - \lambda$ no sea invertible equivale a que su determinante sea 0, que es como decir que $C(X)$, evaluado en $X = \lambda$, debe ser 0. ♣

Esto da un procedimiento para determinar los vectores y valores propios de L

1. Se calcula $C(X)$, procurando tomar una base $\mathcal{U}$ en donde la matriz a de L permita conocer $\det(a - XI)$ con facilidad.
2. Se factoriza $C(X) = (-1)^n (X - \lambda_1)^{m_1} \cdots (X - \lambda_p)^{m_p} Q(X)$ con $Q(X)$ sin raíces en $\Bbbk$. Hay que admitir la posibilidad de que $C(X)$ no tenga raíces en $\Bbbk$ en cuyo caso no hay valores ni vectores propios. En todo caso, determinar esta factorización es técnicamente lo más difícil. Sin embargo, aunque no podamos factorizar sabemos que cualquier polinomio de grado n tiene, contadas con su multiplicidad, n raíces como máximo. Este es el número máximo de valores propios.
3. Para cada λ_i se calcula $\ker(L - \lambda_i)$. La palabra "calcular" es un poco ambigua. Se supone que, tras elegir una base $\mathcal{U}$, se expresa $\ker(L - \lambda_i)$ en implícitas o paramétricas. Suele pedirse hacerlo en paramétricas, más concretamente, dando una base de $\ker(L - \lambda_i)$.

Problema 196 *Calcular los valores y subespacios propios de la matriz real (5.3).* ♦

Solución. Se calculó como ejemplo que $C(X) = X(X-1)^2(X-2)$ que tiene tres raíces $\lambda_1 = 0$, $\lambda_2 = 1$ (doble) y $\lambda_3 = 2$. Para resolver $(a - 0)x = 0$ que nos dará los vectores propios de x, hacemos operaciones fila en $a = a - 0$,

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

que da la solución general

$$\begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} x^1 \\ 0 \\ 0 \\ -x^1 \end{pmatrix} = x^1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

y los vectores propios de 0 son los no nulos proporcionales a $(1, 0, 0, -1)^\top$.

Para $\lambda_2 = 1$ se resuelve $(a - 1)x = 0$ y hacemos operaciones fila en $a - 1$,

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

que lleva al sistema

$$\left\{ \begin{array}{l} x^1 = x^4 \\ x^2 = -x^3 - x^4 \end{array} \right. \quad \text{y la solución general} \quad \begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = \begin{pmatrix} x^4 \\ -x^3 - x^4 \\ x^3 \\ x^4 \end{pmatrix} = x^3 \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + x^4 \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Resolviendo con un cálculo similar $(a-2)\,x = 0$ tenemos los espacios propios

$$\mathbb{E}(0) = \lg\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}\right), \quad \mathbb{E}(1) = \lg\left(\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}\right), \quad \mathbb{E}(2) = \lg\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right). \blacklozenge$$

Problema 197 *Tomamos $\Bbbk = \mathbb{R}$. Calcular (si existen) los valores y vectores propios de las matrices*

$$a = \begin{pmatrix} 1 & p \\ 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

con $p \neq 0$. ¿Qué pasa en el caso de a si se toma $\Bbbk = \mathbb{C}$? ¿Y si en a es $\Bbbk = \mathbb{Z}_5$?

Problema 198 *En $\Bbbk = \mathbb{C}$ determinar los valores y espacios propios de a con todo ceros excepto i en la diagonal y justo por encima de ella. Por ejemplo, para $n = 4$,*

$$a = \begin{pmatrix} i & i & 0 & 0 \\ 0 & i & i & 0 \\ 0 & 0 & i & i \\ 0 & 0 & 0 & i \end{pmatrix}.$$

Problema 199 *Sea $\Bbbk = \mathbb{C}$ y $p, q \in \mathbb{R}$ con $q \neq 0$. Consideramos $L : \mathbb{C}^2 \to \mathbb{C}^2$ dado por la matriz*

$$a = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}.$$

Determinar los valores y vectores propios de L. ¿Cuántos subespacios propios hay?

Problema 200 *Sean S y P la simetría y proyección asociadas a la descomposición $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$. Determinar sus polinomios característicos y subespacios propios*

Problema 201 *Sea $\mathbb{E} = \mathbb{R}_k[t]$, el espacio de los polinomios de grado $\leq k$ y $L = D$, la derivada. Calcular el polinomio característico de L y sus subespacios propios.*

5.4. Valores propios de matrices simétricas y hermitianas

Veremos más adelante la importancia de que el polinomio característico sea linealmente factorizable. Esto sucede siempre si $\Bbbk = \mathbb{C}$ pero puede fallar si $\Bbbk = \mathbb{R}$. Mostraremos que para las matrices simétricas en $\mathbb{R}$, $C(X)$ es linealmente factorizable y para las hemitianas en $\mathbb{C}$, si bien ya sabemos que lo es, se da el hecho adicional de que las raíces son *reales.* Vamos a dar en lo posible un tratamiento conjunto. Si $a \in \mathbb{C}^{n\times n}$ recordamos que la adjunta hermitiana es a^* con $(a^*)^i_j = \overline{\left(a^j_i\right)}$ (se traspone y se conjuga); o sea $a^* = \overline{(a^\top)}$ o, si se prefiere, se conjuga y se traspone). En el problema 28 se prueba para matrices no necesariamente cuadradas que (no olvidar conjugar λ)

$$(a+b)^* = a^* + b^*, \quad (\lambda a)^* = \bar{\lambda} a^*, \quad (ab)^* = b^* a^*, \quad (a^*)^* = a.$$

Además, para matrices cuadradas $\operatorname{tr}(aa^*) = \sum_{i,j=1}^n \left|a^i_j\right|^2 \in \mathbb{R}$, aunque a sea compleja, de donde $\operatorname{tr}(aa^*) = 0$ equivale a $a = 0$. Si consideramos $\mathbb{R}^{m\times n}$ como un subconjunto de $\mathbb{C}^{m\times n}$ viendo los números reales como caso particular de los complejos, tenemos que para $a \in \mathbb{R}^{m\times n}$ es $a^* = a^\top$ y recuperamos

fórmulas como $(ab)^\top = b^\top a^\top$. Conviene en lo sucesivo para $\lambda \in \mathbb{C}$ usar la notación λ^* como sinónimo de $\bar{\lambda}$. Si $a \in \mathbb{C}^{n\times n}$ y $x, y \in \mathbb{C}^n$ se cumple[10] que $x^*ay \in \mathbb{C}$ verifica $(x^*ay)^* = y^*a^*x$ y, si es $x = y$, y a es *hermitiana*, queda $(x^*ax)^* = x^*a^*x = x^*ax$. Como $\mu = \mu^*$ equivale a que μ sea real, tenemos que aunque x y a puedan tener coeficientes *complejos*, x^*ax es un número *real*. Parece algo intrascendente, pero no lo es.

Teorema 97 *Si $a \in \mathbb{C}^{n\times n}$ es hermitiana ($a^* = a$) y $x \in \mathbb{C}^n$*

1. *El número $x^*ax \in \mathbb{C}$ es real. Además, si $a = I$, $x^*x = \sum_{i=1}^n \left|x^i\right|^2$, luego $x^*x = 0$ si y solo si $x = 0$.*
2. *Los valores propios de a son reales. Si a, como caso particular, es simétrica, sus valores propios como matriz compleja son reales.*
3. *El polinomio $C(X)$ de una matriz simétrica es linealmente factorizable como polinomio real.*

Demostración. Ya se ha probado **1**. Sea $ax = \lambda x$. Calculamos

$$x^*ax = x^*(\lambda x) = \lambda(x^*x), \quad x^*ax = x^*a^*x = (ax)^* x = \lambda^* x^* x.$$

Llegamos a $\lambda(x^*x) = \lambda^* x^* x$, pero, al ser x propio, $x \neq 0$ y $x^*x \neq 0$. Por cancelación, $\lambda = \lambda^*$, $\lambda \in \mathbb{R}$, y se tiene **2**. Finalmente, si a es simétrica, también es hermitiana, y como matriz compleja cumple $C(X) = (\lambda_1 - X)^{m_1} \cdots (\lambda_p - X)^{m_p}$. Acabamos de ver que los λ_i son reales, luego esta es la factorización real. ♣

Problema 202 *Consideremos las matrices complejas, siendo a hermitiana y b no (pero sí simétrica),*

$$a = \begin{pmatrix} 1 & i & 0 \\ -i & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

Comprobar que $C_a(X)$, a pesar de ser a compleja, es un polinomio real y que (¡por supuesto!) los valores propios de a son reales, pero que $L : \mathbb{C}^3 \to \mathbb{C}^3$, $L(x) = ax$, tiene vectores propios que sí pueden ser complejos. Probar que b no tiene valores propios reales..

Problema 203 *Probar que si $a \in \mathbb{C}^{n\times n}$ es **antihermitiana** (por definición, $a^* = -a$), $x^*ax \in \mathbb{C}$ es imaginario puro. y que si a es hermitiana, ia es antihermitiana.*

5.5. Diagonalización

Sea $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo y $\dim(\mathbb{E}) = n$. Queremos encontrar una base $\mathcal{W} = (w_1, \ldots, w_n)$ tal que $\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = a$, la matriz de L en la base $\mathcal{W}$, sea lo más sencilla posible. Por supuesto, la palabra "sencilla" es muy imprecisa y la entenderemos como que a debe tener "muchos ceros". Si nos hicieran la pregunta, *parecida pero diferente*, sobre si se pueden elegir *dos bases* $\mathcal{W}$ y $\mathcal{V}$ en $\mathbb{E}$ de modo que $\mathrm{mat}_{\mathcal{W}}^{\mathcal{V}}(L)$ tenga "muchos ceros", ahí la respuesta es clara.

Problema 204 *Completar una base $(w_{r+1}, \ldots, w_n)$ de $\ker(L)$ hasta una base $(w_1, \ldots, w_r, w_{r+1}, \ldots, w_n)$ de todo $\mathbb{E}$; probar que $(L(w_1), \ldots, L(w_r)) = (v_1, \ldots, v_r)$ es independiente; ampliar $(v_1, \ldots, v_r)$ hasta una base $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_n)$ de $\mathbb{E}$ y probar que **(a)** La matriz de L en las bases $\mathcal{W}$ y $\mathcal{V}$ tiene todo ceros excepto los r primeros términos de la diagonal; y **(b)** El número r de unos es el rango de r.*

Desde luego esta matriz es bien simple, pero interesa que sea $\mathcal{W} = \mathcal{V}$ y esto hace el problema más difícil. ¿Y si hubiera una base $\mathcal{W}$ formada tan solo por vectores propios? Sería casi tan bueno como lo anterior, porque la matriz sería diagonal y *solo intervendría una base.* Podemos ser más específicos. Si $\lambda_1, \ldots, \lambda_p$ son los *diferentes* valores propios de $\mathcal{W}$ y suponemos reordenados los vectores de $\mathcal{W}$ de modo que $(w_1, \ldots, w_{m_1})$ sean los vectores con valor propio λ_1; $(w_{m_1+1}, \ldots, w_{m_2})$ sean los vectores con valor propio λ_2; … ; y $\left(w_{m_{(p-1)}+1}, \ldots, w_{m_p}\right)$ (dicho con palabras, juntamos primero los u_j con valor propio

[10] Ha de ponerse x^*ay para que x^*, a, e y sean multiplicables.

λ_1, luego lo hacemos con los de valor propio $\lambda_2, \ldots$ y seguimos hasta acabar con los de valor propio λ_p) se tiene que

$$\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = \begin{pmatrix} \lambda_1 I_1 & & & \\ & \lambda_2 I_2 & & \\ & & \ddots & \\ & & & \lambda_p I_p \end{pmatrix}, \quad \text{siendo } I_j \text{ la matriz unidad } m_j \times m_j. \tag{5.4}$$

Como ejemplo, recordamos que en el problema 196 vimos que

$$a = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{5.5}$$

tenía tres subespacios propios

$$\lg\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}\right) = \lg(w_1), \quad \lg\left(\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}\right) = \lg(w_2, w_3), \quad \lg\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) = \lg(w_4) \tag{5.6}$$

correspondientes a los valores propios $0, 1, 2$. En la base (w_1, w_2, w_3, w_4) el endomorfismo $L(x) = ax$ de $\mathbb{R}^4$ tendrá matriz

$$d = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Diremos que L es **diagonalizable** si existe una base $\mathcal{W}$ de $\mathbb{E}$ formada por vectores propios. Obsérvese que $a = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$ es entonces una matriz diagonal. Reciprocamente, si en una base $\mathcal{W}$ la matriz a de L en esa base es diagonal, se tiene $L(w_j)$ de la forma $\lambda_j w_j$ y queda claro que los w_j son vectores propios. Suele darse la matriz a de L en una base auxiliar $\mathcal{U}$. Si hay una base $\mathcal{W}$ de vectores propios, tenemos

$$d = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = \mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)\,\mathrm{mat}_{\mathcal{W}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = \left(\mathrm{mat}_{\mathcal{W}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right)^{-1}\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)\,\mathrm{mat}_{\mathcal{W}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = c^{-1}ac.$$

Poder diagonalizar equivale a que la matriz a de L en una base arbitraria sea similar o semejante a otra d que sea diagonal. La matriz c en $d = c^{-1}ac$ es interpretable como la matriz que expresa la base $\mathcal{W}$ "buena" (porque diagonaliza y simplifica) en términos de la base "mala" $\mathcal{U}$ (porque en ella L es complicada). Si volvemos al ejemplo de arriba

$$c^{-1}ac = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = d.$$

En él L viene dado por a como en (5.5) pero en la base estándar $\mathcal{E}$ que es la "mala". La base "buena", porque diagonaliza, es $(w_1, w_2, w_3, w_4) = \mathcal{W}$ en (5.6), y si concatenamos (vulgo, pegamos las columnas) de los w_j, sale una matriz c que expresa $\mathcal{W}$ en función de $\mathcal{E}$. Como era de esperar, $c^{-1}ac = d$ (lo ha hecho el ordenador).[11]

Si abstraemos lo que nos dice este ejemplo vemos que afirmar que a es diagonalizable, primeramente definido como que $L(x) = ax$ tiene base de vectores propios, tiene una formulación equivalente: hay una c invertible tal que $c^{-1}ac = d$ es diagonal. *Las matrices diagonalizables son las similares o semejantes a matrices diagonales.*

Hemos advertido que puede suceder que L no sea diagonalizable debido a que, aun teniendo L vectores propios, no tenga los suficientes. El concepto clave es el de **multiplicidad algebraica** de λ, que es la multiplicidad m de λ como raíz de $C(X)$. Recordemos que ya se definió la multiplicidad geométrica de

[11] Si se quiere comprobar a mano que $c^{-1}ac = b$ con matrices concretas, es más cómodo verificar que $ac = cb$, sin calcular inversas.

λ como la dimensión d de su subespacio propio. La relación entre d y m es muy importante. Si tomamos la matriz a como ejemplo,

$$a=\begin{pmatrix}1&1\\0&1\end{pmatrix},\quad C(X)=(X-1)^2,\quad \lambda=1,\quad \mathbb{E}(1)=\lg\begin{pmatrix}1\\0\end{pmatrix},\quad d=1,\quad m=2.$$

Aquí se ve que las dos multiplicidades pueden diferir. Los teoremas principales pendientes de exponer dicen que siempre $d\leq m$ (pero quizás sea $d<m$), y que la coincidencia de multiplicidades es cuestión esencial para que L sea diagonalizable.

Para ver que $d\leq m$, conviene una definición previa. Diremos que $\mathbb{F}$, subespacio de $\mathbb{E}$, es **estable** (por L) si $L(\mathbb{F})\subset\mathbb{F}$. Hay muchos ejemplos de subespacios estables.[12] Los más obvios son $0,\mathbb{E},\operatorname{im}(\mathbb{F})=L(\mathbb{F})$ y $\ker(\mathbb{F})$, pero el que nos va a interesar es $\ker(L-\lambda\operatorname{id})=\mathbb{E}(\lambda)$ cuando λ es un valor propio de L. En efecto, si $x\in\mathbb{E}(\lambda)$,

$$(L-\lambda\operatorname{id})(L(x))=L(L-\lambda\operatorname{id})(x)=L(0)=0.$$

Teorema 98 *Si $\mathbb{F}$ es un subespacio estable, el polinomio característico $Q(X)$ de la restricción $M:\mathbb{F}\to\mathbb{F}$ divide al polinomio característico $C(X)$ de L.*

Demostración. Tomamos una base $\mathcal{U}=(u_1,\ldots,u_m,u_{m+1},\ldots,u_n)$ de $\mathbb{E}$ siendo $\mathcal{V}=(u_1,\ldots,u_m)$ base del subespacio $\mathbb{F}$. Por la estabilidad, si $j\leq m$ debe ser $L(u_j)$ combinación lineal tan solo de $(u_1,\ldots,u_m)$ y

$$\operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L)=\begin{pmatrix}a&b\\0_{(m-n)\times m}&c\end{pmatrix},\qquad a\in\Bbbk^{m\times m},\quad b\in\Bbbk^{m\times(n-m)},\quad c\in\Bbbk^{(n-m)\times(n-m)}.$$

Al calcular por cajas $\det\left(\operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L)-XI_m\right)=C(X)$ resulta

$$C(X)=\det(a-XI_m)\det\left(c-XI_{(n-m)}\right).$$

Pero a es la matriz de la restricción M en $\mathcal{V}$, así que $C(X)=Q(X)\det\left(c-XI_{(n-m)}\right)$ mostrando que $Q(X)$ divide a $C(X)$. ♣

Teorema 99 *La multiplicidad geométrica de cada λ es menor o igual que su multiplicidad algebraica.*

Demostración. Aplicaremos el teorema anterior a $\mathbb{F}=\mathbb{E}(\lambda)$. Sea d su dimensión, que por definición es la multiplicidad geométrica. El polinomio $Q(X)$ de la restricción M de L a $\mathbb{E}(\lambda)$ es en este caso $(\lambda-X)^d$. Esto es inmediato porque todos los vectores de $\mathbb{E}(\lambda)$ son propios y en cualquier base de $\mathbb{E}(\lambda)$ la matriz de M es una matriz con todo ceros excepto todo λ en la diagonal principal. Por el teorema anterior, $C(X)=(\lambda-X)^d R(X)$ y al ser, por definición de multiplicidad algebraica, m el máximo valor tal que $(X-\lambda)^m$ divide a $C(X)$, ha de ser $d\leq m$. ♣

Otro teorema interesante, que es herramienta principal en el teorema de diagonalización, es este.

Teorema 100 *Sea $\{\lambda_1,\ldots,\lambda_q\}$ un subconjunto de valores propios* distintos. *Entonces $\mathbb{E}(\lambda_1)+\cdots+\mathbb{E}(\lambda_q)$ es suma directa. Como consecuencia, si $x_1,\ldots,x_q$ son vectores propios de los valores propios distintos $\lambda_1,\ldots,\lambda_q$ tenemos que son independientes.*

Se imagina enseguida que el teorema se aplicará al caso en el que $\{\lambda_1,\cdots,\lambda_q\}$ sea la totalidad de los valores propios y así es. Damos el teorema en forma más amplia porque se facilita la demostración por inducción sobre q. Advertimos que, aunque $\{\lambda_1,\ldots,\lambda_q\}$ sea el conjunto de todos los valores propios, *no se dice* que sea $\mathbb{E}=\mathbb{E}(\lambda_1)\oplus\cdots\oplus\mathbb{E}(\lambda_q)$, y la suma puede ser un subespacio *distinto de* $\mathbb{E}$.

Demostración. El teorema 44 da varias propiedades equivalentes para que una suma sea directa y por ello, cualquiera sirve como definición. Elegimos para poder decir que sea $\mathbb{E}=\mathbb{F}_1\oplus\ldots\oplus\mathbb{F}_k$ el que si $0=x_1+\ldots+x_k$ con todo $x_i\in\mathbb{F}_i$, tiene que ser $x_1=x_2=\ldots=x_k=0$.

Vamos a probar el teorema por inducción sobre q, siendo obvio si $q=1$. Supongámoslo cierto para $q-1$ y probémoslo para q. Sea $0=x_1+\ldots+x_q$ con $x_i\in\mathbb{E}(\lambda_i)$. Entonces $x_1=-x_2-x_3-\ldots-x_q$ y, al ser los x_i propios,

$$L(x_1)=\begin{cases}\lambda_1x_1=-\lambda_1(x_2+x_3+\ldots+x_q)=-\lambda_1x_2-\ldots-\lambda_1x_q\\ L(-x_2-x_3-\ldots-x_q)=-\lambda_2x_2-\ldots-\lambda_qx_q\end{cases}$$

[12] Muchos autores dicen **invariante** en lugar de **estable.** Preferimos "estable" porque nos parece que "invariante" sugiere más bien que para todo $x\in\mathbb{F}$ es $L(x)=x$, una condición mucho más fuerte que la estabilidad de nuestra definición.

Restamos las expresiones y $0 = (\lambda_1 - \lambda_2)\, x_2 - \ldots (\lambda_1 - \lambda_q)\, x_q$, estando este vector en $\mathbb{E}(\lambda_2) + \ldots + \mathbb{E}(\lambda_q)$. Por la hipótesis inductiva, $(\lambda_1 - \lambda_2)\, x_2 = \ldots = (\lambda_1 - \lambda_q)\, x_q = 0$. Recordemos que los λ_i son distintos, luego todos los $\lambda_1 - \lambda_i$ son no nulos, lo que nos lleva a $x_2 = \ldots = x_q = 0$. La igualdad inicial $0 = x_1 + \ldots + x_q$ se reduce a $x_1 = 0$ y obtenemos lo que deseábamos, que es $x_1 = x_2 = \ldots = x_q = 0$.

La última afirmación se sigue de otra más general: si tenemos la suma *directa* $\mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_k$ y k vectores $x_i \in \mathbb{F}_i$ no nulos, entonces $(x_1, \ldots, x_q)$ es independiente. En efecto, si $\mu_1 x_1 + , \ldots, \mu_k x_k = 0$, por lo dicho al principio de la demostración $\mu_1 x_1 = \ldots = \mu_k x_k = 0$ y, al no ser nulos los x_i, $\mu_1 = \ldots = \mu_k = 0$. ♣

Teorema 101 (de diagonalización) *Sea $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo con polinomio característico $C(X)$ y $\{\lambda_1, \ldots, \lambda_h\}$ el conjunto de sus distintos valores propios. Son equivalentes*

1. *L es diagonalizable; o sea, existe una base $\mathcal{W}$ tal que $\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$ es diagonal.*
2. *El polinomio $C(X)$ se puede escribir como $C(X) = \pm (X - \lambda_1)^{m_1} \cdots (X - \lambda_h)^{m_h}$ y las multiplicidades geométricas $d_i = \dim(\mathbb{E}(\lambda_i))$ coinciden con las algebraicas m_i.*
3. *$\mathbb{E}$ es expresable como suma directa de los subespacios propios; o sea, $\mathbb{E} = \mathbb{E}(\lambda_1) \oplus \ldots \oplus \mathbb{E}(\lambda_h)$.*

Demostración. Daremos demostración circular **1** $\Longrightarrow$ **2** $\Longrightarrow$ **3** $\Longrightarrow$ **1**.

Supongamos que se tiene **1**. Podemos si hace falta reorganizar los vectores de $\mathcal{W}$ de modo que vayan juntos en primer lugar los de valor propio λ_1, luego los de λ_2, y finalmente los de λ_h. Entonces $a = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$ es como en (5.4) y se obtiene $C(X)$ en la forma dicha, siendo m_i el número de veces que λ_i aparece en la diagonal. Evidentemente

$$n = \dim(\mathbb{E}) = \deg(C(X)) = m_1 + m_2 + \ldots + m_h.$$

El rango de $a - \lambda_i$ es $n - m_i$ porque $a - \lambda_i$ se transforma enseguida en matriz escalonada o reducida con exactamente m_i filas nulas. Se tiene probado **2** porque

$$d_i = \dim \ker(L - \lambda_i) = \dim(\mathbb{E}) - \mathrm{rg}(a - \lambda_i) = n - (n - m_i) = m_i.$$

Sea **2** cierto. La factorización de $C(X)$ nos da las raíces $\lambda_1, \ldots, \lambda_h$ que, por el teorema 96, son los valores propios. El teorema 100 nos da la suma directa $\mathbb{E}(\lambda_1) \oplus \cdots \oplus \mathbb{E}(\lambda_h)$. Además,

$$\dim(\mathbb{E}(\lambda_1) \oplus \cdots \oplus \mathbb{E}(\lambda_h)) = \dim(\mathbb{E}(\lambda_1)) + \cdots + \dim(\mathbb{E}(\lambda_h)) = d_1 + \ldots + d_h = m_1 + \ldots + m_h = n = \dim(\mathbb{E})$$

y esto implica que $\mathbb{E}(\lambda_1) \oplus \cdots \oplus \mathbb{E}(\lambda_h) = \mathbb{E}$, la condición **3**.

Probamos que **3** implica **1**. En cada $\mathbb{E}(\lambda_i)$ se toma una base $\mathcal{W}_i = \left(w_{(i)1}, w_{(i)2}, \ldots, w_{(i)d_i}\right)$ que, obviamente, está formada por vectores propios con valor propio λ_i. Si $\mathcal{W}$ viene definida pegando ("yuxtaponiendo" es más culto) las bases, tenemos una base de $\mathbb{E}$ en la que L tiene matriz diagonal. ♣

El teorema 101 se usa en primer lugar para saber si L es diagonalizable y esto se consigue verificando que se cumple **2.** Dejando de lado el cálculo de $C(X)$ y sus raíces, hay que saber si $d_i = m_i$. Dado que d_i es dimensión de un núcleo,

$$d_i = \mathrm{nul}(L - \lambda_i) = n - \mathrm{rg}(L - \lambda_i), \quad \text{luego} \quad d_i = m_i \text{ equivale a } n - \mathrm{rg}(L - \lambda_i) = m_i.$$

Volviendo al problema 196, si queremos saber si la matriz a es diagonalizable, necesitamos primero saber que $C(X) = X(X-1)^2(X-2)$ (ya lo hicimos) y luego calcular

$$a - 0 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad a - 1 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad a - 2 = \begin{pmatrix} -1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 \end{pmatrix}$$

y los rangos respectivos $3, 2$ y 3. Como $4-3$, $4-2$ y $4-3$ son las multiplicidades algebraicas de las raíces $0, 1, 2$ de $C(X)$, el teorema 101 nos dice que a es diagonalizable e incluso cómo será la matriz d obtenida al diagonalizar a. Esto, si no necesitamos la base que diagonaliza, evita resolver los sistemas $(a - \lambda_i)\, x = 0$, pero si se necesita, este trabajo es ineludible.

Hay tres problemas típicos de cálculo relacionados con la diagonalización. Cada uno exige un poco más que el anterior.

1. *Saber si es posible diagonalizar* L. Hay que calcular el polinomio característico $C(X)$, lo que supone calcular un determinante, cosa que puede ser más o menos fácil. Después hay que hallar sus raíces y tal como hemos dicho, es un camino con incertidumbre. Lo mejor es que la matriz sea triangular, pues entonces sale factorizado $C(X) = \left(a_1^1 - X\right) \cdots (a_m^m - X)$. En todo caso, si $C(X)$ no es linealmente factorizable, es imposible diagonalizar. En el caso $\Bbbk = \mathbb{C}$, el teorema fundamental del Álgebra elimina este obstáculo, mostrando una de las ventajas de $\mathbb{C}$ sobre $\mathbb{R}$. No obstante, aunque se pueda escribir $C(X) = \left(a_1^1 - X\right) \cdots (a_m^m - X)$, se necesita además que $m_i = d_i = \dim \ker (L - \lambda_i)$. Si disponemos de la matriz a de L en una base auxiliar $\mathcal{U}$, lo que se ha de verificar es $m_i = n - \operatorname{rg}(a - \lambda_i)$, $i = 1, \ldots, h$.

2. *Sabido que* L *es diagonalizable, saber además su matriz diagonal.* Suele elegirse una ordenación $(\lambda_1, \ldots, \lambda_h)$ de los distintos valores propios y entonces la matriz será de la forma (5.4). Esto supone muy poco trabajo adicional respecto a lo hecho para **1**.

3. *Sabido que* L *es diagonalizable, calcular una base* $\mathcal{W}$ *de vectores propios.* Hemos visto en **1** que había que calcular los rangos $\operatorname{rg}(a - \lambda_i)$ pero no se necesitaba resolver los sistemas homogéneos $(a - \lambda_i) x = 0$. Ahora sí que hay que hacerlo. Si se hace en la forma del capítulo primero, tendremos bases del espacio de soluciones, que supone tener en coordenadas los vectores de una base de $\ker (L - \lambda_i)$. Más concretamente, si $\left(s^1, \ldots, s^n\right)^\top \in \Bbbk^n$ es uno de estos vectores, $s^1 u_1 + \ldots + s^n u_n$ es un vector de la base $\mathcal{W}$ dentro de $\ker (L - \lambda_i)$. Con esto tenemos bases $\left(w_{(i)1}, \ldots, w_{(i)m_i}\right) = \mathcal{W}_{(i)}$ de $\ker (L - \lambda_i)$. Afirmamos que

$$\mathcal{W} = \left(w_{(1)1}, \ldots, w_{(1)m_1}, w_{(2)1}, \ldots, w_{(2)m_2}, \ldots, w_{(h)1}, \ldots, w_{(h)m_h}\right)$$

es base de $\mathbb{E}$. En principio, encadenar sucesiones de vectores independientes no garantiza que la cadena lo sea. No obstante, en este caso, cada $\mathcal{W}_{(i)}$ es base de $\ker (L - \lambda_i) = \mathbb{E}(\lambda_i)$ y hay una suma directa $\mathbb{E} = \mathbb{E}(\lambda_1) \oplus \ldots \oplus \mathbb{E}(\lambda_h)$, garantizan el problema 97 que esta sucesión sí forma base. Con estas observaciones podemos prescindir de muchas comprobaciones de independencia.

El siguiente problema puede ahorrar mucho trabajo y es sencillo.

Problema 205 *Supongamos que* L *tiene* $C(X) = (\lambda_1 - X) \cdots (\lambda_n - X)$ *con* n *raíces distintas siendo* $n = \dim(\mathbb{E})$. *Probar que* L *es diagonalizable.*

Problema 206 *En la base* $\mathcal{U} = (u_1, u_2)$ *se tiene que* $L(u_1) = -2u_1 + 10u_2$ *y* $L(u_2) = -2u_1 + 7u_2$. *Probar que* L *es diagonalizable y dar en función de* $\mathcal{U}$ *una base* $\mathcal{V}$ *que diagonalice* L. ♦

Solución. Es inmediato que

$$\operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = a = u \begin{pmatrix} -2 & -2 \\ 10 & 7 \end{pmatrix}.$$

Se tiene $C(X) = (2 - X)(3 - X)$. Los sistemas $(a - 2) x$ y $(a - 3) x$ tienen como espacios de soluciones

$$\operatorname{lg}\left(\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right) \text{ y } \operatorname{lg}\left(\begin{pmatrix} -2 \\ 5 \end{pmatrix}\right)$$

Los vectores v_1 y v_2 dados por $v_1 = -u_1 + 2u_2$ y $v_2 = -2u_1 + 5u_2$ forman una base que diagonaliza L. La matriz de L en $\mathcal{V}$ es diagonal con $(2, 3)$ en ella. ♦

Problema 207 *Sea* $\mathbb{E} = \mathbb{R}_4[X]$ *y* $L : \mathbb{E} \to \mathbb{E}$ *dada por* $L(P(X)) = P'(X) + P''(X)$ *(las primas son derivadas). Estudiar si* L *es diagonalizable.*

Problema 208 *Sea* $\mathbb{E} = \mathbb{C}_1[X]$ *el espacio de los polinomios* $P(X) = \alpha X + \beta$ *con* $\alpha, \beta \in \mathbb{C}$. *Definimos* $L : \mathbb{E} \to \mathbb{E}$ *por* $L(\alpha X + \beta) = (\alpha + i\beta) X + \beta i$. *Estudiar si* L *es diagonalizable y sus subespacios propios.*

Problema 209 *Sea* $\mathbb{E}$ *el espacio de las matrices triangulares superiores* 2×2. *Damos*

$$L : \mathbb{E} \to \mathbb{E}, \qquad L \begin{pmatrix} p & q \\ 0 & r \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p & q \\ 0 & r \end{pmatrix}.$$

Estudiar las multiplicidades algebraicas y geométricas de los valores propios. ¿Es L *diagonalizable?*

Problema 210 *Hallar los valores y vectores propios de*

$$a=\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix},\quad b=\begin{pmatrix} \alpha & -1 & -1 \\ -1 & \alpha & -1 \\ -1 & 1 & \alpha \end{pmatrix},\quad c=\begin{pmatrix} \alpha & 1 & -1 \\ 1 & \alpha & 1 \\ -1 & 1 & \alpha \end{pmatrix},\quad d=\begin{pmatrix} \alpha & 1 & -1 \\ 1 & \alpha & 1 \\ 1 & -1 & \alpha \end{pmatrix}$$

y si se puede diagonalizar. Nota: no nos dicen si $\Bbbk$ *es* $\mathbb{R}$ *o* $\mathbb{C}$*. Esto no importa mucho para* a, b, c*, pero complica la respuesta para la matriz* d*. Se puede asumir que para* d *el cuerpo es* $\mathbb{R}$*.*

Problema 211 *Determinar el polinomio característico y los valores propios de la matriz*

$$a=\begin{pmatrix} \alpha & 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & 1 & 1 & \cdots & 1 \\ 1 & 1 & \alpha & 1 & \cdots & 1 \\ 1 & 1 & 1 & \alpha & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & \alpha \end{pmatrix}$$

con todo α *en la diagonal principal y los demás coeficientes* 1*. ¿Es* a *diagonalizable? ¿Cómo será su forma diagonal? ¿Y los espacios propios y base diagonalizable? Indicación: Poner* $\alpha = 1+\beta$ *y usar el problema 176 de* Determinantes*.*

Damos ahora problemas menos calculísticos.

Problema 212 *Sea* $P(X)$ *un polinomio y* a *una matriz* $n \times n$*. Sustituimos* X *por* a *y* $P(a) = b$ *es una nueva matriz.*[13] *Probar que si* $b = P(a) \in \Bbbk^{n\times n}$ *y* v *es vector propio de* a *con valor propio* λ*, entonces* v *es también vector propio de* b *con valor propio* $P(\lambda)$*. Aplicar esto para probar que*

1. *Si* $a \neq 0$ *tiene una potencia que se anula; digamos que* $a^k = a \cdot a \cdot \cdots \cdot a$ *(*k *veces) es la matriz cero, entonces* a *no es diagonalizable.*
2. *Si* a *diagonalizable, al sustituir* a *en su polinomio característico* $C(X)$ *sale la matriz cero.*

Se ha probado **2** en el problema suponiendo que a es diagonalizable, pero sucede que esto es cierto *sea como sea la matriz.* Es el fundamental **teorema de Cayley-Hamilton**: *al sustituir la matriz* a *o el endomorfismo* L *en su polinomio característico* $C(X)$ *resulta la matriz o endomorfismo* 0. Es el teorema 102, demostrado aquí solo parcialmente, que luego tendrá demostración completa. Por ahora lo admitiremos y tiene muchas aplicaciones.

Lo aplicamos para obtener una fórmula más económica de la inversa de la matriz a. Si a es invertible, $\det(a) = d \neq 0$. Sabemos por el teorema 95 que d es el término constante de $C(X)$. Por tanto

$$0 = C(a) = d + \alpha_1 a + \alpha_2 a^2 + \ldots + \alpha_n a^n, \quad -d = \left(\alpha_1 + \alpha_2 a + \ldots + \alpha_n a^{n-1}\right) a,$$

y la inversa es

$$a^{-1} = \frac{-1}{d}\left(\alpha_1 + \alpha_2 a + \ldots + \alpha_n a^{n-1}\right).$$

Para matrices 2×2, si t y d son la traza y determinante, se tiene en general y como ejemplo

$$C(X) = X^2 - tX + d, \quad a^2 - ta + d = 0, \quad d = (t-a)\,a, \quad a^{-1} = \frac{1}{d}(t-a),$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = \frac{1}{-2}\left(5\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\right) = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

Otras veces se necesita conocer a^p para muchos valores de p. Inevitablemente ha de calcularse $a^1, a^2, \ldots, a^{n-1}$ pero ya tenemos con el teorema de Cayley-Hamilton que

$$(-1)^n a^n = (-1)^n a^n = -\left(d + \alpha_1 a + \alpha_2 a^2 + \ldots + \alpha_{n-1} a^{n-1}\right),$$

y se puede escribir una fórmula recursiva que nos da cualquier potencia como combinación lineal de las n primeras, incluyendo $a^0 = I$.

[13] Se entiende que el término constante α_0 pasa a ser $\alpha_0 I_n$.

Problema 213 *Calcular una fórmula inductiva para a^k siendo*

$$a = \begin{pmatrix} 0 & h & 0 \\ h & 0 & h \\ 0 & h & 0 \end{pmatrix}, \quad h \in \mathbb{R}.$$

Problema 214 *Sean a y b matrices $n \times n$ relacionadas por $a = c^{-1}bc$ con c invertible ¿Cuál es la relación entre los valores y vectores propios de a y b? Generalizar a endomorfismos L y M de $\mathbb{E}$.*

Este problema da herramientas para ponerse uno mismo problemas de cálculo sabiendo por anticipado la solución. Sea d una matriz diagonal y c invertible. Es muy fácil calcular los vectores propios de d y el polinomio característico $C(X)$. El polinomio característico de $a = cdc^{-1}$ (mejor -1 al final) es también $C(X)$ y si y es vector propio de d, $x = cy$ es vector propio de a. Dado que $\mathcal{E}$ es base que diagonaliza d (puede haber más) los vectores $ce_1, \ldots, ce_n$ forman base que diagonaliza a. Como ce_j es la columna j de c, $(c_1, \ldots, c_n)$ diagonaliza a. Para disponer de c y c^{-1} lo mejor es expresar c como producto de matrices elementales $c = E_1 \cdots E_h$ y como las $(E_j)^{-1}$ son fácilmente calculables, se tiene $c^{-1} = (E_h)^{-1} \cdots (E_1)^{-1}$. Por dar un ejemplo (quizás demasiado) fácil

$$c = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}, \quad c^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -4 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -4 \end{pmatrix},$$

$$d = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}, \quad a = cdc^{-1} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & -4 \end{pmatrix} = \begin{pmatrix} 3 & -4 \\ 0 & 2 \end{pmatrix}.$$

La base $\left((4,1)^\top, (1,0)^\top\right)$ diagonaliza a y, por ejemplo, $(4,1)^\top$ es vector propio con valor propio 2.

Problema 215 *Sea $a \in \Bbbk^{n\times n}$ una matriz triangular superior con todos los términos de la diagonal principal iguales y algún coeficiente no nulo sobre ella. Probar que a no puede ser diagonalizable.*[14]

Problema 216 *En un espacio $\mathbb{E}$ de dimensión n se tienen endomorfismos L y M con valores propios $\{\lambda_1, \ldots, \lambda_n\}$ y $\{\mu_1, \ldots, \mu_n\}$, ambos conjuntos con n elementos. Probar que son equivalentes (a) $L \circ M = M \circ L$ y (b) L y M tienen los mismos vectores propios (pero quizás distintos valores propios).*

Problema 217 *Sea a diagonalizable y $p \in \mathbb{N}$. Demostrar que $b = a^p$ es diagonalizable y determinar sus valores propios. ¿Puede existir a no diagonalizable y p entero positivo tal que b sí sea diagonalizable?*

Problema 218 *Sean $a, b \in \Bbbk^{n\times n}$ con a invertible. Probar que los polinomios característicos de ab y ba son iguales. Más generalmente, si L y M son endomorfismos de $\mathbb{E}$ y L un isomorfismo, los polinomios característicos de $L \circ M$ y $M \circ L$ son el mismo.*

5.6. El teorema de Cayley-Hamilton

Hemos visto que cada matriz a o endomorfismo L tienen un polinomio característico

$$C(X) = c_0 + c_1X + c_2X^2 + \ldots + c_{n-1}X^{n-1} + (-1)^n X^n.$$

Podemos preguntarnos si dado un polinomio de esta forma hay siempre una matriz (y desde luego un endomorfismo en cuanto tenga a a por matriz en una base) tal que el polinomio característico de a sea justamente $C(X)$. La respuesta es que sí. La matriz a que vamos a construir se le llama la **matriz acompañante** de $C(X)$ (algún texto la llama **matriz de Frobenius** también).

Problema 219 *Probar (desarrollando por ¿fila? ¿columna? ¿cuál?) que el polinomio característico de*

$$a = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -c_2 \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 & -c_{n-3} \\ 0 & 0 & 0 & \cdots & 1 & 0 & -c_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -c_{n-1} \end{pmatrix}$$

[14] Quizás ayude empezar con a matriz 4×4.

es $(-1)^n \left[c_0 + c_1X + c_2X^2 + \ldots + c_{n-1}X^{n-1} + X^n\right]$.

Podemos también preguntarnos qué tipo de endomorfismo es el que tiene una matriz con la forma de matriz acompañante. Vamos a responderlo. Fijemos $w \in \mathbb{E}$ no nulo y consideremos la sucesión de vectores

$$L^0(w) = w,\ L^1(w) = L(w),\ L^2(w),\ L^3(w),\ \ldots,\ L^k(w),\ \ldots$$

Como $\mathbb{E}$ es de dimensión finita n hay un *primer* k tal que $L^k(w)$ depende linealmente de los vectores precedentes; o sea,

$$L^k(w) = -c_0L^0(w) - c_1L^1(w) - c_2L^2(w) - \ldots - c_{k-1}L^{k-1}(w). \tag{5.7}$$

Afirmamos que los vectores $L^0(w) = w,\ L^1(w) = L(w),\ L^2(w),\ L^3(w),\ \ldots, L^{k-1}(w)$, en número k, son independientes. Supongamos en efecto que para coeficientes λ_i se tuviera

$$0 = \lambda_0L^0(w) + \lambda_1L^1(w) + \lambda_2L^2(w) + \ldots + \lambda_{k-1}L^{k-1}(w)$$

y sea λ_r el último coeficiente no nulo. En este caso

$$-\frac{\lambda_0}{\lambda_r}L^0(w) - \frac{\lambda_1}{\lambda_r}L^1(w) - \ldots\frac{\lambda_{r-1}}{\lambda_r}L^{r-1}(w) = L^r(x),$$

y al ser $r < k$ tendríamos una contradicción con la definición de k. El espacio

$$\mathbb{Z}(w) = \lg\left(L^0(w), L^1(w), L^2(w), \ldots, L^{k-1}(w)\right)$$

es el **subespacio cíclico generado por** w y se ha visto que $\left(L^0(w), L^1(w), L^2(w), \ldots, L^{k-1}(w)\right)$ es una base. Se tiene

$$\begin{cases} L\left(L^j(w)\right) = L^{j+1}(w) \text{ si } j < k-1. \\ L\left(L^{k-1}(w)\right) = L^k(w) = -c_0L^0(w) - c_1L^1(w) - \ldots - c_{k-1}L^{k-1}(w) \end{cases}$$

Esto prueba que $\mathbb{Z}(w)$ es estable y que la matriz de la restricción M de L a $\mathbb{Z}(w)$ es la matriz acompañante del problema 219 con k en vez de n. Resumiendo: las matrices de acompañantes corresponden a funciones lineales que actúan sobre un espacio tipo $\mathbb{Z}(w)$ con $L^k(w)$ cumpliendo (5.7).

Los espacios cíclicos permiten una demostración muy sencilla del teorema 102 de Cayley-Hamilton.

Teorema 102 *Al sustituir la matriz a en su polinomio característico resulta la matriz* 0. *Más generalmente, al sustituir el endomorfismo L en su polinomio característico. resulta el endomorfismo* 0.

Demostración. Sea $C(X)$ el polinomio característico de L y $w \in \mathbb{E}$ no nulo. Ha de mostrarse que $C(L)(w) = 0$. Consideremos $\mathbb{F} = \mathbb{Z}(w)$ que es estable por L, y sea $M : \mathbb{F} \to \mathbb{F}$ obtenido por restricción de L a $\mathbb{F}$. Si $C_{\mathbb{F}}(X)$ es el polinomio característico de M hemos visto en el teorema 100 que se puede expresar $C(X) = Q(X)C_{\mathbb{F}}(X)$. Afirmamos que $C_{\mathbb{F}}(L)(w) = 0$, lo que bastará para que sea $C(L)(w) = 0$. En la base $\left(L^0(w), L^1(w), L^2(w), \ldots, L^{k-1}(w)\right)$, M tiene matriz a como en el problema 219 (hay que cambiar n por k) y $C_{\mathbb{F}}(X) = (-1)^k\left[c_0 + c_1X + c_2X^2 + \ldots + c_{k-1}X^{k-1} + X^k\right]$. Entonces,

$$\begin{aligned} C_{\mathbb{F}}(L)(w) &= (-1)^k\left[c_0 + c_1L + c_2L^2 + \ldots + c_{k-1}L^{k-1} + L^k\right](w) \\ &= (-1)^k\left[c_0L^0(w) + c_1L^1(w) + c_2L^2(w) + \ldots + c_{k-1}L^{k-1}(w) + L^k(w)\right] \\ &= (-1)^k\left[c_0L^0(w) + c_1L^1(w) + \ldots + c_{k-1}L^{k-1}(w)\right] \\ &\quad - (-1)^k\left[c_0L^0(w) + c_1L^1(w) + \ldots + c_{k-1}L^{k-1}(w)\right] \end{aligned}$$

que sin duda es cero. Esto acaba la demostración del teorema de Cayley-Hamilton. ♣

Problema 220 *Sea* $a \in \Bbbk^{n\times n}$. *Consideramos el subespacio* $\mathbb{S}$ *de* $\Bbbk^{n\times n}$ *generado por* $\left\{a^0, a^1, \ldots, a^k, \ldots\right\}$. *¿Cuál es la máxima dimensión posible de* $\mathbb{S}$? *¿Como se extiende esto si en vez de a tenemos un endomorfismo L de* $\mathbb{E}$ *y* $n = \dim(\mathbb{E})$.

5.7. Diagonalización y fórmulas recursivas

Se puede definir una sucesión en $\mathbb{R}$ recursivamente en la forma $x_n = f(x_{n-1})$, siendo f una función fija y eligiendo x_0 arbitrariamente. El caso más sencillo es $f(x) = px$ y $x_n = px_{n-1} = p^2 x_{n-2} = \ldots = p^n x_0$ (hay más rigor en una demostración inductiva de $x_n = p^{n-1}x_1$). Si tomamos p, q podemos definir $x_n = px_{n-1} + qx_{n-2}$ para $n \geq 2$ y eligiendo x_0, x_1 arbitrariamente. Ahora es más difícil lograr una fórmula "cerrada" del tipo $x_n = g(x_0, x_1, p, q, n)$ que, por una simple sustitución, nos dé x_n para cualquier n. (El ejemplo inicial tenía la solución $x_n = g(x_0, p, n) = p^n x_0$.) Hay a veces ingeniosos procedimientos que permiten conseguirlo. Para tener una referencia concreta ofrecemos al lector la **sucesión de Fibonacci**, caso particular de $x_n = px_{n-1} + qx_{n-2}$ con $p = q = 1$ y $x_0 = 0$, $x_1 = 1$.[15]

Se introduce una variable adicional $y_n = x_{n-1}$ para $n \geq 1$, con lo que la definición inductiva es

$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \begin{pmatrix} px_{n-1} + qx_{n-2} \\ x_{n-1} \end{pmatrix} = \begin{pmatrix} px_{n-1} + qy_{n-1} \\ x_{n-1} \end{pmatrix} = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}.$$

El cálculo de la fórmula cerrada se basa en que

$$\text{si } a = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix}, \text{ entonces } \begin{pmatrix} x_n \\ y_n \end{pmatrix} = a \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix} = a^2 \begin{pmatrix} x_{n-2} \\ y_{n-2} \end{pmatrix} = \ldots = a^{n-1} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}.$$

Hace falta una fórmula cerrada de a^k para $k \in \mathbb{N}$ y para ello diagonalizamos a. Se calcula primero

$$C(X) = \begin{vmatrix} p - X & q \\ 1 & -X \end{vmatrix} = X^2 - pX - q \text{ con raíces } r, s = \frac{p \pm \sqrt{p^2 + 4q}}{2},$$

que supondremos reales y distintas en adelante. Desde luego es lo que pasa si $p = q = 1$. Calculamos ahora una base de vectores propios. Si λ es una de las raíces r o s, calculamos $\ker(a - \lambda)$ para lo que operamos con

$$\begin{pmatrix} p - \lambda & q \\ 1 & -\lambda \end{pmatrix} \to \begin{pmatrix} 1 & -\lambda \\ p - \lambda & q \end{pmatrix} \to \begin{pmatrix} 1 & -\lambda \\ 0 & q + \lambda(p - \lambda) \end{pmatrix} = \begin{pmatrix} 1 & -\lambda \\ 0 & 0 \end{pmatrix}.$$

Se ha usado al final que dado que λ es raíz, $\lambda^2 - p\lambda - q = -(q + \lambda(p - \lambda)) = 0$. Entonces los vectores $(x, y)^\top \in \mathbb{R}^2$ que cumplen $x - \lambda y = 0$ forman $\ker(a - \lambda)$. Por la teoría general, la base $\mathcal{U} = \left((r, 1)^\top, (s, 1)^\top\right)$ diagonaliza a y si $L : \mathbb{R}^2 \to \mathbb{R}^2$ se identifica con a,

$$\begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix} = d = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id})\, \mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(L)\, \mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}) = \begin{pmatrix} r & s \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r & s \\ 1 & 1 \end{pmatrix} = c^{-1}ac.$$

Calculamos a^k por $a^k = \left(cdc^{-1}\right)^k = cd^k c^{-1}$, que detallado nos da

$$a^k = \begin{pmatrix} r & s \\ 1 & 1 \end{pmatrix} \begin{pmatrix} r^k & 0 \\ 0 & s^k \end{pmatrix} \begin{pmatrix} r & s \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{r - s} \begin{pmatrix} r & s \\ 1 & 1 \end{pmatrix} \begin{pmatrix} r^k & 0 \\ 0 & s^k \end{pmatrix} \begin{pmatrix} 1 & -s \\ -1 & r \end{pmatrix},$$

$$a^k = \frac{1}{r - s} \begin{pmatrix} r^{k+1} - s^{k+1} & rs^{k+1} - r^{k+1}s \\ r^k - s^k & rs^k - r^k s \end{pmatrix}.$$

Recordemos que buscamos x_n, que es la *primera componente* de $a^{n-1}(x_1, x_0)^\top$. Obviamente es

$$x_n = \frac{1}{r - s}(r^n - s^n, rs^n - r^n s) \begin{pmatrix} x_1 \\ x_0 \end{pmatrix} = \frac{(r^n - s^n)x_1 + (rs^n - r^n s)x_0}{r - s}, \tag{5.8}$$

que se puede simplificar con $r - s = \sqrt{p^2 + 4q}$ y $rs = q$. Y más aún con la sucesión de Fibonacci,

$$x_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right] = \frac{r^n - s^n}{\sqrt{5}}$$

[15] Con la definición tradicional, los primeros términos son 1, 1, 2, 3, 5, 8, 13, 21, 34,..., luego $x_1 = x_2 = 1$. Aquí introducimos $x_0 = 0$ (fuera de la sucesión) para facilitar los cálculos.

Hay un modo interesante de calcular x_n aplicable a (5.8) basada en que una de las raíces, digamos s, cumpla $|s|<1$. (En la sucesión de Fibonacci, $\frac{1-\sqrt{5}}{2}\approx -0{,}618$.) Se observa que para n "un poco grande", $s^n\approx 0$. Por consiguiente, (5.8) nos da, con la simplificación $rs=q$,

$$x_{n+1}=\frac{\left(r^n-s^n\right)x_1+\left(rs^n-r^ns\right)x_0}{\sqrt{p^2+4q}}\approx\frac{r^nx_1-r^nsx_0}{\sqrt{p^2+4q}}=r^{n-1}\frac{rx_1-qx_0}{\sqrt{p^2+4q}}=\widetilde{x}_n.$$

Desde luego, $\widetilde{x}_n$ no es un entero pero se aproximará a un entero que, necesariamente, debe ser x_n. Por ejemplo,

$$\widetilde{x}_8=\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^8\approx 21{,}010$$

y se puede calcular directamente que $x_8=21$ en la sucesión de Fibonacci.

5.8. Ecuaciones diferenciales matriciales. El caso fácil

5.8.1. El caso real

Sea $x=\left(x^1,\ldots,x^n\right)^\top$ una función de $\mathbb{R}$ en $\mathbb{R}^n$ cuyas componentes son las $x^i:\mathbb{R}\to\mathbb{R}$. Por ejemplo, $x\left(t\right)=\left(\cos t,\sin t\right)^\top$ o $x\left(t\right)=\left(t,\log\left(1+t^2\right),1/\left(1+t^2\right)\right)^\top$. Representamos las derivadas con un punto encima para no interferir con los índices. Por ejemplo, en el primer caso $\dot{x}\left(t\right)=\left(-\sin t,\cos t\right)^\top$ y en el segundo $\dot{x}^3\left(t\right)=-2t/\left(t^2+1\right)^2$. Estudiamos ecuaciones diferenciales del tipo

$$\left\{\begin{array}{l}\dot{x}^1=a_1^1x^1+\ldots+a_n^1x^n\\ \dot{x}^2=a_1^2x^1+\ldots+a_n^2x^n\\ \qquad\vdots\\ \dot{x}^n=a_1^nx^1+\ldots+a_n^nx^n\end{array}\right.\quad\text{o, abreviadamente,}\quad\begin{pmatrix}\dot{x}^1\\ \vdots\\ \dot{x}^n\end{pmatrix}=\begin{pmatrix}a_1^1&\cdots&a_n^1\\ \vdots&\ddots&\vdots\\ a_1^n&\cdots&a_n^n\end{pmatrix}\begin{pmatrix}x^1\\ \vdots\\ x^n\end{pmatrix}.$$

Puede abreviarse todavía más a $\dot{x}=ax$, siendo $a\in\mathbb{R}^{n\times n}$ y $x=\left(x^1,\ldots,x^n\right)^\top$ una función de $\mathbb{R}$ en $\mathbb{R}^n$. Se dice que $\dot{x}=ax$ es una **ecuación diferencial lineal con coeficientes constantes** y x es una **solución de la ecuación** si la verifica. Por ejemplo, $x\left(t\right)=\left(\cos t,\sin t\right)^\top$ es solución de

$$\begin{pmatrix}\dot{x}^1\\ \dot{x}^2\end{pmatrix}=\begin{pmatrix}0&-1\\ 1&0\end{pmatrix}\begin{pmatrix}x^1\\ x^2\end{pmatrix}\text{ porque }\left\{\begin{array}{l}\frac{d}{dt}\cos t=0\cdot\cos t+(-1)\cdot\sin t\\ \frac{d}{dt}\sin t=1\cdot\cos t+0\cdot\cos t\end{array}\right.$$

Nos interesa el caso particular en el que a es diagonalizable; es decir, existe c invertible tal que $c^{-1}ac=d\in\mathbb{R}^{n\times n}$ es diagonal. Aquí es muy fácil dar la solución completa de la ecuación. Abordamos primero el caso particular $\dot{y}=dy$ con d diagonal. Tendremos que resolver n ecuaciones independientes $\dot{y}^h\left(t\right)=\lambda_hy^h\left(t\right)$ y empezamos por $n=1$.

Teorema 103 *Para $\lambda\in\mathbb{R}$, la solución general de la ecuación $\dot{u}\left(t\right)=\lambda u\left(t\right)$ viene dada el conjunto de funciones $u:\mathbb{R}\to\mathbb{R}$ de la forma $u\left(t\right)=ke^{\lambda t}$, con $k\in\mathbb{R}$ constante y $e^\bullet$ la exponencial.*

Demostración. Evidentemente $\frac{d}{dt}\left(ke^{\lambda t}\right)=k\lambda e^{t\lambda}=\lambda\left(ke^{t\lambda}\right)$ y las funciones $u\left(t\right)=ke^{\lambda t}$ son solución de la ecuación. Sea $w:\mathbb{R}\to\mathbb{R}$ una solución de $\dot{u}\left(t\right)=\lambda u\left(t\right)$, quizás de forma distinta. Definimos $v\left(t\right)=w\left(t\right)e^{-\lambda t}$ y

$$\dot{v}\left(t\right)=\dot{w}\left(t\right)e^{-\lambda t}-w\left(t\right)\lambda e^{-\lambda t}=\left[\dot{w}\left(t\right)-\lambda w\left(t\right)\right]e^{-\lambda t}=0\cdot e^{-\lambda t}=0,$$

ya que w es solución. Queda pues $\dot{v}=0$ así que v toma valor constante $k\in\mathbb{R}$. De $w\left(t\right)e^{-\lambda t}=k$ obtenemos $w\left(t\right)=ke^{\lambda t}$ y este es el tipo más general de solución. ♣

Aplicando este teorema a las diversas ecuaciones $\dot{y}^h\left(t\right)=\lambda_hy^h\left(t\right)$ resulta que

$$y\left(t\right)=\left(k^1e^{\lambda_1t},\ldots,k^ne^{\lambda_nt}\right)^\top,\quad k^1,\ldots,k^n\in\mathbb{R},$$

da la totalidad de soluciones de $\dot{y} = dy$ con d la matriz de diagonal $(\lambda_1, \ldots, \lambda_n)$. Tratamos $\dot{x} = ax$, siendo $c^{-1}ac = d$. Afirmamos que definiendo $x = cy$; o si se quiere con diverso detalle

$$\begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}, \text{ o bien } \begin{cases} x^1 = c_1^1 y^1 + \ldots + c_n^1 y^n \\ \vdots\vdots \\ x^n = c_1^n y^1 + \ldots + c_n^n y^n \end{cases}$$

se cumple que x es solución de $\dot{x} = ax$. En efecto, como las c_j^i son constantes es muy fácil probar que de $x = cy$ sale $\dot{x} = c\dot{y}$ y $\dot{x} = c\dot{y} = cdy = acy = ax$. Hasta aquí se tiene que si y es solución de $\dot{y} = dy$, $x = cy$ lo es de $\dot{x} = ax$. Hay una afirmación dual que dice que si x es solución de $\dot{x} = ax$ entonces $y = c^{-1}x$ lo es de $\dot{y} = dy$. La demostración es similar. Queda probado que se pueden transportar soluciones de $\dot{y} = dy$ a soluciones de $\dot{x} = ax$ y viceversa multiplicando por c o c^{-1} y hay correspondencia biyectiva entre los dos espacios de soluciones.[16] Casi hemos probado el siguiente teorema en donde $\mathbb{E}$ es el espacio de las funciones $x : \mathbb{R} \to \mathbb{R}^n$ con componentes infinitamente derivables de $\mathbb{R}$ en $\mathbb{R}$.

Teorema 104 *El conjunto de soluciones de la ecuación $\dot{x} = ax$ con $a \in \mathbb{R}^{n\times n}$ diagonalizable es un subespacio vectorial $\mathbb{S}$ de $\mathbb{E}$, que es de dimensión finita (aunque $\mathbb{E}$ no lo es). Si para c invertible se tiene $c^{-1}ac = d$ y $\left(\lambda^1, \ldots, \lambda^n\right)$ es la diagonal de d, una base de $\mathbb{S}$ está formada por las soluciones*

$$s_1(t) = e^{\lambda_1 t}c_1,\ s_2(t) = e^{\lambda_2 t}c_2, \quad \ldots \quad s_n(t) = e^{\lambda_n t}c_n,$$

siendo $c_1, \ldots, c_n \in \mathbb{R}^n$ las columnas de c.

Demostración. La solución general de $\dot{y} = dy$ se vio que era

$$y(t) = \begin{pmatrix} k_1 e^{\lambda_1 t} \\ \vdots \\ k_n e^{\lambda_n t} \end{pmatrix} = k^1 e^{\lambda_1 t} e_1 + k^2 e^{\lambda_2 t} e_2 + \ldots + k^n e^{\lambda_n t} e_n,$$

siendo $(e_1, \ldots, e_n)$ la base estándar de $\mathbb{R}^n$. También sabemos que cualquier solución x de $\dot{x} = ax$ es de la forma

$$x(t) = cy(t) = c\left(k^1 e^{\lambda_1 t} e_1 \ldots + k^n e^{\lambda_n t} e_n\right) = k^1 e^{\lambda_1 t}\left[ce_1\right] + \ldots + k^n e^{\lambda_n t}\left[ce_n\right] = k^1 e^{\lambda_1 t} c_1 + \ldots + k^n e^{\lambda_n t} c_n,$$

ya que ce_i es c_i, la columna i de c. Así pues, $(s_1(t), \ldots, s_n(t))$ genera $\mathbb{S}$. Veamos la independencia. Si

$$0 = k^1 e^{\lambda_1 t} c_1 + \ldots + k^n e^{\lambda_n t} c_n$$

resulta, por ser c invertible, que $(c_1, \ldots, c_n)$ es independiente, así que $k^1 e^{\lambda_1 t} = \ldots = k^n e^{\lambda_n t} = 0$. Como las exponenciales son > 0, ha de ser $k^1 = \ldots = k^n = 0$. ♣

Problema 221 *Resolver los sistemas $\dot{x} = ax$ siendo respectivamente*

$$a = \begin{pmatrix} 3 & -4 \\ 0 & 2 \end{pmatrix}, \quad a = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix}. \blacklozenge$$

Solución parcial. En realidad hicimos en la página 157 un cálculo para ver que

$$\text{si } c = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} \text{ y } d = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}, \text{ entonces } a = cdc^{-1} = \begin{pmatrix} 3 & -4 \\ 0 & 2 \end{pmatrix}.$$

Se supone que el lector ha tenido que calcular c y d. Las soluciones generales de $\dot{y} = dy$ y $\dot{x} = ax$ son

$$\begin{pmatrix} y^1(t) \\ y^2(t) \end{pmatrix} = \begin{pmatrix} k^1 e^{2t} \\ k^2 e^{3t} \end{pmatrix}, \quad \begin{pmatrix} x^1(t) \\ x^2(t) \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} k^1 e^{2t} \\ k^2 e^{3t} \end{pmatrix} = \begin{pmatrix} 4k^1 e^{2t} + k^2 e^{3t} \\ k^1 e^{2t} \end{pmatrix},$$

[16] Quizás le resulte al lector un lenguaje demasiado elaborado, pero lo que hemos probado es que si $\mathbb{S}_a$ y $\mathbb{S}_d$ son los espacios de soluciones de $\dot{x} = ax$ e $\dot{y} = dy$, las funciones $y \to cy$ y $x \to c^{-1}x$, establecen isomorfismos que son inversos uno de otro entre $\mathbb{S}_a$ y $\mathbb{S}_d$.

aunque quizás prefiera el lector poner la segunda como

$$\begin{pmatrix} x^1(t) \\ x^2(t) \end{pmatrix} = k^1 e^{2t} \begin{pmatrix} 4 \\ 1 \end{pmatrix} + k^2 e^{3t} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

El segundo caso queda para el lector pero le ayudamos un poco diciendo que los valores propios de a son $0, 1, 2$ con vectores propios respectivos $u_1 = (1,1,0)^\top$, $u_2 = (-1,-1,1)^\top$ y $u_3 = (-1,0,1)^\top$. ♦

Frecuentemente las ecuaciones representan la evolución de un fenómeno físico, siendo t el tiempo y $x^i(t)$ el valor de los parámetros que configuran el fenómeno en el instante t. Las a^i_j son constantes más o menos experimentales y $\dot{x} = ax$ es la formulación de la ley física que explica la evolución del fenómeno en forma de ecuación diferencial. Muchísimas leyes físicas son matemáticamente así con diversa complejidad para la ecuación diferencial. Normalmente no interesa tanto la totalidad de las soluciones sino una concreta fijada unívocamente por ciertas condiciones. Es un hecho de gran trascendencia física y matemática que el vector $x(0) = \left(x^1(0), \ldots, x^n(0)\right)^\top$, que representa el **estado inicial del fenómeno**, determina de modo unívoco una solución. Si tal como decimos $x(0)$ determina $x(t)$ para cualquier t, quiere decirse que hay un determinismo extremo, pues si, junto con la ley física (que es a) se conoce el estado del sistema para $t = 0$, conocemos el estado en cualquier instante pasado o futuro, que esto es lo que da $x(t)$. Es matemáticamente muy sencillo ver que $x(0) \in \mathbb{R}^n$ *determina unívocamente la función* $x \in \mathbb{E}$ porque, dado que $e^0 = 1$, si $x(t) = k^1 e^{\lambda_1 t} c_1 + \ldots + k^n e^{\lambda_n t} c_n$, entonces $x(0) = k^1 c_1 + \ldots + k^n c_n$. Queda mejor en forma matricial,

$$\begin{pmatrix} x^1(0) \\ \vdots \\ x^n(0) \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \begin{pmatrix} k^1 \\ \vdots \\ k^n \end{pmatrix} \text{ luego } \begin{pmatrix} k^1 \\ \vdots \\ k^n \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix}^{-1} \begin{pmatrix} x^1(0) \\ \vdots \\ x^n(0) \end{pmatrix}.$$

En el caso primero del problema 221 la solución que para $t = 0$ toma los valores $(4,-2)^\top$ es la que tiene $\left(k^1, k^2\right)^\top = (4,-2)^\top$ verificando

$$\begin{pmatrix} 4 \\ -2 \end{pmatrix} = k^1 \begin{pmatrix} 4 \\ 1 \end{pmatrix} + k^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} k^1 \\ k^2 \end{pmatrix},$$

y por tanto

$$\begin{pmatrix} k^1 \\ k^2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ -2 \end{pmatrix} = \begin{pmatrix} -2 \\ 12 \end{pmatrix}.$$

5.8.2. El caso complejo

Si volvemos al ejemplo inicial de la sección con solución $x(t) = (\cos t, \sin t)^\top$, observamos que no queda cubierto por el trabajo realizado ya que la correspondiente matriz tiene $C(X) = X^2 + 1$ sin raíces reales. Sin embargo se pueden usar casi todas las ideas del trabajo precedente con tal de sustituir $C^\infty(\mathbb{R},\mathbb{R})$, el **espacio de las funciones infinitamente derivables de $\mathbb{R}$ en $\mathbb{R}$** por $C^\infty(\mathbb{R},\mathbb{C})$, el **espacio de las funciones infinitamente derivables de $\mathbb{R}$ en $\mathbb{C}$**. Allí vimos que si $\mathbb{E}$ era $C^\infty(\mathbb{R},\mathbb{R})^n$ los $x \in \mathbb{E}$ solución de $ax = \dot{x}$ formaban un subespacio $\mathbb{S}$ de $\mathbb{E}$ de dimensión finita. Ahora pretendemos probar que si $a \in \mathbb{C}^{n\times n}$ tiene sentido $ax = \dot{x}$ con tal de que conozcamos $C^\infty(\mathbb{R},\mathbb{C})$ y asumamos que una solución es un elemento de $\mathbb{E} = C^\infty(\mathbb{R},\mathbb{C})^n$. Hay que aclarar que una función $f : \mathbb{R} \to \mathbb{C}$ se puede escribir como $f(t) = u(t) + iv(t)$ con $u, v : \mathbb{R} \to \mathbb{R}$ y que si u y v son derivables, la derivada de f es $f'(t) = u'(t) + iv'(t)$ por definición. Según esto, $C^\infty(\mathbb{R},\mathbb{C})$ es el espacio de aquellas $f : \mathbb{R} \to \mathbb{C}$ tales que u y v están en $C^\infty(\mathbb{R},\mathbb{R})$.

Hemos dado por conocida la función exponencial $t \to e^t$ en $C^\infty(\mathbb{R},\mathbb{R})$. Se puede generalizar a otra **exponencial compleja** de $\mathbb{C}$ en $\mathbb{C}$ que definimos. Si $z = u + iv \in \mathbb{C}$ con $u, v \in \mathbb{R}$, se define $e^z = e^u(\cos v + i \sin v) \in \mathbb{C}$. Advertimos al lector que, como queremos trabajar en $\mathbb{E} = C^\infty(\mathbb{R},\mathbb{C})$, nos van a interesar las funciones de $\mathbb{R}$ en $\mathbb{C}$ (y no de $\mathbb{C}$ en $\mathbb{C}$) construidas con la exponencial de la forma

$$E : \mathbb{R} \to \mathbb{C}, \quad E(t) = e^{t\alpha} = e^{t(u+iv)} = e^{tu}(\cos tv + i \sin tv). \tag{5.9}$$

El lector puede admitir, aunque tienen comprobación sencilla pero pesada, los resultados siguientes que son análogos a lo que pasa en $C^\infty(\mathbb{R},\mathbb{R})$.

1. La suma, producto y producto por $\lambda \in \mathbb{C}$ de funciones de $C^\infty(\mathbb{R},\mathbb{C})$ están en $C^\infty(\mathbb{R},\mathbb{C})$ pues las derivadas se calculan con $(f+g)'(t) = f'(t)+g'(t)$, $(fg)'(t) = f'(t)g(t)+f(t)g'(t)$, $(\lambda f)'(t) = \lambda f'(t)$. Como consecuencia, $\mathbb{E} = C^\infty(\mathbb{R},\mathbb{C})$ es un espacio vectorial *complejo*. Las funciones polinómicas $P(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_n t^n$ con los $\alpha_j \in \mathbb{C}$ están en $C^\infty(\mathbb{R},\mathbb{C})$ y este espacio es de dimensión infinita.

2. La condición $f' = 0$ equivale a que f es constante.

3. La exponencial compleja verifica $e^{z+w} = e^z e^w$, $e^0 = 1$, y $(e^z)^{-1} = e^{-z}$.

4. Si E viene dada por (5.9) se tiene que $E'(t) = \frac{d}{dt}(e^{t\alpha}) = \alpha e^{t\alpha} = \alpha E(t)$.

Para $a \in \mathbb{C}^{n\times n}$ podemos considerar la **ecuación diferencial lineal con coeficientes constantes** $\dot{x} = ax$, siendo una **solución** una función $x : \mathbb{R} \to \mathbb{C}^n$ con n componentes $x^i : \mathbb{R} \to \mathbb{C}$ que verifique la ecuación. Nos interesa el caso particular en el que a es diagonalizable; es decir, existe $c \in \mathbb{C}^{n\times n}$ invertible tal que $c^{-1}ac = d \in \mathbb{C}^{n\times n}$ es diagonal. En tal caso es muy fácil dar la solución completa de la ecuación. Abordemos primero el caso particular $\dot{y} = dy$ con d diagonal. En este caso, hay que resolver n ecuaciones independientes $\dot{y}^h(t) = \lambda_h y^h(t)$ variando solo la situación respecto al caso real ya estudiado en que los λ_h son complejos. Sin embargo, no cuesta nada conjeturar y verificar el siguiente teorema calcado del teorema 103.

Teorema 105 *Para $\lambda \in \mathbb{C}$, la solución general de la ecuación $\dot{u}(t) = \lambda u(t)$ viene dada el conjunto de funciones $u : \mathbb{R} \to \mathbb{C}$ de la forma $u(t) = ke^{\lambda t}$, con $k \in \mathbb{C}$ constante.*

Aplicando este teorema a las diversas ecuaciones $\dot{y}^h(t) = \lambda_h y^h(t)$ resulta que

$$y(t) = \left(k^1 e^{\lambda_1 t}, \ldots, k^n e^{\lambda_n t}\right)^\top, \quad k^1, \ldots, k^n \in \mathbb{R}$$

da la totalidad de soluciones de $\dot{y} = dy$ con $(\lambda_1, \ldots, \lambda_n)$ en la diagonal de la matriz diagonal d. Si nuestra matriz a es diagonalizable, $c^{-1}ac = d$ con c invertible, afirmamos que definiendo $x = cy$ se cumple que x es solución de $\dot{x} = ax$. El pequeño cálculo es idéntico al que se hizo para el caso real. Se tiene también que si x es solución de $\dot{x} = ax$ entonces $y = c^{-1}x$ lo es de $\dot{y} = dy$. Queda probado que se pueden transportar soluciones de $\dot{y} = dy$ a soluciones de $\dot{x} = ax$ y viceversa multiplicando por c o c^{-1} y hay correspondencia biyectiva entre los dos espacios de soluciones. Se tiene también un análogo del teorema 104 donde ahora $\mathbb{E} = C^\infty(\mathbb{R},\mathbb{C})^n$.

Teorema 106 *El conjunto de soluciones de la ecuación $\dot{x} = ax$ con $a \in \mathbb{C}^{n\times n}$ diagonalizable es un subespacio vectorial $\mathbb{S}$ de $\mathbb{E}$. Este subespacio $\mathbb{S}$ es de dimensión finita (aunque $\mathbb{E}$ no lo es). Si para c invertible se tiene $c^{-1}ac = d$ con d diagonal y $\left(\lambda^1, \ldots, \lambda^n\right)$ es la diagonal de d, una base de $\mathbb{S}$ está formada por las soluciones*

$$s_1(t) = e^{\lambda_1 t}c_1,\ s_2(t) = e^{\lambda_2 t}c_2, \quad \ldots \quad s_n(t) = e^{\lambda_n t}c_n,$$

siendo $c_1, \ldots, c_n \in \mathbb{C}^n$ las columnas de c.

Como en el caso real, la condición inicial $x(0) \in \mathbb{C}^n$ *determina unívocamente la función* $x \in \mathbb{E} = C^\infty(\mathbb{R},\mathbb{C})^n$ porque dado que $e^0 = 1$, si $x(t) = k^1 e^{\lambda_1 t}c_1 + \ldots + k^n e^{\lambda_n t}c_n$, entonces $x(0) = k^1 c_1 + \ldots + k^n c_n$. Queda mejor en forma matricial,

$$\begin{pmatrix} x^1(0) \\ \vdots \\ x^n(0) \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \begin{pmatrix} k^1 \\ \vdots \\ k^n \end{pmatrix} \text{ luego } \begin{pmatrix} k^1 \\ \vdots \\ k^n \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix}^{-1} \begin{pmatrix} x^1(0) \\ \vdots \\ x^n(0) \end{pmatrix}.$$

Problema 222 *Determinar el espacio de soluciones de $ax = \dot{x}$ siendo*

$$a = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

y la solución concreta tal que $x(0) = (1,0)^\top$. ♦

Solución. Los valores propios de a son $+i$ y $-i$ con vectores propios $(i,1)^\top$ y $(-i,1)$. Se tiene pues

$$\begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = d.$$

Entonces $e^{it} = \cos t + i \sin t$ y $e^{-it} = \cos(-t) + i \sin(-t) = \cos t - i \sin t$. La solución general es toda función lineal que sea combinación *compleja* con coeficientes k^1 y k^2 de

$$(\cos t + i \sin t) \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} i \cos t - \sin t \\ \cos t + i \sin t \end{pmatrix} \quad \text{y} \quad (\cos t - i \sin t) \begin{pmatrix} -i \\ 1 \end{pmatrix} = \begin{pmatrix} -i \cos t - \sin t \\ \cos t - i \sin t \end{pmatrix}.$$

Los coeficientes k^i para que sea $x(0) = (1,0)^\top$ deben verificar

$$\begin{pmatrix} x^1(0) \\ x^2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \begin{pmatrix} k^1 \\ k^2 \end{pmatrix} \text{ y } \begin{pmatrix} k^1 \\ k^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -i & 1 \\ i & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}i \\ \frac{1}{2}i \end{pmatrix}.$$

La solución buscada es (¡oh, maravilla!)

$$x(t) = -\frac{1}{2}i \begin{pmatrix} i \cos t - \sin t \\ \cos t + i \sin t \end{pmatrix} + \frac{1}{2}i \begin{pmatrix} -i \cos t - \sin t \\ \cos t - i \sin t \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}. \blacklozenge$$

Por supuesto, es mucho más fácil adivinar a ojo que $x(t) = (\cos t, \sin t)^\top$ es la solución requerida.

5.9. El polinomio minimal

Si $a \in \Bbbk^{n\times n}$, las matrices $a^0 = I, a^1 = a, a^2, \ldots . a^{n^2}$ (los exponentes indican potencias, no fila) de $\Bbbk^{n\times n}$ son en número n^2+1. Al ser $\dim(\Bbbk^{n\times n}) = n^2$, existen $p_0, p_1, \ldots, p_{n^2}$ en $\Bbbk$ tales que

$$p_0 I + p_1 a + p_2 a^2 + \ldots + p_{n^2} a^{n^2} = 0, \text{ equivalente a } P(a) = 0 \text{ si } P(X) = p_0 + p_1 X + p_2 X^2 + \ldots + p_{n^2} X^{n^2}.$$

Hay pues polinomios $P(X)$ tales que $P(a) = 0$. Recordamos el teorema 102 (de Cayley-Hamilton) según el cual el polinomio característico $C(X)$ de a cumple $C(a) = 0$. Como consecuencia hay polinomios de grado $\leq n$ tales que $P(a) = 0$. Esto permite definir el **polinomio minimal** (de a)[17] como el polinomio $M(X)$ mónico y de mínimo grado tal que $M(a) = 0$. Puede ser $\deg(C(X)) > \deg(M(X))$.

Teorema 107 *Los polinomios $P(X)$ que cumplen $P(a) = 0$ verifican respecto a $M(X)$ que*

1. *Son múltiplos de $M(X)$.*
2. *Tienen entre sus raíces a los valores propios de a.*
3. *$M(X)$ es único y tiene las mismas raíces que $C(X)$; o sea, los valores propios de a.*

Demostración. Se escribe $P(X) = M(X)Q(X) + R(X)$ con $R(X) = 0$ o $\deg R(X) < \deg M(X)$. Lo segundo no puede darse porque $0 = P(a) = M(a)Q(a) + R(a) = R(a)$, que contradice la definición de $M(X)$. Queda probado **1**. Para probar **2** se comprueba primero fácilmente que si x es vector propio de λ se cumple que $P(a)(x) = P(\lambda)x$ sea como sea $P(X)$. Si $P(a) = 0$ queda $P(\lambda)x = 0$, y al ser $x \neq 0$, es $P(\lambda) = 0$.

Si $M_1(X)$ y $M_2(X)$ cumplen la definición de polinomio minimal se tendrá por **1** que $M_1(X) = M_2(X)Q_2(X)$ y $M_2(X) = M_1(X)Q_1(X)$. Como deben tener $M_1(X)$ y $M_2(X)$ el mismo grado se sigue que $Q_2(X)$ es de grado 0. Entonces, como $M_1(X)$ y $M_2(X)$ son mónicos, llegamos a que $Q_2(X) = 1$ y $M_1(X) = M_2(X)$. Acabamos **3**. Por el teorema de Cayley-Hamilton, $C(a) = 0$ y **1** da que $M(X)$ divide a $C(X)$, luego las raíces de $M(X)$ lo son de $C(X)$. Por otra parte, $M(a) = 0$ y **2** implican que las raíces de $M(X)$ deben incluir a los valores propios, que son las raíces de $C(X)$. ♣

[17] Es mucho más frecuente decir "polinomio mínimo". Si un conjunto está ordenado, un elemento mínimo es el que es menor que cualquier otro y un elemento minimal es aquel para el que no se encuentran otros estrictamente menores (sutil diferencia). Lo mínimo es minimal, pero no al revés. En inglés se dice *minimal polinomial* y su dominio nos hace elegir *polinomio minimal*. ¿Es euclídeo o euclidiano, polinomial o polinómico, hermítico o hermitiano? Al menos en los dos primeros casos la RAE da como correcta la segunda opción. Reconozcámoslo.

Si tenemos L endomorfismo de $\mathbb{E}$, se puede ampliar esto a L y el **polinomio minimal de** L será el que es mónico y de grado mínimo verificando $M(L)=0$. Se ve enseguida que si a es la matriz de L en una base $\mathcal{U}$, los polinomios mínimo y característico de L y los valores propios de L y a son los mismos. Por ejemplo, si $\lambda\in\Bbbk$ es valor propio de L con x como vector propio asociado, el vector $\mathrm{mat}^{\mathcal{U}}(x)=\left(x^1,\ldots,x^n\right)^\top$ es vector propio de a con valor propio λ y viceversa.

El polinomio minimal tiene gran interés teórico pero no tiene una fórmula definida de cálculo como el característico. Si tenemos $C(X)$ linealmente factorizado como $(X-\lambda_1)^{m_1}\cdots(X-\lambda_p)^{m_p}$, sabemos por el teorema 107 que $M(X)$ es de la forma $(X-\lambda_1)^{s_1}\cdots(X-\lambda_p)^{s_p}$ con cada $s_j\leq m_j$, ya que $M(X)$ divide a $C(X)$, y $s_j\geq 1$ ya que cada raíz de $C(X)$ lo es de $M(X)$. El procedimiento para calcular $M(X)$ en esta situación es un puro tanteo. Se calcula $(a-\lambda_1)\cdots(a-\lambda_p)=b$; si es cero, $M(X)=(X-\lambda_1)\cdots(X-\lambda_p)$ con todos los factores lineales con exponente 1. Si este producto es no nulo, se multiplica b por nuevas matrices $a-\lambda_j$ hasta llegar al primer producto no nulo. Sustituyendo a por X obtenemos el polinomio minimal. Se comprende que el trabajo puede ser grande en cuanto se complique $C(X)$. Si $C(X)$ está factorizado pero no con factores lineales, vale el mismo tanteo.

Problema 223 *Calcular el polinomio minimal de las matrices con* $\Bbbk=\mathbb{R}$,

$$a=\begin{pmatrix}2&0&-1&-1\\1&0&-1&-1\\0&1&1&0\\0&0&0&1\end{pmatrix},\quad b=\begin{pmatrix}2&0&0\\1&0&-1\\0&1&2\end{pmatrix}.\ \blacklozenge$$

Solución. En el primer caso, $C(X)=(X-1)^4$. Entonces se comprueba que $(a-1)$ y $(a-1)^2$ son no nulas pero sí lo es $(a-1)^3$, luego $M(X)=(X-1)^3\neq C(X)$.

En el segundo caso, $C(X)=(X-2)(X-1)^2$ y se comprueba que $(a-1)(a-2)\neq 0$ luego necesariamente será $M(X)=C(X)$. $\blacklozenge$

Problema 224 *Ídem para las matrices (¡prácticamente idénticas!)*

$$a=\begin{pmatrix}1&0&0&1\\0&1&0&0\\0&0&1&0\\1&0&0&1\end{pmatrix},\quad b=\begin{pmatrix}1&0&0&-1\\0&1&0&0\\0&0&1&0\\1&0&0&1\end{pmatrix}$$

Distinguir el caso $\Bbbk=\mathbb{R}$ *del caso* $\Bbbk=\mathbb{C}$. *Ayudamos con* $C_b(X)=\left(X^2-2X+2\right)(X-1)^2$.

El lector puede ponerse todos los problemas de cálculo numérico que quiera, basado en que matrices semejantes tienen el mismo polinomio minimal. La receta es esta. Se toma una matriz invertible p cuya inversa sea fácil de calcular; por ejemplo,

$$p=\begin{pmatrix}1&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix},\quad p^{-1}=\begin{pmatrix}1&-1&-1&-1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

y una matriz t cuyos polinomios característico y minimal sean fáciles también de calcular, siendo lo mejor que t sea triangular. Se define $a=ptp^{-1}$ y ya se sabe que $M_a(X)=M_t(X)$. La matriz a de un problema de más arriba se ha obtenido como

$$\begin{pmatrix}1&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}1&0&0&0\\1&1&0&0\\0&1&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}1&1&1&1\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}^{-1}=\begin{pmatrix}2&0&-1&-1\\1&0&-1&-1\\0&1&1&0\\0&0&0&1\end{pmatrix}.$$

5.10. Una segunda visita a la diagonalización

Conocer el polinomio minimal permite conocer si un endomorfismo es diagonalizable o no con un criterio distinto a los del teorema 101 y construir bases de vectores propios. Es lo que mostramos en esta sección.

Teorema 108 *Sea $L:\mathbb{E}\to\mathbb{E}$ un endomorfismo cuyos polinomios característico y minimal sean linealmente factorizables (si lo es el uno, lo es el otro). Para que L sea diagonalizable, es necesario y suficiente que sea $M(X)=(X-\lambda_1)\cdots(X-\lambda_p)$; o sea, que todos los exponentes de los factores sean* 1.

Demostración. Sea L diagonalizable y $\mathbb{E}=\mathbb{E}(\lambda_1)\oplus\ldots\oplus\mathbb{E}(\lambda_p)$ la suma directa de los subespacios propios (teorema 101). Bastará ver que el polinomio $P(X)=(X-\lambda_1)\cdots(X-\lambda_p)$ cumple $P(L)(x)=0$ para todo $x\in\mathbb{E}$ pues entonces, $P(X)=M(X)$. Justificación: $M(X)$ dividirá a $P(X)$ (**1** en el teorema 107) y al tener $M(X)$ las mismas raíces que $P(X)$ (**2** en ese teorema) ha de ser $M(X)=P(X)$. Para ver que $P(L)(x)=0$ basta que sea $x=x_j\in\mathbb{F}_j$. Permutando a conveniencia los factores de $P(X)$,

$$\begin{aligned}P(L)(x_j)&=(L-\lambda_1)\circ\cdots\circ\widehat{(L-\lambda_j)}\circ\cdots\circ(L-\lambda_p)\circ(L-\lambda_j)(x_j)\\&=(L-\lambda_1)\circ\cdots\circ\widehat{(L-\lambda_j)}\circ\cdots\circ(L-\lambda_p)(0)=0.\end{aligned}$$

Sea ahora $M(X)=(X-\lambda_1)\cdots(X-\lambda_p)$. Tenemos la identidad

$$\begin{aligned}1=&\frac{\widehat{(X-\lambda_1)}(X-\lambda_2)\cdots(X-\lambda_p)}{\widehat{(\lambda_1-\lambda_1)}(\lambda_1-\lambda_2)\cdots(\lambda_1-\lambda_p)}+\ldots\\&+\frac{(X-\lambda_1)\cdots\widehat{(X-\lambda_j)}\cdots(X-\lambda_p)}{(\lambda_j-\lambda_1)\cdots\widehat{(\lambda_j-\lambda_j)}\cdots(\lambda_j-\lambda_p)}+\ldots+\frac{(X-\lambda_1)\cdots(X-\lambda_{p-1})\widehat{(X-\lambda_p)}}{(\lambda_p-\lambda_1)\cdots(\lambda_p-\lambda_{p-1})\widehat{(\lambda_p-\lambda_p)}}\end{aligned}$$

donde el tejadillo indica que se quita lo que está bajo él. La forma de probar la identidad es curiosa. El lado derecho $P(X)$ es un polinomio de grado $p-1$ y $Q(X)=P(X)-1$ es un polinomio de grado $p-1$ que se anula en n valores *distintos* $\lambda_1,\ldots,\lambda_p$. Tiene que ser $Q(X)=0$, que es la identidad anunciada.

Con esta identidad $1=P(X)$, se tiene $x=P(L)(x)$ para todo $x\in\mathbb{E}$. Con más detalle,

$$x=x_1+\ldots+x_p,\text{ siendo }x_j=\frac{(L-\lambda_1)\circ\cdots\circ\widehat{(L-\lambda_j)}\circ\cdots\circ(L-\lambda_p)}{(\lambda_j-\lambda_1)\cdots\widehat{(\lambda_j-\lambda_j)}\cdots(\lambda_j-\lambda_p)}(x).$$

Obsérvese que x_j es vector propio con valor propio λ_j porque

$$\begin{aligned}(L-\lambda_j)(x)&=\mu_j(L-\lambda_j)\circ(L-\lambda_1)\circ\cdots\circ\widehat{(L-\lambda_j)}\circ\cdots\circ(L-\lambda_p)(x)\\&=\mu_j(L-\lambda_1)\circ\cdots\circ(L-\lambda_j)\circ\cdots\circ(L-\lambda_p)(x)=\mu_jM(L)(x)=0,\end{aligned}$$

siendo $\mu_j=\left[(\lambda_j-\lambda_1)\cdots\widehat{(\lambda_j-\lambda_j)}\cdots(\lambda_j-\lambda_p)\right]^{-1}$. Hemos probado hasta aquí que cada x es suma de vectores propios, $x=x_1+\ldots+x_p$ con x_j vector propio de λ_j. Así pues, $\mathbb{E}=\mathbb{E}(\lambda_1)+\ldots+\mathbb{E}(\lambda_p)$ pero, al ser los λ_j distintos, la suma es directa (teorema 100) y L es diagonalizable. ♣

Este teorema 108 y su naturaleza constructiva permiten abordar de nuevo la diagonalización. Dada a y conocidos los valores propios $\lambda_1,\ldots,\lambda_p$, se calcula $(a-\lambda_1)\cdots(a-\lambda_p)=b$ y si $b=0$ esto es condición necesaria y suficiente para su diagonalización. Los subespacios y bases de vectores propios se pueden calcular porque $\mathbb{E}(\lambda_j)$ es la imagen de $(a-\lambda_1)\cdots\widehat{(a-\lambda_j)}\cdots(a-\lambda_p)=b_j$; es decir, el espacio de columnas de b_j y un conjunto máximo de columnas independientes de b_j es una base de $\mathbb{E}(\lambda_j)$. Yuxtaponiendo estas bases de los $\mathbb{E}(\lambda_j)$ tenemos una base de $\mathbb{E}$ de vectores propios. Volvemos con este método a estudiar algunos problemas antes vistos en este capítulo.

Problema 225 *Supongamos que L tiene $C(X)=(\lambda_1-X)\cdots(\lambda_n-X)$ con n raíces distintas siendo $n=\dim(\mathbb{E})$. Probar que L es diagonalizable. (Es el antiguo problema 205).* ♦

Solución. En cualquier caso $M(X)$ debe dividir a $C(X)$ pero como todos los $(X-\lambda_j)$ en $C(X)$ tienen exponente 1, es $C(X)=M(X)$ y el teorema 108 da que L es diagonalizable. ♦

Problema 226 *No es diagonalizable $L:\mathbb{R}_4[X]\to\mathbb{R}_4[X]$, $L(P(X))=P'(X)+P''(X)$. ¿Por qué?* ♦

Solución. Su matriz en la base estándar $\left(1,X,\ldots,X^4\right)$ es

$$a=\begin{pmatrix}0&1&2&0&0\\0&0&2&6&0\\0&0&0&3&12\\0&0&0&0&4\\0&0&0&0&0\end{pmatrix}.$$

Obviamente $C(X) = -X^5$ pero como es estrictamente triangular, $a^3 = 0$ luego $M(X) = X^3$. Debería ser $M(X) = X$ para que fuese posible diagonalizar. ♦

Problema 227 *Probar que $L : \mathbb{E} \to \mathbb{E}$ que en una base $\mathcal{U} = (u_1, u_2)$ tiene matriz*

$$a = \begin{pmatrix} -2 & -2 \\ 10 & 7 \end{pmatrix}$$

es diagonalizable y dar una base de vectores propios. ♦

Solución. Se tiene $C(X) = (2 - X)(3 - X)$ y esto implica $M(X) = C(X)$. Además

$$a - 2 = \begin{pmatrix} -4 & -2 \\ 10 & 5 \end{pmatrix}, \quad a - 3 = \begin{pmatrix} -5 & -2 \\ 10 & 4 \end{pmatrix}$$

y $\mathbb{E}(2) = \operatorname{im}(a - 3)$ y $\mathbb{E}(3) = \operatorname{im}(a - 2)$. Como base de $\mathbb{E}(2)$ vale cualquiera de las columnas de $a - 3$ y como base de $\mathbb{E}(3)$ cualquiera de las columnas de $a - 2$. Al tratar este problema hace algún tiempo se tomó la base $\left((-1, 2)^\top, (-2, 5)^\top\right)$. ♦

Problema 228 *Dar una base de vectores propios, cuando sea diagonalizable, de*

$$a = \begin{pmatrix} \alpha & 1 & -1 \\ 1 & \alpha & 1 \\ -1 & 1 & \alpha \end{pmatrix}.$$

Para no alargar el cálculo decimos que $C(X) = (Y - 2)(Y + 1)^2$, siendo $Y = \alpha - X$.

Problema 229 *Dar una base de vectores propios, suponiendo $\Bbbk = \mathbb{C}$, $\beta \neq 0$ para*

$$a = \begin{pmatrix} 1 & 0 & 0 & \beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \alpha \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

cuando sea posible.

Problema 230 *Calcular una base de vectores propios para la matriz (ya apareció en el problema 211),*

$$a = \begin{pmatrix} \alpha & 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & 1 & 1 & \cdots & 1 \\ 1 & 1 & \alpha & 1 & \cdots & 1 \\ 1 & 1 & 1 & \alpha & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & \alpha \end{pmatrix}.$$

Ayudamos diciendo que $C(X) = ((\alpha - 1) - X)^{n-1}((\alpha - 1 + n) - X)$.

Hemos visto antes que las matrices simétricas y hermitianas tenían valores propios "especiales", en el sentido de ser reales, y que por ello el polinomio característico era linealmente factorizable. El polinomio minimal de estas matrices, que debe ser linealmente factorizable al dividir al característico, es también "especial" porque vamos a probar que es de la forma $M(X) = (X - \lambda_1) \cdots (X - \lambda_p)$; o sea, que todos los exponentes de los factores sean 1. Por aplicación del teorema 108 deduciremos que las matrices simétricas o hermitianas son diagonalizables. La clave para ver que los factores de $M(X)$ tienen exponente 1 es este teorema.

Teorema 109 *Sea $a \in \mathbb{C}^{nxn}$ hermitiana y $x \in \mathbb{C}^n$. Si para $k \geq 2$ se tiene $a^k x = 0$, entonces $ax = 0$.*

Demostración. Como $2(k - 1) \geq k$ debe ser $a^{k-1}a^{k-1}x = 0$. Entonces, por ser $a = a^*$,

$$0 = x^*\left(a^{k-1}a^{k-1}x\right) = x^*\left(\left(a^{k-1}\right)^* a^{k-1}x\right) = \left(a^{k-1}x\right)^* a^{k-1}x.$$

Vimos que para $y \in \mathbb{C}^n$, $y^*y = 0$ implica $y = 0$. En particular, $a^{k-1}x = 0$ y repitiendo el argumento $a^{k-1}x, \ldots, ax$ son cero. ♣

Teorema 110 *Los polinomios minimales de las matrices simétricas y hermitianas son del tipo* $M(X) = (X-\lambda_1)\cdots(X-\lambda_p)$. *Todas estas matrices son diagonalizables.*

Demostración. Sea $M(X) = (X-\lambda_1)^{m_1}\cdots(X-\lambda_p)^{m_p}$ con algún $m_i \geq 2$. Llegaremos a contradicción. Supongamos por simplicidad $m_1 \geq 2$. Tomemos $y \in \mathbb{C}^n$ arbitrario y $x = (a-\lambda_1)^{m_2}\cdots(a-\lambda_p)^{m_p} y$. Por definición de polinomio minimal, $(a-\lambda_1)^{m_1} x = 0$. Pero $a-\lambda_1$ es hermitiana, así que el teorema precedente nos da $(a-\lambda_1)x = 0$ y $P(X) = (X-\lambda_1)(X-\lambda_2)^{m_2}\cdots(X-\lambda_p)^{m_p}$ de grado inferior a $M(X)$ anula a todo vector de $\mathbb{C}^n$, contradiciendo que $M(X)$ es el polinomio minimal.

La última afirmación se sigue del teorema 108. ♣

Quizás haya observado el lector que hemos hablado de *matrices* simétricas y hermitianas y no de *endomorfismos* simétricos o hermitianos; de hecho no se han definido. Es conveniente hacerlo, pero será al tratar los espacios euclidianos y unitarios, y de momento nos contentamos con esta exposición restringida.

5.11. Triangulación

Hemos visto qué significa que la base $\mathcal{U} = (u_1, \ldots, u_n)$ diagonalice el endomorfismo L, pudiéndose enunciar de varias formas, siendo una de ellas que la matriz de L en $\mathcal{U}$ sea diagonal. Diremos que $\mathcal{U}$ **triangula** L si la matriz en $\mathcal{U}$ es triangular (superior o inferior). La traducción con símbolos es

$$\begin{cases} L(u_1) = a_1^1 u_1 \\ L(u_2) = a_2^1 u_2 + a_2^2 u_2 \\ \vdots \\ L(u_j) = a_j^1 u_2 + a_j^2 u_2 + \ldots + a_j^j u_j \\ \vdots \\ L(u_n) = a_n^1 u_1 + a_n^2 u_2 + \ldots + a_n^n u_n \end{cases} \qquad \begin{cases} L(u_1) = a_1^1 u_1 + a_1^2 u_2 + \ldots + a_1^n u_n \\ L(u_2) = a_2^2 u_2 + a_2^3 u_3 + \ldots + a_1^n u_n \\ \vdots \\ L(u_j) = a_j^j u_2 + a_j^{j+1} u_{j+1} + \ldots + a_j^n u_n \\ \vdots \\ L(u_n) = a_n^n u_n \end{cases}$$

o más brevemente, $L(u_j) = \sum_{1\leq i\leq j} a_j^i u_i$ y $L(u_j) = \sum_{j\leq i\leq n} a_j^i u_i$ para $j = 1, 2, \ldots, n$. Antes de seguir le informamos al lector que si se sustituye la base $\mathcal{U} = (u_1, u_2, \ldots, u_n)$ por $\mathcal{V} = (u_n, u_{n-1}, \ldots, u_1)$ y la matriz de L en $\mathcal{U}$ es triangular superior, la de L en $\mathcal{V}$ es triangular inferior y viceversa.[18] Esto implica que si se quiere probar que L es triangulable para $\mathcal{U}$, basta hacerlo para (digamos) el caso "superior", pues con muy poco esfuerzo resulta el caso "inferior". Para adaptarnos al desarrollo de la forma de Jordan vamos a tratar la triangulación inferior.

Teorema 111 *Para* $L : \mathbb{E} \to \mathbb{E}$ *o* $a \in \Bbbk^{n\times n}$ *son equivalentes* ***(a)*** L *o* a *son triangulables, y* ***(b)*** *sus polinomios* $C(X)$ *y* $M(X)$ *(característico y minimal) son linealmente factorizables.*

Demostración. Como $M(X)$ divide a $C(X)$ y ambos tienen las mismas raíces, basta que uno sea linealmente factorizable para que lo sea el otro. Si L es triangulable tiene matriz t triangular en la base $\mathcal{U}$. Como $C(X)$ no depende de la base, $C(X) = \left(t_1^1 - X\right)\left(t_2^2 - X\right)\cdots(t_n^n - X)$ y **(a)** implica **(b)**.

Probamos que **(b)** implica **(a)** por inducción sobre n, siendo la implicación verdadera si $n = 1$. Supongamos $n > 1$ y cierto el teorema para todo endomorfismo M de un espacio $\mathbb{F}$ con $\dim(\mathbb{F}) < n$. Si $C(X) = \pm(X-\lambda_1)^{m_1}\cdots(X-\lambda_p)^{m_p}$, no todos los subespacios $(L-\lambda_i)(\mathbb{E})$ pueden ser iguales a $\mathbb{E}$, porque sería $C(L)(\mathbb{E}) = \mathbb{E}$, pero el teorema 102 (de Cayley-Hamilton) nos dice que $C(L) = 0$, luego $\mathbb{E} = 0$, incompatible con $n > 1$. Sea $(L-\lambda_i)(\mathbb{E}) = \mathbb{F}$ uno de estos espacios. Es estable por L ya que

$$L(\mathbb{F}) = (L\circ(L-\lambda_i))(\mathbb{E}) = ((L-\lambda_i)\circ L)(\mathbb{E}) \subset (L-\lambda_i)(\mathbb{E}) = \mathbb{F}.$$

El teorema 98 nos dice que si $M : \mathbb{F} \to \mathbb{F}$ es la restricción de L, el polinomio característico de M divide al de L, luego es también linealmente factorizable. Se puede aplicar la hipótesis inductiva a M y habrá una base $(u_{m+1}, \ldots, u_n)$ de $\mathbb{F}$ tal que

$$L(u_j) = M(u_j) = \sum_{m+1\leq i\leq n} a_j^i u_i, \quad j = m+1, \ldots, n.$$

[18] Se puede probar como ejercicio de virtuosismo con el manejo de índices que, al sustituir $\mathcal{U}$ por $\mathcal{V}$, las matrices a y b de L para $\mathcal{U}$ y $\mathcal{V}$ se obtienen una de otra "rotándolas 180 grados en torno al punto central", lo que generaliza dicho sobre matrices triangulares. Nos excusamos por la falta de rigor del entrecomillado.

Completamos $(u_{m+1},\dots,u_n)$ hasta una base $(u_1,\dots,u_m,u_{m+1},\dots,u_n)=\mathcal{U}$ de $\mathbb{E}$. En esta base L es triangulable, porque $(L-\lambda_i)(u_k)\in\mathbb{F}$ para $k\leq m$ da sucesivamente

$$(L-\lambda_i)(u_k)=\sum_{m+1\leq i\leq n}b_k^iu_i,\qquad L(u_k)=\lambda_iu_k+\sum_{m+1\leq i\leq n}b_k^iu_i,$$

y estas últimas son las condiciones de triangulación. En efecto, $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$ es

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)=\begin{pmatrix}\lambda_i&\cdots&&&&\\ \vdots&\ddots&\vdots&&&\\ 0&\cdots&\lambda_i&&&\\ b_1^{m+1}&\cdots&b_m^{m+1}&a_{m+1}^{m+1}&\cdots&\\ \vdots&\ddots&\vdots&\vdots&\ddots&\vdots\\ b_1^n&\cdots&b_m^n&a_{m+1}^n&\cdots&a_n^n\end{pmatrix},$$

y tiene todo ceros sobre la diagonal. ♣

La demostración elegida es muy rápida pero quizás no la mejor para calcular. Para los casos $n=2,3$, interesantes en la práctica, merece la pena un ataque directo que es parte de algo más general a tratar en el futuro: el teorema de Jordan. Mejora la triangulación con bases que hacen aparecer numerosos ceros, aunque si $n=2,3$ la mejora se nota menos. Si los valores propios son diferentes entre sí, L es diagonalizable (teorema 108) *e ignoraremos en adelante este caso.* Usaremos la abreviatura $L_\sigma=L-\sigma$, $\sigma\in\Bbbk$, y será de uso continuo que $L_\sigma(x)=y$ equivale a $L(x)=\sigma x+y$.

Teorema 112 *Si $n=2$ y L no es diagonalizable, se tiene $M(X)=(X-\lambda)^2$. En* todas *las bases $\mathcal{B}_u=(u,L_\lambda(u))$ la matriz de L es*

$$\begin{pmatrix}\lambda&0\\1&\lambda\end{pmatrix}.$$

Demostración. Es inmediato comprobar que en las bases $\mathcal{B}_v$ se tiene esa matriz. Mostramos que existen las bases $\mathcal{B}_v$. No puede ser $L_\lambda=0$ pues sería $M(X)\neq(X-\lambda)^2$. Tomamos u tal que $L_\lambda(u)=v\neq0$. Afirmamos que $\mathcal{B}_u=(u,v)=(u,L_\lambda(u))$ es base para lo que es suficiente mostrar la independencia. Si se tiene $0=\alpha u+\beta L_\lambda(u)$, aplicamos L_λ y como $M(L)=0$, obtenemos $0=\alpha v$ y $\alpha=0$. De $0=\beta v$ resulta $\beta=0$. ♣

Pasamos al caso $n=3$ basándonos el polinomio minimal. Como L no es diagonalizable, de acuerdo con el teorema 108, solo quedan tres posibilidades para $M(X)$ (se supone $\lambda\neq\mu$)

$$(\mathbf{a})\ (X-\lambda)^2(X-\mu),\quad(\mathbf{b})\ (X-\lambda)^2,\quad(\mathbf{c})\ (X-\lambda)^3. \tag{5.10}$$

Teorema 113 *Sea L no diagonalizable con $M(X)$ de una de las formas en (5.10). Hay bases $\mathcal{B}=(u,v,w)$ en donde la matriz de L es de la forma respectiva*[19]

$$(\mathbf{a})\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&0&\mu\end{pmatrix},\qquad(\mathbf{b})\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&0&\lambda\end{pmatrix},\qquad(\mathbf{c})\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&1&\lambda\end{pmatrix}.$$

Demostración. Caso **(a)** $M(X)=(\lambda-X)^2(\mu-X)$. No puede ser $L_\lambda^2=0$ o $L_\mu=0$ porque solo λ o μ serían valores propios. Tomemos u y w no nulos tales que $L_\lambda^2(u)=L_\mu(w)=0$ y sea $\mathcal{B}=(u,v,w)=(u,L_\lambda(u),w)$. Afirmamos que $\mathcal{B}$ es base. En efecto, sea $0=\alpha u+\beta L_\lambda(u)+\gamma w$ a apliquemos L_λ. No puede ser $L_\lambda(w)=0$ porque $w\neq0$ sería vector propio con valores distintos λ y μ. Tenemos $0=\alpha L_\lambda(u)+\gamma L_\lambda(w)$. Los vectores $L_\lambda(u)$ y $L_\lambda(w)$ son propios con valores propios *distintos* λ y μ (porque $L_\mu\circ L_\lambda(w)=L_\lambda\circ L_\mu(w)=0$) y entonces son independientes. Tenemos por ello $\alpha=\gamma=0$ y enseguida $\beta=0$ también. Una vez comprobado que $\mathcal{B}$ es base, resulta de $L_\lambda(u)=v$ y $L_\lambda(v)=L_\mu(w)$ que $L(u)=\lambda u+v$, $L(v)=\lambda v$ y $L(w)=\mu w$, y la matriz de L en $\mathcal{B}$ es **(a).**

Caso **(b)** $M(X)=(\lambda-X)^2$. No puede ser $L_\lambda=0$ porque L sería diagonalizable. Tomemos pues u tal que $L_\lambda(u)=v\neq0$ pero observando que $M(X)=(\lambda-X)^2$ implica que $L_\lambda(v)=0$. Tiene que

[19] Estas bases están el la tabla (5.11) un poco más adelante

ser $\operatorname{rg}(L_\lambda)=1$. Desde luego no puede ser 0 ni 3 y $\operatorname{rg}(L_\lambda)=2$ lleva a contradicción. Habría en efecto vectores independientes $L_\lambda(p)$ y $L_\lambda(q)$ y en cualquier base $\mathcal{U}=(p,q,r)$ las matríces de L y L_λ serían

$$\operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L)=\begin{pmatrix}\lambda&0&\alpha\\0&\lambda&\beta\\0&0&\gamma\end{pmatrix},\quad \operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L_\lambda)=\begin{pmatrix}0&0&\alpha\\0&0&\beta\\0&0&\gamma-\lambda\end{pmatrix}$$

siendo imposible $\operatorname{rg}(L_\lambda)=2$. Visto que $\operatorname{rg}(L_\lambda)=1$, deducimos que $\dim\ker(L_\lambda)=2$, y podemos tomar $w\in\ker(L_\lambda)$ independiente de w. Afirmamos que $\mathcal{B}=(u,v,w)=(u,L_\lambda(u),w)$ es base, para lo que basta que sea independiente. Si se tiene $0=\alpha u+\beta L_\lambda(u)+\gamma w$, se aplica L_λ y queda $0=\alpha v$, de donde $\alpha=0$ y $0=\beta L_\lambda(u)+\gamma w$. La independencia de $v=L_\lambda(u)$ y w da $\beta=\gamma=0$. Sabiendo que $\mathcal{B}$ es base, las condiciones $L_\lambda(u)=v$, $L_\lambda(v)=0$ y $L_\lambda(w)=0$ se reformulan con $L(u)=\lambda u+v$, $L(v)=\lambda v$ y $L(w)=\lambda w$, y la matriz de L en $\mathcal{B}$ es **(b).**

Caso **(c)** $M(X)=(\lambda-X)^3$. No puede ser $L_\lambda^2=0$ porque sería $M(X)\neq(\lambda-X)^3$. Tomamos u tal que sea $L_\lambda^2(u)=w\neq0$ y definamos $v=L_\lambda(u)$. Afirmamos que $\mathcal{B}=(u,v,w)=\left(u,L_\lambda(u),L_\lambda^2(u)\right)$ es base para lo que es suficiente probar la independencia. Si se tiene $0=\alpha u+\beta L_\lambda(u)+\gamma L_\lambda^2(u)$, se aplica L_λ^2 y $0=\alpha w$, luego $\alpha=0$ y $0=\beta L_\lambda(u)+\gamma L_\lambda^2(u)$. Aplicando otra vez L_λ queda $\beta w=0$ y $\beta=0$. Finalmente, de $\gamma w=0$ queda $\gamma=0$. Probado que $\mathcal{B}$ es base, Las condiciones $L_\lambda(u)=v$ y $L_\lambda(v)=w$ llevan enseguida como en los casos anteriores a que la matriz de L en $\mathcal{B}$ sea **(c).** ♣

Podemos hacer la siguiente tabla que sintetiza el teorema

$M(X)$	$\mathcal{B}=(u,v,w)$
$(X-\lambda)^2(X-\mu)$	$v=L_\lambda(u),\ w\in\ker L_\mu$
$(X-\lambda)^2$	$v=L_\lambda(u),\ w\in\ker(L_\lambda)$ independiente de v
$(X-\lambda)^3$	$v=L_\lambda(u),\ w=L_\lambda^2(u)$

(5.11)

Aunque los problemas de cálculo son muy pesados, vamos a decirle al lector cómo ponerse tantos como quiera, sabiendo de antemano cuál es la respuesta correcta. Abstraemos con $\lambda\neq\mu$ arbitrarios porque quizás se sigue mejor el argumento, pero se pueden poner números concretos *distintos.* Se toma una matriz $b\in\Bbbk^{3\times3}$ de una de las formas **(a)**, **(b)** o **(c)** y otra invertible c (preferiblemente con c^{-1} fácilmente calculable). Entonces, definiendo $a=c^{-1}bc$ tendremos que en la base formada por las tres columnas de c^{-1} (¡no de c!) $L(x)=ax$ tiene la matriz "bonita" b. Un ejemplo puede ser

$$\begin{pmatrix}1&0&0\\-1&1&0\\-1&0&1\end{pmatrix}\begin{pmatrix}\mu&0&0\\0&\lambda&1\\0&0&\lambda\end{pmatrix}\begin{pmatrix}1&0&0\\1&1&0\\1&0&1\end{pmatrix}=\begin{pmatrix}\mu&0&0\\\lambda-\mu+1&\lambda&1\\\lambda-\mu&0&\lambda\end{pmatrix}=a.$$

Problema 231 *Hacer la tarea propuesta para $L:\Bbbk^3\to\Bbbk^3$ con matriz a en la base estándar.* ♦

Solución. Como matrices semejantes tienen el mismo polinomio característico, se puede contar con el dato adicional $C(X)=-(X-\lambda)^2(X-\mu)$. Si no conocieramos el origen de a, la "trampa", necesitaríamos calcularlo, pero es un caso fácil. Tenemos

$$(a-\lambda)(a-\mu)=\begin{pmatrix}\mu-\lambda&0&0\\\lambda-\mu+1&0&1\\\lambda-\mu&0&0\end{pmatrix}\begin{pmatrix}0&0&0\\\lambda-\mu+1&\lambda-\mu&1\\\lambda-\mu&0&\lambda-\mu\end{pmatrix}=\begin{pmatrix}0&0&0\\\lambda-\mu&0&\lambda-\mu\\0&0&0\end{pmatrix}.$$

Si tomamos $u=(0,0,1)^\top$ y $v=(0,1,0)^\top=(a-\lambda)(0,0,1)^\top=(a-\lambda)u$ ya tenemos los dos primeros vectores de la base. Hay que completarla con w tal que $(a-\mu)w=0$ y $(-1,1,1)^\top$ sirve como w. Efectivamente,

$$\begin{pmatrix}0&0&-1\\0&1&1\\1&0&1\end{pmatrix}^{-1}\begin{pmatrix}\mu&0&0\\\lambda-\mu+1&\lambda&1\\\lambda-\mu&0&\lambda\end{pmatrix}\begin{pmatrix}0&0&-1\\0&1&1\\1&0&1\end{pmatrix}=\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&0&\mu\end{pmatrix},$$

que es la b de la preparación del problema "reorganizada". ♦

Problema 232 *Hacer otro problema similar para a, siendo*

$$\begin{pmatrix}1&0&0\\1&1&0\\1&0&1\end{pmatrix}^{-1}\begin{pmatrix}1&0&1\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\1&1&0\\1&0&1\end{pmatrix}=\begin{pmatrix}2&0&1\\0&1&0\\-1&0&0\end{pmatrix}=a.$$

5.12. Un primer contacto con la forma de Jordan

Al triangular matrices en la sección anterior buscando las mejores bases, vimos que había en ellas sectores de la forma u, $L_\lambda(u)=v$, $L_\lambda^2(u)=w$. Para $u\in\mathbb{E}$ supongamos que se tiene una sucesión de vectores independientes $(u_0,u_1,\dots,u_{q-1})=Z(u)$, llamado el **ciclo de** u, definido por

$$u_0=L_\lambda^0(u),\ u_1=L_\lambda^1(u),\ u_2=L_\lambda^2(u),\ \dots\ ,u_{q-1}=L_\lambda^{q-1}(u),\ u_q=L_\lambda^q(u)=0$$

o, de modo equivalente, por

$$u_0,\ u_1=L_\lambda(u_0),\ u_2=L_\lambda(u_1),\ \dots\ ,\ u_{q-1}=L_\lambda(u_{q-2}),\ u_q=L_\lambda(u_{q-1})=0.$$

Al ser $L_\lambda(x)=(L-\lambda)x=y$ deducimos que $L(x)=\lambda x+y$ y las $L(u_j)$ se expresan como

$$L(u_0)=\lambda u_0+u_1,\ L(u_1)=\lambda u_1+u_2,\ \dots\ ,\ L(u_{q-2})=\lambda u_{q-2}+u_{q-1},\ L(u_{q-1})=\lambda u_{q-1}.$$

Si suponemos que $Z(u)$ es una base de $\mathbb{E}$ (luego $q=n$), la matriz de L en $\mathcal{B}$ es de la forma

$$J_\lambda(q)=\begin{pmatrix}\lambda&&&&&\\1&\lambda&&&&\\&1&\lambda&&&\\&&&\ddots&&\\&&&&\lambda&\\&&&&1&\lambda\end{pmatrix}\in\mathbb{k}^{q\times q}.$$

Es sencillo constatar que si en vez de $(u_0,u_1,\dots,u_{q-1})$ utilizamos como base $(u_{q-1},u_{q-2},\dots,u_1,u_0)$ la matriz de L es la traspuesta de la anterior con "los unos sobre las lambdas".[20] Estas matrices se llaman **bloques de Jordan de dimensión** q **para el valor propio** λ. Es excepcional tener una base tan buena para L, pero el **teorema de Jordan** dice que *si* $C(X)$ *es linealmente factorizable,*[21] *podemos yuxtaponer bases del tipo* $Z(u)$, *con* λ *recorriendo los valores propios de* L, *de modo que se obtiene una base de todo* $\mathbb{E}$. En esa base, que se llama **base de Jordan** (y no es única), la matriz de L tiene forma semidiagonal

$$\begin{pmatrix}J_\bullet(\bullet)&&&\\&\ddots&&\\&&J_\bullet(\bullet)\end{pmatrix}$$

llamándose **matrices de Jordan** a estas matrices. Las matrices diagonales son casos particulares con todas las $J_\bullet(\bullet)\in\mathbb{k}^{1\times1}$ y una base que diagonalice L (si existe) es base de Jordan.

Aunque hay un número finito de posibilidades, porque hay un número finito de valores propios, la simple enumeración de posibilidades es una dura tarea y el conocer la **matriz de Jordan de** L, la que le corresponde, es aún más dura. La teoría general es objeto de otro capítulo pero para $n=\dim(\mathbb{E})=2,3$ tenemos ya el trabajo hecho. *Las bases de Jordan y sus matrices son las bases del teorema* 113 y son consecuencia suya estas versiones "pequeñas" del teorema de Jordan.[22]

Teorema 114 *Si* $n=2$ *y* L *no es diagonalizable, se tiene* $M(X)=(X-\lambda)^2$, *hay una base de Jordan* $\mathcal{J}$ *de la forma* $(u,(L-\lambda)(u))$, *y la matriz de Jordan es*

$$\mathrm{mat}_{\mathcal{J}}^{\mathcal{J}}(L)=\begin{pmatrix}\lambda&0\\1&\lambda\end{pmatrix}. \tag{5.12}$$

Teorema 115 *Para* $n=3$ *y* L *no diagonalizable, según sea el polinomio minimal*

$$\textbf{(a)}\ (X-\lambda)^2(X-\mu),\quad \textbf{(b)}\ (X-\lambda)^2,\quad \textbf{(c)}\ (X-\lambda)^3$$

existe una base de Jordan $\mathcal{B}=(u,v,w)$ *en donde, respectivamente,* L *tiene matriz*

$$\textbf{(a)}\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&0&\mu\end{pmatrix},\qquad \textbf{(b)}\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&0&\lambda\end{pmatrix},\qquad \textbf{(c)}\begin{pmatrix}\lambda&0&0\\1&\lambda&0\\0&1&\lambda\end{pmatrix}. \tag{5.13}$$

[20] Cada autor tiene una preferencia. Preferimos "los unos *bajo* las lambdas" porque en las bases de Jordan están compuestas por segmentos $Z(u)$ y con la convención "los unos *sobre* las lambdas" habría que invertirlos.

[21] En $\mathbb{k}=\mathbb{C}$ esto se tiene siempre, pero si $\mathbb{k}\neq\mathbb{C}$ puede ser imposible obtener la base y matriz de Jordan.

[22] Insistimos en que si L es diagonalizable su matriz en la forma diagonal y las bases en que esto sucede, son de Jordan.

Antes de entrar en problemas de cálculo, debemos hacer una reflexión importante: sabemos que L determina unívocamente a su polinomio mínimal, que el polinomio minimal nos permite escribir la matriz de Jordan,[23] y que la matriz de Jordan nos permite recuperar el polinomio minimal. Por tanto, para cada L hay una única matriz de Jordan si ignoramos reordenaciones de valores propios o de los bloques de Jordan. Esto dice, al menos por lo visto para $n = 2, 3$ (y cierto en general, pero cuesta más probarlo) que *la matriz de Jordan clasifica endomorfismos.* Aclaramos que se trata de los endomorfismos tales que $C(X)$ sea linealmente factorizable, pero en $\mathbb{k} = \mathbb{C}$ esto se cumple siempre. Sabemos por ejemplo que, una vez fijados los valores propios, hay para $n = 2$ tres tipos de matrices de Jordan, siendo (5.12) la única no diagonalizable. Si $n = 3$, hay tres tipos no diagonalizables (5.13) y tres diagonalizables, estando en la diagonal $(\lambda, \lambda, \lambda)$, (λ, λ, μ) y (λ, μ, θ). Los problemas de puro cálculo si $n = 2, 3$ no tienen dificultad en cuanto al procedimiento a seguir. Es mucho más fácil saber cual será la matriz de Jordan de L o de a que determinar la base de Jordan. Lo primero (insistimos, si $n = 2, 3$) se tiene nada más disponer de $M(X)$ factorizado. Para la base hay que calcularla con el teorema 113 o la tabla (5.11).

Problema 233 *Dar la matriz y una base de Jordan para L que en la base estándar de $\mathbb{R}^n$ tiene matrices*

$$a = \begin{pmatrix} 0 & 2 \\ -2 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}. \blacklozenge$$

Solución. El polinomio característico de a es $C(X) = (X-2)^2$. Se comprueba que $a - 2 \neq 0$ luego $M(X) = C(X) = (X-2)^2$. La matriz de Jordan y $a - 2$ son respectivamente

$$\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}, \quad a - 2 = \begin{pmatrix} -2 & 2 \\ -2 & 2 \end{pmatrix}.$$

La base de Jordan es $\mathcal{J} = (u, v) = (u, L_\lambda(u))$ con $v \neq 0$. Lo más sencillo es tomar $u = (1, 0)^\top$ y $v = (-2, -2)$. Efectivamente,

$$\begin{pmatrix} 1 & -2 \\ 0 & -2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}.$$

Para b tenemos $C(X) = (3 - X)^3$ y $(b-3)^2 = 0$ y $M(X) = (X-3)^2$ ya que $b - 3 \neq 0$. Según el teorema 115 tendremos una base de Jordan (u, v, w) tomando u tal que $u = L_\lambda(v) = L_3(v)$ sea no nulo y añadiendo como w un vector en $\ker(L_\lambda)$ independiente de u. Calculamos y elegimos

$$b - 3 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v = (b-3)\, u = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Como $\ker(b-3)$ está generado por $(0, 1, 0)^\top$ y $(1, 0, -1)^\top$ es natural elegir $w = (1, 0, -1)^\top$. La base de Jordan es (u, v, w). Para nuestra tranquilidad el ordenador constata que

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}. \blacklozenge$$

Problema 234 *Calcular la matriz de Jordan y una base de Jordan para a_1 y a_2 definidas por*

$$a_1 = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ -2 & 1 & 1 \end{pmatrix}.$$

Problema 235 *¿Es posible calcular matriz y base de Jordan para*

$$a = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \ ?$$

[23] Si $n \geq 4$ el mero conocimiento de $M(X)$ no determina unicamente la forma de Jordan.

Problema 236 *Calcular base y matriz de Jordan para*

$$a=\begin{pmatrix}\alpha & 0 & \alpha-2\\ 1 & 1 & 1\\ 0 & 0 & 2\end{pmatrix}$$

cuando ***no sea*** *diagonalizable.*

Problema 237 *Pedimos la matriz de Jordan (aunque no la base de Jordan) de*

$$a=\begin{pmatrix}1 & 0 & -\alpha^2+1\\ 1 & \alpha & 1\\ 0 & 0 & \alpha^2\end{pmatrix}$$

según el valor de $\alpha\in\mathbb{R}$.

5.13. Triangulación en el caso real

Vimos en el teorema 111 que L es triangulable si y solo si $C(X)$ es linealmente factorizable. Si $\Bbbk=\mathbb{R}$ esto no está garantizado porque puede haber factores tipo $(X-p)^2+q^2$. Hay sin embargo una posibilidad de triangular L, algo peor que la anterior pero que permite hallar una base donde la matriz, si no diagonal, al menos tenga cajas 2×2 como máximo en ella y ceros bajo esta "diagonal ampliada".

Teorema 116 *Sea L un endomorfismo cuyo polinomio característico se factoriza en la forma*[24]

$$C(X)=(X-\lambda_1)\cdots(X-\lambda_p)\left(X^2+\alpha_1X+\beta_1\right)\cdots\left(X^2+\alpha_qX+\beta_q\right),$$

con los $X^2+\alpha X+\beta$ no linealmente factorizables. Hay una base donde la matriz es

$$a=\begin{pmatrix}\ddots & & & * & *\\ & H_i & & & *\\ & & \ddots & & \\ 0 & & & K_j & \\ 0 & 0 & & & \ddots\end{pmatrix},\quad \textit{siendo}\quad H_i=(\lambda_i)\in\mathbb{R}^{1\times1},\quad \textit{y}\quad K_j=\begin{pmatrix}0 & -\alpha_j\\ 1 & -\beta_j\end{pmatrix}\in\mathbb{R}^{2\times2}.$$

Hay ceros bajo las matrices H y K y coeficientes arbitrarios sobre ellas.

Demostración. Admitamos que se da una conclusión más débil del teorema, y es que sea $H_i=(r_i)$ y

$$K_j=\begin{pmatrix}0 & -s_j\\ 1 & -t_i\end{pmatrix}$$

sin presuponer que sea $r_i=\lambda_i$, $s_j=\alpha_j$ y $t_j=\beta_j$ en la matriz a de L. Se tendría entonces

$$C(X)=C_a(X)=(X-r_1)\cdots(X-r_p)\left(X^2+s_1X+t_1\right)\cdots\left(X^2+s_qX+t_q\right)$$

y, por la unicidad de la factorización de $C(X)$ debe cumplirse $r_i=\lambda_i$, $s_j=\alpha_j$ y $t_j=\beta_j$. Es por tanto suficiente el probar la versión más débil del teorema, cosa que vamos a hacer por inducción sobre $n=\dim(\mathbb{E})$ y siendo obvio si $n=1$.

Observamos que no todos los subespacios $(L-\lambda_i)(\mathbb{E})$ y $\left(L^2+\alpha_jL+\beta_j\right)(\mathbb{E})$ pueden ser iguales a $\mathbb{E}$, porque si esto sucediera $C(L)(\mathbb{E})=\mathbb{E}$ y $\mathbb{E}=0$. Sea i o j uno de estos índices de modo que $\mathbb{F}=(L-\lambda_i)(\mathbb{E})$ o $\mathbb{F}=\left(L^2+\alpha_jL+\beta_j\right)(\mathbb{E})$ cumpla $\mathbb{F}\neq\mathbb{E}$. Puede verse como en la demostración del teorema 111 que $L(\mathbb{F})\subset\mathbb{F}$. De esto deriva una idea sutil que hace que la demostración de este teorema sea algo distinta de la del teorema 111. Consideramos el conjunto de los subespacios $\mathbb{G}$ de $\mathbb{E}$ tales que **(a)** $\mathbb{F}\subset\mathbb{G}\neq\mathbb{E}$ y **(b)** $L(\mathbb{G})\subset\mathbb{G}$, que es no vacío pues $\mathbb{F}$ está en él. Allí tomsmos un subespacio $\mathbb{G}$ de *dimensión máxima,* que cumple pues **(a)**, **(b)**, y $\mathbb{G}\neq\mathbb{E}$ (podría ser $\mathbb{F}=\mathbb{G}$). Distinguimos según se pueda elegir $\mathbb{F}$

[24] Interesa para mejor comprender el teorema poner todos los factores de $C(X)$ con exponente 1, pero advirtiendo que los $(X-\lambda_i)$ o $\left(X^2+\alpha_jX+\beta_j\right)$ pueden estar repetidos.

1. $\mathbb{F} = (L - \lambda_i)(\mathbb{E})$. Se toma $v \in \mathbb{E} - \mathbb{G}$ y el subespacio $\mathbb{H}$ generado por v y $\mathbb{G}$. Claramente,

$$(L - \lambda_i)(v) \in \mathbb{F} \subset \mathbb{G} \subset \mathbb{H} \text{ y } L(v) = \lambda_i v + (L - \lambda_i)(v) \in \mathbb{H}.$$

Esto nos dice que $L(\mathbb{H}) \subset \mathbb{H}$ y, como $\dim(\mathbb{H}) > \dim(\mathbb{G})$, para no caer en contradicción con la definición de $\mathbb{G}$, solo queda la posibilidad $\mathbb{H} = \mathbb{E}$. Por hipótesis inductiva hay una base $(u_1, \ldots, , u_{n-1})$ de $\mathbb{G}$ tal que la restricción de L a $\mathbb{G}$ tiene matriz en la forma deseada. Ampliamos $(u_1, \ldots, , u_{n-1})$ a $(u_1, \ldots, , u_{n-1}, v)$ que será base de $\mathbb{E}$. Para conocer la matriz de L en ella se ve que $L(v) = \lambda_i v + (L - \lambda_i)(v)$ y $(L - \lambda_i)(v)$ es combinación de $(u_1, \ldots, , u_{n-1})$, por lo que la matriz es

$$\begin{pmatrix} H_1 & & * & * & * \\ & H_2 & & * & * \\ 0 & & \ddots & & \\ 0 & 0 & & H_\ell & * \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix}.$$

2. $\mathbb{F} = \left(L^2 + \alpha_j L + \beta_j\right)(\mathbb{E})$. Observar que en este caso, $\beta_j \neq 0$ pues si no sería $X^2 + \alpha_j X + \beta_j$ linealmente factorizable. Se toma $v \in \mathbb{E} - \mathbb{G}$ y se define $\mathbb{H} = \lg(v, L(v)) + \mathbb{G}$. En particular,

$$\left(L^2 + \alpha_j L + \beta_j\right)(v) = w \in \mathbb{F} \text{ implica que } L^2(v) = -\alpha_j L(v) - \beta_j v + w \in \mathbb{H}.$$

Con esto, es fácil ver que $\mathbb{H}$ cumple $\mathbb{F} \subset \mathbb{H}$ y $L(\mathbb{H}) \subset \mathbb{H}$ y $\dim(\mathbb{H}) \geq \dim(\mathbb{G})+1$. Para no contradecir la definición de $\mathbb{G}$ hay que admitir que $\mathbb{H} = \mathbb{E}$. Caben dos posibilidades

a) Si $L(v) = \sigma v$, se tiene que $\dim(\mathbb{E}) = \dim(\mathbb{G}) + 1$ y se amplía una base $(u_1, \ldots, u_{n-1})$ de $\mathbb{G}$, donde la restricción de L a $\mathbb{G}$ tenga la forma deseada, a una base $(u_1, \ldots, u_{n-1}, v)$ de $\mathbb{E}$. Es inmediato que en esa base L tiene matriz de la forma deseada y (σ) es la última caja 1×1 en vez de (λ_i) como en la matríz de arriba.

b) Si v y $L(v)$ fuesen independientes sería $\dim(\mathbb{E}) = \dim(\mathbb{G}) + 2$ y se extendería una base previa $(u_1, \ldots, u_{n-2})$ de $\mathbb{G}$ donde la restricción de L a $\mathbb{G}$ tenga la forma deseada, a una base $(u_1, \ldots, u_{n-2}, v, L(v)) = (u_1, \ldots, u_{n-2}, u_{n-1}, u_n)$ de $\mathbb{E}$. En esa base $L(u_{n-1}) = u_n$ y

$$L(u_n) = L^2(v) = \left(L^2 + \alpha_j L + \beta_j\right)(v) - \alpha_j L(v) - \beta_j v = \left(L^2 + \alpha_j L + \beta_j\right)(v) - \alpha_j u_n - \beta_j u_{n-1}$$

y como $\left(L^2 + \alpha_j L + \beta_j\right)(v) \in \mathbb{G}$ es combinación lineal de $(u_1, \ldots, u_{n-2})$, la matriz

$$\begin{pmatrix} H_1 & & * & * & & * \\ & H_2 & & * & & * \\ 0 & & \ddots & & & \\ 0 & 0 & & H_\ell & & * \\ 0 & 0 & & 0 & 0 & -\alpha_j \\ 0 & 0 & \cdots & 0 & 1 & -\beta_j \end{pmatrix}$$

es de la forma deseada.

Probada la versión más débil, se tiene, como dijimos, la más fuerte. ♣

5.14. Ecuaciones lineales homogéneas con coeficientes constantes

Esta sección podía figurar al final del capítulo tercero. La ponemos aquí porque la sección *Ecuaciones diferenciales matriciales. El caso fácil* ha tratado de la función exponencial y del concepto de ecuación diferencial, y ayudará a leer esta en cuánto a qué es una ecuación diferencial y su espacio de soluciones.

Sea un polinomio $P(X) = a_0 + a_1 X + a_2 X^2 + \ldots + a_{n-1} X^{n-1} + X^n$ con coeficientes *complejos* y $\mathbb{E}$ el espacio $C^\infty(\mathbb{R}, \mathbb{C})$ de las funciones infinitamente diferenciables $x : \mathbb{R} \to \mathbb{C}$. Sea $D : \mathbb{E} \to \mathbb{E}$ la derivada y $x^{(k)}$ la derivada k de la función x con el convenio $x^{(0)} = x$. La función

$$P(D) : \mathbb{E} \to \mathbb{E}, \quad P(D)(x) = a_0 x^{(0)} + a_1 x^{(1)} + a_2 x^{(2)} + \ldots + a_{n-1} x^{(n-1)} + x^{(n)},$$

es una función lineal. Por ejemplo, si $P(X) = X^2 - 3X + 2$ y $x(t) = \cos t$, tenemos una nueva función

$$(P(D)(x))(t) = \frac{d}{dt}\frac{d}{dt}(\cos t) - 3\frac{d}{dt}(\cos t) + 2\cos t = \cos t + 3\sin t.$$

Una **ecuación lineal homogénea** es la de la forma $P(D)(x) = 0$ y una solución de ella es una función que la verifica.[25] Para $P(X) = X^2 - 3X + 2$, $x(t) = \cos t$ no es solución pero $x(t) = e^t$ sí lo es. Se trata de dar todas las soluciones de $P(D)(x) = 0$. Será esencial el poder factorizar $P(X)$ como

$$P(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_k)^{m_k}, \quad \lambda_1, \ldots, \lambda_k \in \mathbb{C},$$

y usaremos con frecuencia las fórmulas

$$(D - \alpha)\left(e^{\alpha t}\right) = 0, \quad (D - \alpha)\left(e^{\beta t}\right) = (\beta - \alpha)e^{\beta t}. \tag{5.14}$$

Son inmediatas porque $(D - \alpha)(e^{\alpha t}) = D(e^{\alpha t}) - \alpha(e^{\alpha t}) = \alpha e^{\alpha t} - \alpha e^{\alpha t} = 0$. Por otra parte,

$$(D - \alpha)\left(e^{\beta t}\right) = (D - \beta + \beta - \alpha)\left(e^{\beta t}\right) = (D - \beta)\left(e^{\beta t}\right) + (\beta - \alpha)e^{\beta t} = 0 + (\beta - \alpha)e^{\beta t}.$$

Teorema 117 *La ecuación $(D - \alpha)(x) = 0$ tiene como espacio de soluciones $\lg(e^{\alpha t})$, la recta en $\mathbb{E} = C^\infty(\mathbb{R}, \mathbb{C})$ que genera la función $e^{\alpha t}$. Por otra parte, $(D - \alpha) : \mathbb{E} \to \mathbb{E}$ es suprayectiva.*

Demostración. De acuerdo con (5.14), $e^{\alpha t}$ es solución. Si x es otra solución, $y(t) = x(t)e^{-\alpha t}$ cumple $y' = 0$, luego y toma valor constante $k \in \mathbb{C}$ y $x(t) = ke^{\alpha t}$.

Sea $y \in \mathbb{E}$. Debemos encontrar x tal que $(D - \alpha)(x) = y$. Buscamos la solución en la forma $x(t) = z(t)e^{\alpha t}$ con z a determinar. Calculamos

$$(D - \alpha)(x(t)) = x'(t) - \alpha x(t) = z'(t)e^{\alpha t} + z(t)\alpha e^{\alpha t} - \alpha z(t)e^{\alpha t} = z'(t)e^{\alpha t}.$$

Esto indica que $x(t) = z(t)e^{\alpha t}$ cumple $(D - \alpha)(x) = y$ si $z'(t) = y(t)e^{-\alpha t}$. Se puede conseguir z integrando las componentes real y compleja de $y(t)e^{-\alpha t}$. ♣

Ilustramos esto último para $y(t) = \frac{t^p}{p!}e^{\alpha t}$. ¿Cómo es $x(t)$ tal que $(D - \alpha)(x) = y$? Debe cumplir $x(t) = z(t)e^{\alpha t}$ y $z'(t) = y(t)e^{-\alpha t} = t^p/p!$, luego $z(t) = t^{p+1}/(p+1)!$ y

$$(D - \alpha)\left(\frac{t^{p+1}}{(p+1)!}e^{\alpha t}\right) = \frac{t^p}{p!}e^{\alpha t}. \tag{5.15}$$

Problema 238 *Determinar una función x tal que $(D - i)^2(x) = t^3 e^{it}$.* ♦

Solución. Sea $(D - i)(x) = u$. Determinamos u tal que $(D - i)(u) = t^3 e^{it} = 3!\left(t^3/3!\right)e^{it}$ y con esto

$$u(t) = 3!\frac{t^4}{4!}e^{it}, \quad x(t) = 3!\frac{t^5}{5!}e^{it} = \frac{t^5}{20}e^{it}.$$ ♦

Si nos preguntan lo mismo pero con $y(t) = \sqrt{1 + t^2}$ en vez de $y(t) = t^3 e^{it}$ hay que calcular integrales, con el trabajo que supone. Sin embargo, si nos limitamos al caso de que sea $y(t) = Q(t)e^{\lambda t}$ con $Q(X)$ un polinomio, es fácil encontrar con (5.15) $x(t)$ tal que $(D - \lambda)(x) = y$. Es más laborioso, como luego veremos, resolver $(D - \lambda)(x) = y$ con $y(t) = Q(t)e^{\beta t}$ y $\beta \neq \lambda$, pero hay una fórmula directa que evita el cálculo de integrales. El siguiente teorema es de tipo técnico.

Teorema 118 *Sea $\mathbb{E}$ un espacio sobre $\Bbbk$ arbitrario, quizás de dimensión infinita. Sean L y M endomorfismos de $\mathbb{E}$ cuyos núcleos tienen dimensión finita p y q y, además, M es suprayectivo. Entonces* $\dim(L \circ M) = \dim(L) + \dim(M)$.

Demostración. Sea $(u_1, \ldots, u_p)$ una base de $\ker(L)$ que, como M es suprayectiva, escribiremos como $(M(v_1), \ldots, M(v_p))$. Sea $(w_1, \ldots, w_q)$ base de $\ker(M)$. Mostramos que $\mathcal{B} = (v_1, \ldots, v_p, w_1, \ldots, w_q)$ es base de $\ker(L \circ M)$ con lo que el teorema es inmediato. Sin duda, $(L \circ M)(v_i) = L(u_i) = 0$ y $(L \circ M)(w_i) = L(0) = 0$. Si $(L \circ M)(x) = 0$ se tiene $M(x) \in \ker(L)$, lo que nos da sucesivamente

$$M(x) = \lambda^1 M(v_1) + \ldots + \lambda^p M(v_p), \quad x - \lambda^1 v_1 - \ldots \lambda^p v_p \in \ker(M), \quad x - \lambda^1 v_1 - \ldots \lambda^p v_p = \mu^1 w_1 + \ldots + \mu^p w_q,$$

[25] Lo de "coeficientes constantes" del título de la sección se debe a los coeficientes constantes del polinomio. Si en $P(X)$ los a_i fuesen funciones de t, tendría sentido $P(D)(x) = 0$, pero sería una ecuación mucho más difícil de resolver.

y por consiguiente, $\mathcal{B}$ genera $\ker(L \circ M)$. También se da que $\mathcal{B}$ es independiente porque si

$$\lambda^1 v_1 + \ldots \lambda^p v_p + \mu^1 w_1 + \ldots + \mu^p w_q = 0,$$

se aplica M y $0 = \lambda^1 u_1 + \ldots \lambda^p u_p$. La independencia de $(u_1, \ldots, u_p)$ da $\lambda^1 = \ldots = \lambda^p = 0$ luego $\mu^1 w_1 + \ldots + \mu^p w_q = 0$, Ahora, la independencia de $(w_1, \ldots, w_q)$ da $\mu^1 = \ldots = \mu^p = 0$. ♣

Teorema 119 *El espacio de soluciones* $\mathbb{S}$ *de* $P(D)(x) = 0$ *tiene dimensión* $n = \deg(P(X))$.

Demostración. Hay funciones x tales que $P(D)(x) = 0$, pues, vista la factorización, basta que sea $(D - \lambda_k)(x) = 0$ para que se tenga $P(D)(x) = 0$. Probamos por inducción sobre n que $\dim(\mathbb{S}) = n$. Es obvio si $n = 1$ por el teorema 117. Si es cierto el teorema para $n - 1$ lo probamos para n con el teorema 118. Factorizamos $P(X) = Q(X)(D - \lambda_k)$. Claramente $L = Q(D)$ y $M = D - \lambda_k$ son endomorfismos de $\mathbb{E} = C^\infty(\mathbb{R}, \mathbb{C})$, siendo $L \circ M = P(D)$ y con sus núcleos de dimensiones $n - 1$ (hipótesis inductiva) y 1 (teorema 117). El teorema 118 dice que $\mathbb{S} = \ker(L \circ M)$ tiene dimensión $(n - 1) + 1 = n$ y el paso inductivo está completo. ♣

Queremos encontrar una base de $\mathbb{S}$, el espacio de soluciones de $P(D)(x) = 0$. Hay un caso sencillo.

Teorema 120 *Si* $P(X)$ *tiene las* n *raíces* simples*, la sucesión de funciones* $\left(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}\right)$ *es base de* $\mathbb{S}$.

Demostración. Primeramente, cada $e^{\lambda_i t}$ está en $\mathbb{S}$ porque

$$\begin{aligned} P(D)\left(e^{\lambda_i t}\right) &= (D - \lambda_1) \circ \cdots \circ (D - \lambda_i) \circ \cdots (D - \lambda_n)\left(e^{\lambda_i t}\right) \\ &= (D - \lambda_1) \circ \cdots \circ \widehat{(D - \lambda_i)} \circ \cdots (D - \lambda_n) \circ (D - \lambda_i)\left(e^{\lambda_1 t}\right) \\ &= (D - \lambda_1) \circ \cdots \circ \widehat{(D - \lambda_i)} \circ \cdots (D - \lambda_n)(0) = 0. \end{aligned}$$

Probamos por inducción sobre n la independencia de $\left(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}\right)$, siendo trivial si $n = 1$. Supuesto cierta para $n - 1$ consideramos una combinación $0 = \mu^1 e^{\lambda_1 t} + \ldots + \mu_n e^{\lambda_n t}$. Con (5.14) se tiene que

$$\begin{aligned} 0 &= (D - \lambda_n)\left(\mu^1 e^{\lambda_1 t} + \ldots + \mu_n e^{\lambda_n t}\right) = \mu^1 (D - \lambda_n) e^{\lambda_1 t} + \ldots + \mu_n (D - \lambda_n) e^{\lambda_n t} \\ &= \mu_1 (\lambda_1 - \lambda_n) e^{\lambda_1 t} + \ldots + \mu_{n-1} (\lambda_1 - \lambda_n) e^{\lambda_{n-1} t} + 0. \end{aligned}$$

Por hipótesis inductiva $\mu_1 (\lambda_1 - \lambda_n) = \ldots = \mu_{n-1} (\lambda_1 - \lambda_n) = 0$ y, al ser los λ distintos entre sí, $\mu_1 = \ldots = \mu_{n-1} = 0$. Queda $0 = \mu_n e^{\lambda_n t}$ de donde $\mu_n = 0$. Probada la independencia y sabido que que $\dim(\mathbb{S}) = n$ (teorema 119), $\left(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}\right)$ necesariamente es base. ♣

Problema 239 *Sea* $\lambda \in \mathbb{C}$ *no real. Dar la solución de* $(D - \lambda) \circ \left(D - \bar{\lambda}\right)(x) = 0$ *tal que* $x(0) = 2i$, $\dot{x}(0) = 0$. ♦

Solución. Directamente, la solución general es $x(t) = \sigma e^{\lambda t} + \tau e^{\bar{\lambda} t}$, y $\dot{x}(t) = \sigma \lambda e^{\lambda t} + \tau \bar{\lambda} e^{\bar{\lambda} t}$ debemos encontrar σ y τ de forma que sea

$$\begin{cases} \sigma + \tau = 2i \\ \lambda\sigma + \bar{\lambda}\tau = 0 \end{cases} \text{ que tiene solución } \sigma = \frac{1}{v}(2iv - 1),\ \tau = \frac{1}{v}$$

para $\lambda = u + vi$ con $u, v \in \mathbb{R}$ y $v \neq 0$. Esto se ve con el cálculo

$$\begin{pmatrix} 1 & 1 & 2i \\ \lambda & \bar{\lambda} & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2i \\ 0 & -2vi & -2i \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2i \\ 0 & 1 & \frac{1}{v} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & \frac{1}{v}(2iv - 1) \\ 0 & 1 & \frac{1}{v} \end{pmatrix}. \blacklozenge$$

Problema 240 *Determinar todas las funciones* reales *(si existen) tales que* $x'' + x = 0$ *y* $x(0) = 0$.

Como siguiente paso vamos a estudiar el espacio de soluciones $\mathbb{S}$ cuando $P(X) = (X - \lambda)^m$.

Teorema 121 *El espacio de soluciones* $\mathbb{S}$ *cuando* $P(X) = (X - \lambda)^m$ *tiene base*

$$\mathcal{B} = \left(\frac{1}{0!}e^{\lambda t}, \frac{1}{1!}te^{\lambda t}, \frac{1}{2!}t^2 e^{\lambda t}, \ldots, \frac{1}{(m-1)!}t^{m-1}e^{\lambda t}\right).$$

Demostración. Sean $z_0(t), \ldots, z_{n-1}(t)$ las funciones de $\mathcal{B}$. Con (5.15) obtenemos enseguida que $(D-\lambda)^h (z_k(t)) = z_{k-h}(t)$, entendiendo que $z_j(t) = 0$ si $j < 0$. Queda claro que $(D-\lambda)^m (z_k(t)) = 0$ para todo k y los $z_k(t) \in \mathcal{B}$. Para que $\mathcal{B}$ sea base es suficiente probar la independencia porque hay m funciones z y $\dim(\mathbb{S}) = m$ (teorema 119). Si se tuviera

$$\begin{aligned} 0 &= \mu^0 \frac{1}{0!} e^{\lambda t} + \mu^1 \frac{1}{1!} t e^{\lambda t} + \mu^2 \frac{1}{2!} t^2 e^{\lambda t} + \ldots + \mu^{m-1} \frac{1}{(m-1)!} t^{m-1} e^{\lambda t} \\ &= e^{\lambda t} \left(\mu^0 + \mu^1 t + \mu^2 \frac{1}{2!} t^2 + \ldots + \mu^{m-1} \frac{1}{(m-1)!} t^{m-1} \right) \end{aligned}$$

la función del paréntesis sería nula, pues la exponencial no se anula, y $\mu^0 = \mu^1 = \ldots = \mu^{m-1} = 0$. ♣

Problema 241 *Dar todas las soluciones de* $(D-i)^3 (x)$ *tales que* $x(0) = x'(0) = 0$.

Teorema 122 *Sea* $P(X) = (X-\lambda_1)^{m_1} \cdots (X-\lambda_k)^{m_k}$. *El espacio de soluciones* $\mathbb{S}$ *de* $P(D)(x) = 0$ *tiene base*

$$\mathcal{B} = \left(\frac{1}{0!} e^{\lambda_1 t}, \frac{1}{1!} t e^{\lambda_1 t}, \ldots, \frac{1}{(m_1-1)!} t^{m_1-1} e^{\lambda_1 t}, \ldots, \frac{1}{0!} e^{\lambda_k t}, \frac{1}{1!} t e^{\lambda_k t}, \ldots, \frac{1}{(m_k-1)!} t^{m_k-1} e^{\lambda_k t} \right).$$

Demostración. Como en casos particulares anteriores lo esencial es ver que los vectores de la supuesta base son independientes, pues al ser $\dim(\mathbb{S}) = m_1 + \ldots + m_k = n$ (teorema 119), ha de ser sucesión generadora. Lo hacemos por inducción sobre n, siendo trivial si $n = 1$. Supongámoslo cierto para $n-1$. Si tenemos

$$0 = \sum_{i=0}^{m_1-1} \mu^{(1)i} \frac{1}{i!} t^i e^{\lambda_1 t} + \sum_{i=0}^{m_2-1} \mu^{(2)i} \frac{1}{i!} t^i e^{\lambda_2 t} + \ldots + \sum_{i=0}^{m_k-1} \mu^{(k)i} \frac{1}{i!} t^i e^{\lambda_k t},$$

aplicamos $(D-\lambda_1)$ y, con (5.15),

$$0 = \sum_{i=0}^{m_1-1} \mu^{(1)i} \frac{1}{(i-1)!} t^{i-1} e^{\lambda_1 t} + (D-\lambda_1) \left(\sum_{i=0}^{m_2-1} \mu^{(2)i} \frac{1}{i!} t^i e^{\lambda_2 t} \right) + \ldots + (D-\lambda_1) \left(\sum_{i=0}^{m_k-1} \mu^{(k)i} \frac{1}{i!} t^i e^{\lambda_k t} \right). \tag{5.16}$$

En el primer sumando solo están $\frac{1}{0!} e^{\lambda_1 t}, \frac{1}{1!} t e^{\lambda_1 t}, \ldots, \frac{1}{(m_1-2)!} t^{m_1-2} e^{\lambda_1 t}$, que supone $m_1 - 1$ funciones. Por otra parte,

$$\begin{aligned} (D-\lambda_1) \left(\frac{1}{i!} t^i e^{\lambda_h t} \right) &= (D - \lambda_h + \lambda_h - \lambda_1) \left(\frac{1}{i!} t^i e^{\lambda_h t} \right) \\ &= (D-\lambda_h) \left(\frac{1}{i!} t^i e^{\lambda_h t} \right) + (\lambda_h - \lambda_1) \left(\frac{1}{i!} t^i e^{\lambda_h t} \right) \\ &= \left(\frac{1}{(i-1)!} t^{i-1} e^{\lambda_h t} \right) + (\lambda_h - \lambda_1) \left(\frac{1}{i!} t^i e^{\lambda_h t} \right). \end{aligned}$$

Hemos expresado (5.16) en la forma

$$0 = \sum_{i=0}^{m_1-1} \mu^{(1)i} \frac{1}{(i-1)!} t^{i-1} e^{\lambda_1 t} + H(t), \tag{5.17}$$

siendo $H(t)$ una combinación lineal de las funciones de $\mathcal{B}$ *distintas de* $\frac{1}{0!} e^{\lambda_1 t}, \frac{1}{1!} t e^{\lambda_1 t}, \ldots, \frac{1}{(m_1-1)!} t^{m_1-1} e^{\lambda_1 t}$, que son en total $m_2 + m_3 + \cdots + m_k$. Entonces hay en (5.17) $(m_1 - 1) + m_2 + \cdots + m_k = n - 1$ funciones del tipo $\frac{1}{i!} t^i e^{\lambda_h t}$ y podemos aplicar la hipótesis inductiva, que lleva a que los coeficientes de las funciones sean nulos, aunque a nosotros nos basta asegurar que $\mu^{(1)0} = \mu^{(1)1} = \ldots = \mu^{(1)(m_1-1)} = 0$. Se ha simplificado la combinación inicial a

$$0 = \sum_{i=0}^{m_2-1} \mu^{(2)i} \frac{1}{i!} t^i e^{\lambda_2 t} + \ldots + \sum_{i=0}^{m_k-1} \mu^{(k)i} \frac{1}{i!} t^i e^{\lambda_k t},$$

y otra aplicación de la hipótesis inductiva nos da que los restantes $\mu^{(h)i}$ con $h \geq 2$ son también 0. ♣

Este teorema 122 nos dice que cualquier solución de $P(D)(x) = 0$ es una combinación lineal con coeficientes *complejos* de las funciones de $\mathcal{B}$. Puede decirse que hemos resuelto $P(D)(x) = 0$. Sin embargo nos gustaría saber si hay una manera de dar las soluciones por una condición inicial, semejante a lo que pasaba con las ecuaciones $\dot{x} = ax$, donde $x(0) = \xi$ fijaba una y solo una solución. La respuesta es sí, definiendo la **condición inicial** como una sucesión de valores $x^{(0)}(\tau) = \xi^0$, $x^{(1)}(\tau) = \xi^1$, ... , $x^{(n-1)}(\tau) = \xi^{n-1}$ para la solución x.

Teorema 123 *Sea x una solución de $P(D)(x) = 0$ tal que $x^{(0)}(0) = x^{(1)}(0) = \ldots x^{(n-1)}(0) = 0$. Entonces $x = 0$. Como consecuencia tenemos un isomorfismo*

$$\Phi : \mathbb{S} \longrightarrow \mathbb{C}^n, \quad \Phi(x) = \left(x^{(0)}(0), x^{(1)}(0), \ldots, x^{(n-1)}(0)\right)^\top,$$

y a cada $\left(\xi^0, \ldots, \xi^{n-1}\right)^\top \in \mathbb{C}^n$ corresponde una y solo una solución tal que $x^{(i)}(0) = \xi^i$, $i = 0, 1, \ldots, n-1$.

Demostración. Probamos lo primero por inducción sobre n para la solución con los $x^{(j)}(0) = 0$. Si $n = 1$, luego $P(D) = D - \lambda$, la solución general es $x(t) = ke^{\lambda t}$ y si $0 = x(0) = k$ queda $x = 0$. Supongamos cierto el teorema para $n-1$ y lo probamos para n. Factorizamos $P(X) = (X - \lambda_1) Q(X)$ y sea $y = Q(D)(x)$. Ciertamente $(D - \lambda_1)(y) = 0$. Obsérvese que si

$$Q(X) = q_0 + q_1 X + \ldots + q_{n-2} X^{n-2} + X^{n-1},$$

entonces

$$y(0) = q_0 x^{(0)}(0) + q_1 x^{(1)}(0) + \ldots + q_{n-2} x^{(n-2)}(0) + x^{(n-1)}(0) = 0.$$

El teorema para $n = 1$ nos da $y = 0$. Si $y = 0$, $Q(D)(x) = 0$, y el teorema para $n-1$ da $x = 0$.

La segunda parte es inmediata porque Φ es lineal y, además, inyectiva porque $\Phi(x) = 0$ implica que $x = 0$, por el teorema precedente. Al tener los espacios la misma dimensión n, es un isomorfismo. ♣

Hay una pequeña desilusión porque no hay una fórmula fácil con la que, si x se escribe con los coeficientes k en la base $\mathcal{B}$ del teorema 122, nos aparezcan estos k como función de los $x^{(i)}(0)$. Sin embargo supone un alivio saber que el proceso, que involucra resolver sistemas lineales, nunca va a quedar bloqueado porque alguno sea incompatible.

Problema 242 *Para $P(X) = (X-1)^2 (X-2)$, dar la solución con condiciones iniciales $(2, 1, 0)$.* ♦

Solución. La solución general es $x(t) = pe^t + qte^t + re^{2t}$. Se tiene

$$x'(t) = (p+q)e^t + qte^t + 2re^{2t}, \quad x''(t) = (p+2q)e^t + qte^t + 4re^{2t}$$

El sistema a resolver es (bastaría con $(u, v, w) = (2, 1, 0)$ pero vamos a por lo más general)

$$\left\{\begin{array}{l} p + r = u \\ p + q + 2r = v \\ p + 2q + 4r = w \end{array}\right. \quad \text{equivalente a} \quad \left(\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{array}\right)\left(\begin{array}{c} p \\ q \\ r \end{array}\right) = \left(\begin{array}{c} u \\ v \\ w \end{array}\right).$$

La solución general del sistema es

$$\left(\begin{array}{c} p \\ q \\ r \end{array}\right) = \left(\begin{array}{c} 2v - w \\ -2u + 3v - w \\ u - 2v + w \end{array}\right) \text{ y en el caso particular } \left(\begin{array}{c} p \\ q \\ r \end{array}\right) = \left(\begin{array}{c} 2 \\ -1 \\ 0 \end{array}\right).$$

La solución del sistema fijada por $(2, 1, 0)$ es $x(t) = 2e^t - te^t$. ♦

5.15. Ecuaciones lineales no homogéneas

Nos planteamos si $P(D)(x) = y$ tiene solución para $y \in \mathbb{E} = C^\infty(\mathbb{R}, \mathbb{C})$ arbitraria. La existencia de solución es sencilla de probar, pero el cálculo de una concreta puede ser muy laborioso.

Teorema 124 *Dada $P(D)(x) = y$ con ecuación homogénea asociada $P(D)(x) = 0$, se tiene que si $u(t)$ es una solución* concreta *de $P(D)(x) = y$ y $v(t)$ otra de $P(D)(x) = 0$, la suma $x(t) = u(t) + v(t)$ es solución de $P(D)(x) = y$. Al recorrer $v(t)$ el conjunto de soluciones de $P(D)(x) = 0$ resultan, sumándoles $u(t)$,* todas *las soluciones de $P(D)(x) = y$.*[26]

Siempre hay al menos una solución concreta u de $P(D)(x) = y$

Demostración. La linealidad de $P(D)$ nos da $P(D)(u+v) = P(D)(u) + P(D)(v) = y + 0 = y$. Si x es solución arbitraria de $P(D)(x) = y$, tomamos $v = x - u$ y $P(D)(x-u) = P(D)(x) - P(D)(u) = y - y = 0$ y $x = u + v$.

Si factorizamos $P(X)$ como *producto* de monomios $(X - \lambda)$ con λ raíz de $P(X)$ vemos que $P(D)$ es *composición* de funciones $D - \lambda : \mathbb{E} \to \mathbb{E}$. Al ser $D - \lambda$ suprayectiva por el teorema 117, obtenemos que $P(D)$, como función de funciones suprayectivas, es a su vez suprayectiva y para cualquier y hay al menos un x tal que $P(D)(x) = y$. ♣

A la hora de resolver $P(D)(x) = y$ podemos decir que hacerlo con la ecuación homogénea asociada $P(D)(x) = 0$ es cosa fácil usando el teorema 122. Hallar una solución concreta u de $P(D)(x) = y$ es bastante asequible si y tiene la forma restringida $y(t) = Q(t)e^{t\mu}$ siendo $Q(X)$ un polinomio sin límite de grado. Supondremos en adelante que $P(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_p)^{m_p}$, permitiéndose que μ no sea uno de los λ_i. Si fijado y determinamos soluciones u_i en cadena de las p ecuaciones

$$(D - \lambda_1)^{m_1}(u_1) = u_0 = y,\ (D - \lambda_2)^{m_2}(u_2) = u_1,\ (D - \lambda_3)^{m_3}(u_3) = u_2,\ \ldots,(D - \lambda_3)^{m_p}(u_p) = u_{p-1}$$

se cumple que $u_p = u$ sirve como solución de $P(D)(x) = y$ ya que

$$\begin{aligned}(D - \lambda_1)^{m_1} \circ \cdots \circ (D - \lambda_p)^{m_p}(u_p) &= (D - \lambda_1)^{m_1} \circ \cdots \circ (D - \lambda_{p-1})^{m_{p-1}}(u_{p-1}) \\ &= (D - \lambda_1)^{m_1} \circ \cdots \circ (D - \lambda_{p-2})^{m_{p-2}}(u_{p-2}) \\ &= (D - \lambda_1)^{m_1} \circ \cdots \circ (D - \lambda_{p-3})^{m_{p-3}}(u_{p-3}) \\ &= \ldots = (D - \lambda_1)^{m_1}(u_{p-1}) = u_0 = y.\end{aligned}$$

Se simplifica el problema si $y(t) = Q(t)e^{t\mu}$. Si fuera μ uno de los λ_1, se tiene renumerando que es $\mu = \lambda_1$. Entonces resolvemos $(D - \lambda_1)^{m_1}(u_1) = u_0 = y$ y quedan pendientes las restantes $(D - \lambda_j)^{m_j}(u_j) = u_{j-1}$. Si μ no es ningún λ_i está incluso pendiente $(D - \lambda_1)^{m_1}(u_1) = u_0 = y$. No es difícil el caso $(D - \alpha)^k(u) = Q(t)e^{t\alpha}$, que se resuelve con (5.3), generalizable a

$$(D - \alpha)^k\left(\left(\frac{t^h}{h!}e^{\alpha t}\right)\right) = \frac{t^{h-k}}{(h-k)!}e^{\alpha t}.$$(se entiende que si $k > h$ el valor es cero).

Por tanto, si $Q(X) = q_0 + q_1 X + \ldots + q_{m-1}X^{m-1} + X^m$, entonces,

$$(D - \alpha)^k\left(q_0\frac{t^k}{k!}e^{\alpha t} + q_1\frac{t^{k+1}}{(k+1)!}e^{\alpha t} + \ldots + q_{m-1}\frac{t^{k+(m-1)}}{(k+(m-1))!}e^{\alpha t} + \frac{t^{k+m}}{(k+m)!}e^{\alpha t}\right) = Q(X)e^{\alpha t}.$$

Tiene más dificultad encontrar soluciones concretas de $(D - \lambda)(x) = Q(t)e^{t\mu}$ con $\lambda \neq \mu$. Primero una fórmula técnica importante. Sea $E^\alpha : \mathbb{E} \to \mathbb{E}$ la función lineal que a la *función* x le asigna otra *función* $E^\alpha(x)$ definida por $E^\alpha(x)(t) = e^{\alpha t}x(t)$. Se tiene que

$$(D - \alpha)^n = E^\alpha \circ D^n \circ E^{-\alpha}. \tag{5.18}$$

Esto es fácil porque para $n = 1$, $D - \alpha = E^\alpha \circ D \circ E^{-\alpha}$ puesto que

$$E^\alpha \circ D \circ E^{-\alpha}(x) = e^{\alpha t}\left(e^{-\alpha t}x\right)' = e^{\alpha t}\left(-\alpha e^{-\alpha t}x + e^{-\alpha t}x'\right) = -\alpha x + x' = (D - \alpha)(x),$$

y, para n arbitrario,

$$(D - \alpha)^n = \left(E^\alpha \circ D \circ E^{-\alpha}\right) \circ \cdots \circ \left(E^\alpha \circ D \circ E^{-\alpha}\right) = E^\alpha \circ D^n \circ E^{-\alpha}.$$

Definimos unos subespacios de $\mathbb{E} = C^\infty(\mathbb{R}, \mathbb{C})$ (luego sus elementos son *funciones*)

$$\mathbb{P}(k, \beta) = \left\{x \in \mathbb{E} = C^\infty(\mathbb{R}, \mathbb{C}) \mid x(t) = Q(t)e^{\beta t},\ Q(X) \text{ polinomio de grado } \leq k\right\}.$$

[26] Puede decirse que si las soluciones de $P(D)(x) = 0$ constituían un subespacio vectorial $\mathbb{S}$ de $\mathbb{E} = C^\infty(\mathbb{R}.\mathbb{C})$, las soluciones de $P(D)(x) = y$ constituyen un subespacio afín de $\mathbb{E}$, que es $u + \mathbb{S}$ siendo u una solución concreta de $P(D)(x) = y$.

1. Para cualquier elección de α y β, $D-\alpha$ lleva $\mathbb{P}(k,\beta)$ en $\mathbb{P}(k,\beta)$.
2. Si $\alpha=\beta$, $(D-\alpha)^{k+1}(\mathbb{P}(k,\alpha))=0$.

Lo primero se sigue de (5.11) porque

$$(D-\alpha)\left(Q(t)\,e^{\beta t}\right)=Q'(t)\,e^{\beta t}+Q(t)\,\beta e^{\beta t}-\alpha Q(t)\,e^{\beta t}=\left(Q'(t)+(\beta-\alpha)\,Q(t)\right)e^{\beta t}$$

y $\deg\left(Q'(X)+(\beta-\alpha)\,Q(X)\right)\leq\deg(Q(X))\leq k$. Para lo otro, si $\deg(Q(X))\leq k$, utilizando (5.18),

$$(D-\beta)^{k+1}\left(Q(t)\,e^{\beta t}\right)=\left(E^{\beta}\circ D^{k+1}\circ E^{-\beta}\right)\left(Q(t)\,e^{\beta t}\right)=\left(E^{\beta}\circ D^{k+1}\right)(Q(t))=E^{\beta}(0)=0,$$

ya que $k<n$ implica que $D^{k+1}(Q(t))=0$. Sin embargo, lo más importante está en el teorema 125, que usa otro más general (teorema 126).

Desde aquí ya se ve el fin del túnel. Si tenemos $P(D)(x)=Q(t)\,e^{\beta t}$ siendo $Q(t)\,e^{\beta t}\in\mathbb{P}(k,\beta)$ y μ no es ninguna de las raíces λ_i de $P(X)$, vamos a probar que $P(D)$ es un automorfismo de $\mathbb{P}(k,\beta)$ (¡ojo!, no de $\mathbb{E}$) luego tendremos una solución $u(t)=R(t)\,e^{\beta t}\in\mathbb{P}(k,\beta)$ tal que $P(D)\left(R(t)\,e^{\beta t}\right)=Q(t)\,e^{\beta t}$; es decir, *una solución concreta de* $P(D)(x)=y$. A estas alturas el lector ya sabrá la diferencia entre probar que $P(D)$ es un automorfismo y calcular su inverso, que es lo que se precisa para conocer $R(t)\,e^{\beta t}$. Veremos que, dentro de lo que son estas cosas, es un cálculo asequible.[27]

Teorema 125 *$D-\lambda:\mathbb{P}(k,\mu)\to\mathbb{P}(k,\mu)$ es un isomorfismo si $\lambda\neq\mu$. (La fórmula explícita de $(D-\lambda)^{-1}$ es (5.19) más abajo.)*

Teorema 126 *Sea $L:\mathbb{E}\to\mathbb{E}$ un endomorfismo tal que $L^p=0$. Entonces $\mathrm{id}_{\mathbb{E}}-L$ es un endomorfismo invertible*[28] *siendo*

$$\left(\mathrm{id}_{\mathbb{E}}-L\right)^{-1}=\mathrm{id}_{\mathbb{E}}+L+L^2+\ldots+L^{p-1}.$$

Con más generalidad, para $\mu\in\Bbbk$ no nulo,

$$\left(\mu\,\mathrm{id}_{\mathbb{E}}-L\right)^{-1}=\frac{1}{\mu}\left(\mathrm{id}_{\mathbb{E}}+\frac{L}{\mu}+\frac{L^2}{\mu^2}+\ldots+\frac{L^{p-1}}{(\mu-1)^{p-1}}\right)=\frac{1}{\mu}\sum_{i=0}^{p-1}\left(\frac{L}{\mu}\right)^{i}.$$

Demostración. Tenemos que

$$\begin{aligned}\left(\mathrm{id}_{\mathbb{E}}-L\right)\left(\mathrm{id}_{\mathbb{E}}+L+L^2+\ldots+L^{p-1}\right)&=\left(\mathrm{id}_{\mathbb{E}}+L+L^2+\ldots+L^{p-1}\right)-\left(L+L^2+\ldots+L^{p-1}+L^p\right)\\&=\mathrm{id}_{\mathbb{E}}-L^p=\mathrm{id}_{\mathbb{E}},\end{aligned}$$

ya que $L^p=0$. Para el caso general observamos que $(1/\mu)\,L$ también cumple $((1/\mu)\,L)^p=0$. Por tanto,

$$\left(\mu\,\mathrm{id}_{\mathbb{E}}-L\right)^{-1}=\left(\mu\left(\mathrm{id}_{\mathbb{E}}-\frac{L}{\mu}\right)\right)^{-1}=\frac{1}{\mu}\left(\mathrm{id}_{\mathbb{E}}-\frac{L}{\mu}\right)^{-1},$$

y la fórmula general sale del caso particular anterior. ♣

La fórmula del teorema 126 es particularmente útil cuando nos interesa, no ya $(1-L)^{-1}$, sino una potencia de $(1-L)^{-1}$. Si $L^3=0$, y queremos $(1-L)^{-6}$, utilizamos $(1-L)^{-1}=1+L+L^2$ y

$$(1-L)^{-2}=\left(1+L+L^2\right)^2=L^4+2L^3+3L^2+2L+1=3L^2+2L+1,$$

ya que $L^3=L^4=0$. Seguimos ahora con

$$(1-L)^{-6}=\left(3L^2+2L+1\right)^3=27L^6+54L^5+63L^4+44L^3+21L^2+6L+1=21L^2+6L+1,$$

(porque $L^q=0$ si $q\geq 3$) que es el valor buscado. Lo ventajoso es no necesitar las fórmulas explícitas con los L^q para todo q, sino basta hacer las cuentas ignorando los términos con $q\geq 3$.

Demostración. (del teorema 125) Sabemos que $D-\beta:\mathbb{P}(k,\beta)\to\mathbb{P}(k,\beta)$ cumple $(D-\beta)^{k+1}=0$. Con $\mathbb{E}=\mathbb{P}(k,\beta)$ y $L=(D-\beta)$ en el teorema 126 tenemos

$$(D-\lambda)^{-1}=-\left((\lambda-\beta)-(D-\beta)\right)^{-1}=\frac{-1}{\lambda-\beta}\left(1+\frac{D-\beta}{\lambda-\beta}+\cdots+\frac{(D-\beta)^k}{(\lambda-\beta)^k}\right),\tag{5.19}$$

aunque esta inversa es *la inversa de la restricción a* $\mathbb{P}(k,\beta)$. (No es la inversa de $D-\lambda:\mathbb{E}\to\mathbb{E}$.) ♣

[27] Damos un primer procedimiento más atractivo desde el punto de vista teórico, pero para el puro cálculo puede ser mejor el que se describe en el párrafo final de la sección.

[28] Ponemos $\mathrm{id}_{\mathbb{E}}-L$, pero como $\lambda\in\Bbbk$ se identifica con $\lambda\,\mathrm{id}_{\mathbb{E}}$, aparecerá $\lambda-L$ en los cálculos prácticos.

Problema 243 *Dar todas las soluciones de* $\left(\left(D-1\right)^2 D\right)(x) = t^3$. ♦

Solución. De acuerdo con el teorema 122 la solución general de la ecuación homogénea asociada es $x(t) = \sigma e^t + \theta t e^t + \varphi$. Vamos a buscar una solución concreta de $\left(\left(D-1\right)^2 D\right)(x) = t^3$. Según se explicó había que resolver en cadena

$$\left(D-1\right)^2(u_1) = u_0 = t^3 \quad \text{y} \quad (D-0)(u_2) = u_1,$$

siendo $u_2 = u$ la solución concreta buscada. Como $y(t) = t^3 \in \mathbb{P}(3,0)$ se sabe que $(D-1)^{3+1}$ es nula sobre $\mathbb{P}(3,0)$ y, por el teorema 125, donde $\lambda = 1$, $\mu = 0$ y $k = 3$,

$$\begin{aligned}(D-1)^{-1} &= -\left((1-\beta)-(D-0)\right)^{-1} = -\left(1+D+D^2+D^3\right),\\ (D-1)^{-2} &= \left(-\left(1+D+D^2+D^3\right)\right)^2 = 4D^3+3D^2+2D+1,\end{aligned}$$

ya que sobre $\mathbb{P}(3,0)$ (¡pero no sobre $\mathbb{E}$!) se tiene $D^h = 0$ para $h \geq 4$. De acuerdo con esto

$$\begin{aligned}u_1(t) &= \left(4D^3+3D^2+2D+1\right)\left(t^3\right) = 24+18t+6t^2+t^3,\\ u_2(t) &= \frac{1}{4}t^4+2t^3+9t^2+24t.\end{aligned}$$

La solución general es

$$x(t) = x(t) = \sigma e^t + \theta t e^t + \varphi + \frac{1}{4}t^4+2t^3+9t^2+24t. \blacklozenge$$

Problema 244 *Dar todas las soluciones de* $x''(t) - x(t) = 2te^t$.

Se puede obtener una solución concreta de $P(D)(x) = y$ como hemos detallado con las inversas de las $(D-\lambda)^k$. *Pero hay un camino alternativo una vez que uno sabe que la solución existe* (quizás más agradable. Se puede escribir con coeficientes indeterminados $u(t) = \left(q_0 + q_1 t + q_2 t^2 + \ldots + q_k t^k\right) e^{\mu t}$ y con el sistema lineal que aparece en $P(D)(u) = y$, calcular los q_j.

Problema 245 *Determinar una solución concreta de* $\left(D-1\right)^2(x) = t^2$. *Ídem para el problema 244.*

Solución. Será de la forma $u(t) = p + qt + rt^2$. Todo es cuestión de operar

$$\begin{aligned}&\tfrac{d}{dt}\left(p+qt+rt^2\right) - \left(p+qt+rt^2\right) = -rt^2 - t(q-2r) + q - p\\ &\tfrac{d}{dt}\left(-rt^2-t(q-2r)+q-p\right) - \left(-rt^2-t(q-2r)+q-p\right) = rt^2+(q-4r)t+p-2q+2r\end{aligned}$$

y debe resolverse $rt^2 + t(q-4r) + p - 2q + 2r = t^2$, lo que da $r = 1$, $q = 4$ y $p = 6$.

Es $u(t) = 6 + 4t + t^2$ la solución. El otro caso queda para el lector. ♦

Capítulo 6

Potencias y exponencial de endomorfismos

Sea $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo del espacio vectorial $\mathbb{E}$ sobre el cuerpo $\Bbbk$ y de dimensión n. *Supondremos salvo aviso contrario que su polinomio característico $C(X)$ es linealmente factorizable*; digamos que $C(X) = (-1)^n (X - \lambda_1)^{m_1} \cdots (X - \lambda_p)^{m_p}$.[1] Interesa poder calcular del modo más efectivo posible las potencias $L^k = L \circ \cdots \circ L$ (k veces) para k arbitrario.[2] El teorema esencial es este

Teorema 127 *Si $C(X)$ es linealmente factorizable, se puede descomponer L de modo único en la forma $L = D + N$ siendo D un endomorfismo diagonalizable y N un endomorfismo nilpotente cumpliéndose además que $D \circ N = N \circ D$.*

Se dice que $N : \mathbb{E} \to \mathbb{E}$ es **nilpotente** si existe un $\ell \in \mathbb{N}$ tal que $N^\ell = 0$ pero $N^{\ell-1} \neq 0$. A ℓ se le llama el **índice de nilpotencia** (de N). Con vistas a nuestro problema de calcular L^k usamos el binomio de Newton, *que puede hacerse porque $D \circ N = N \circ D$*, con lo que

$$L^k = (D + N)^k = \sum_{i=0}^{k} \binom{k}{i} D^i N^{k-i}.$$

Esta fórmula es manejable porque solo deberán conocerse $N^1, N^2, \ldots, N^{\ell-1}$ y, al ser D diagonalizable, se calculan bien sus potencias. Veremos más adelante con detalle cómo disponer los cálculos. Hay algo que no aparece en el enunciado del teorema 127 pero es de gran importancia práctica y es que D y N son expresiones polinomiales de L con unos polinomios calculables en función de cómo sea la factorización de $C(X)$. Por supuesto que es difícil factorizar con precisión $C(X)$, pues lo usual es tener solo valores aproximados para los λ_j, pero esta imprecisión es medible, y como consecuencia la de D y N. No trataremos esta parte del problema, suponiendo siempre, y eso sucederá en los ejemplos, que $C(X)$ factorizado puede conocerse con absoluta exactitud.

6.1. Sobre polinomios

Sea $P(X) = (X - \lambda_1)^{m_1} \cdots (X - \lambda_p)^{m_p}$ un polinomio factorizable con factores lineales, con los λ_i *distintos entre sí.* Definimos también

$$\phi_i(X) = \frac{P(X)}{(X - \lambda_i)^{m_i}} = (X - \lambda_1)^{m_1} \cdots \widehat{(X - \lambda_i)^{m_i}} \cdots (X - \lambda_p)^{m_p}.$$

Teorema 128 *Supuesto $p \geq 2$, hay polinomios $Q_1(X), \ldots, Q_p(X)$ de forma que*

$$1 = Q_1(X)\,\phi_1(X) + \ldots + Q_p(X)\,\phi_p(X), \quad \deg Q_i(X) < m_i,\ i = 1, \ldots, p.$$

[1] Es fácil que el lector olvide esto y hay que insistir en que los teoremas no valen en general para cualquier endomorfismo sino solo para los que tienen $C(X)$ (o $M(X)$ que lo mismo da) linealmente factorizable. Si $\Bbbk = \mathbb{C}$, siempre pasa esto y L puede ser cualquier endomorfismo.

[2] La razón es que la **exponencial** de L dada por la serie $\sum_{k=0}^{\infty} \frac{L^k}{k!}$ es sumamente importante para la solución de sistemas lineales con coeficientes constantes.

Demostración. Consideramos el conjunto $\mathcal{P}$ de polinomios de la forma $\psi_1(X)\phi_1(X)+\ldots+\psi_p(X)\phi_p(X)$ y tomamos en el uno *mónico y de grado mínimo que llamaremos* $D(X)$. Afirmamos que $D(X)$ divide a todos los $\phi_i(X)$. Si fuera esto falso utilizamos la división euclidiana y, suponiendo para simplificar que $D(X)$ no divide a $\phi_1(X)$, disponemos de $\phi_1(X)=A(X)D(X)+R(X)$ con $R(X)\neq 0$ y $\deg R(X)<\deg D(X)$. Entonces,

$$\begin{aligned}R(X)&=\phi_1(X)-A(X)D(X)=\phi_1(X)-A(X)\left[\psi_1(X)\phi_1(X)+\ldots+\psi_p(X)\phi_p(X)\right]\\&=(1-A(X))\psi_1(X)\phi_1(X)+A(X)\psi_2(X)\phi_2(X)+\ldots+A(X)\psi_p(X)\phi_p(X),\end{aligned}$$

lo que muestra que $R(X)\in\mathcal{P}$. Dividiendo por el coeficiente principal de $R(X)$ obtenemos un polinomio mónico en $\mathcal{P}$ de grado estrictamente menor que el de $D(X)$. Contradicción.

Afirmamos por último que $D(X)=1$ y habremos concluido, En efecto, como $D(X)$ divide a $\phi_1(X)$, ha de ser de la forma $D(X)=(X-\lambda_1)^{h_2}\cdots(X-\lambda_p)^{h_p}$, con las $h_j\leq m_j$, pero si algún h_j no fuese 0 sería imposible que $D(X)$ dividiese a $\phi_j(X)$. Por consiguiente, $D(X)=1$. ♣

Vamos a ver cómo se puede hacer el cálculo explícito. Empezamos con el caso muy sencillo $P(X)=(X-\alpha)(X-\beta)$, $\alpha\neq\beta$. Debemos determinar $p=Q_1(X)$ y $q=Q_2(X)$ en

$$1=p(X-\beta)+q(X-\alpha)=(p+q)X-q\alpha-p\beta.$$

Esto lleva al sistema con incógnitas $p,q\in\Bbbk$,

$$\left\{\begin{array}{c}p+q=0\\ \beta p+\alpha q=-1\end{array}\right. \text{ y como } \left(\begin{array}{cc|c}1&1&0\\ \beta&\alpha&-1\end{array}\right)\to\left(\begin{array}{cc|c}1&1&0\\ 0&\alpha-\beta&-1\end{array}\right)\to\left(\begin{array}{cc|c}1&1&0\\ 0&1&\frac{1}{\beta-\alpha}\end{array}\right)$$

resultan las soluciones $p=\frac{1}{\alpha-\beta}$, $q=-\frac{1}{\alpha-\beta}$. Efectivamente,

$$1=\frac{X-\alpha}{\alpha-\beta}+\frac{X-\beta}{\beta-\alpha}.$$

De hecho, una verificación directa, *sin pasar por sistemas lineales*, generaliza esto muchísimo para cuando todos los m_i son 1, porque si $P(X)=(X-\lambda_1)\cdots(X-\lambda_p)$,

$$1=\sum_{i=1}^{p}\frac{1}{(\lambda_i-\lambda_1)\cdots\widehat{(\lambda_i-\lambda_i)}\cdots(\lambda_i-\lambda_p)}(X-\lambda_1)\cdots\widehat{(X-\lambda_i)}\cdots(X-\lambda_p)\tag{6.1}$$

y por tanto los $Q_i(X)$ de grado <1 son $\left[(\lambda_i-\lambda_1)\cdots\widehat{(\lambda_i-\lambda_i)}\cdots(\lambda_i-\lambda_p)\right]^{-1}$. Recordamos que esta fórmula se vio en *Autovalores y autovectores*, siendo la allí idea de la demostración que la diferencia de los dos lados era un polinomio de grado $p-1$ con p raíces $\lambda_1,\ldots,\lambda_p$, luego debe ser el polinomio nulo.

Para cálculos explícitos hay una *simplificación esencial*: escribir los polinomios $Q_i(X)$ como polinomios en potencias de $(X-\lambda_i)$. Vamos a dar más detalle en el siguiente problema.

Problema 246 *Hacer la descomposición del teorema 128 para* $(X-\alpha)^2(X-\beta)$. ♦

Solución. Han de calcularse p,q,r para que sea

$$1=N(X)=Q_1(X)(X-\beta)+Q_2(X)(X-\alpha)^2=(p(X-\alpha)+q)(X-\beta)+r(X-\alpha)^2$$

Un cálculo sistemático hace primero las sustituciones

$$\left\{\begin{array}{l}1=N(\alpha)=q(\alpha-\beta)=-qd\\ 1=N(\beta)=r(\beta-\alpha)^2=rd^2\end{array}\right. \text{ luego } q=\frac{-1}{d},\quad r=\frac{1}{d^2}.$$

Queda calcular p. Aunque pueden seguirse varios métodos (por ejemplo, trabajar con $1=N(0)$) lo mejor es derivar.[3] Tenemos

$$0=N'(X)=p(X-\beta)+(p(X-\alpha)+q)+2r(X-\alpha),$$

y entonces $0=N'(\alpha)=-pd+q$. En definitiva (y con esto la descomposición es inmediata)

$$p=\frac{q}{d}=\frac{-1}{d^2},\quad q=\frac{-1}{d},\quad r=\frac{1}{d^2}\quad\text{y}\quad 1=-\left(\frac{1}{d^2}(X-\alpha)+\frac{1}{d}\right)(X-\beta)+\frac{1}{d^2}(X-\alpha)^2.\ \blacklozenge$$

[3] Aunque sea $\Bbbk\neq\mathbb{R}$, sirve la fórmula de la derivada de un polinomio y las fórmulas bien conocidas para derivar sumas y productos.

Problema 247 *Ídem para* $(X-\alpha)^2(X-\beta)^2$.

Problema 248 *Ídem para* $(X-\alpha)(X-\beta)(X-\gamma)^2$ *con* α,β,γ *distintos entre sí.*

Problema 249 *Ídem para* $(X-\alpha)^3(X-\beta)$. *Si quiere el lector puede elegir* $(X-2)^3(X-1)$ *pero no creemos que suponga una gran ventaja conceptual.*

Si el lector ha resuelto este problema y otro anterior habrá visto que para $k=2,3$ al descomponer $1/(X-\alpha)^k(X-\beta)$ se obtiene para $d=\beta-\alpha$,

$$\begin{cases} 1=-\left(\frac{1}{d^2}(X-\alpha)+\frac{1}{d}\right)(X-\beta)+\frac{1}{d^2}(X-\alpha)^2 & \text{si } k=2 \\ 1=-\left[\frac{1}{d^3}(X-\alpha)^2+\frac{1}{d^2}(X-\alpha)+\frac{1}{d}\right](X-\beta)+\frac{1}{d^3}(X-\alpha)^3 & \text{si } k=3 \end{cases}$$

lo que parece sugerir una fórmula para k arbitrario.

Teorema 129 *La descomposición del teorema 128 para* $(X-\alpha)^k(X-\beta)$ *con* $\alpha\neq\beta$ *es (6.2) más abajo.*

Demostración. Hay que determinar $P(X)$ de grado $k-1$ y $s\in\mathbb{R}$ para que sea

$$1=P(X)(X-\beta)+s(X-\alpha)^k=\left(\sum_{j=0}^{k-1}p_j(X-\alpha)^j\right)(X-\beta)+s(X-\alpha)^k.$$

Si $\beta=0$ queda $s=1/d^k$ siendo $d=\beta-\alpha$. La pregunta es si podremos calcular $P(X)$ de modo que

$$\left(\frac{X-\alpha}{d}\right)^k-1=-P(X)(X-\beta).$$

Sí, se puede hacer. Usamos la identidad $y^k-1=\left(\sum_{j=0}^{k-1}y^j\right)(y-1)$ (es la clave, si bien es algo bastante estándar) con lo que

$$\begin{aligned}\left(\frac{X-\alpha}{d}\right)^k-1&=\left(\sum_{j=0}^{k-1}y^j\right)(y-1)=\left(\sum_{j=0}^{k-1}\left(\frac{X-\alpha}{d}\right)^j\right)\left(\frac{X-\alpha}{\beta-\alpha}-1\right)\\&=\left(\sum_{j=0}^{k-1}\left(\frac{X-\alpha}{d}\right)^j\right)\left(\frac{X-\beta}{\beta-\alpha}\right)=\sum_{j=0}^{k-1}\frac{(X-\alpha)^j}{d^{j+1}}.\end{aligned}$$

Con este cálculo se tiene la fórmula deseada con $d=\beta-\alpha$,

$$1=\left(\frac{X-\alpha}{d}\right)^k-\left(\sum_{j=0}^{k-1}\frac{(X-\alpha)^j}{d^{j+1}}\right)(X-\beta)\quad\text{y}\quad P(X)=-\sum_{j=0}^{k-1}\frac{(X-\alpha)^j}{d^{j+1}}. \tag{6.2}$$

♣

6.2. Expresión de un endomorfismo con proyecciones

Recordamos que si $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$, la proyección P sobre $\mathbb{F}$ a lo largo de $\mathbb{G}$ se define con $P(x)=y$, supuesto que $x=y+z$ con $y\in\mathbb{F}$ y $z\in\mathbb{G}$. Esto se generaliza a $\mathbb{E}=\mathbb{F}_1\oplus\cdots\oplus\mathbb{F}_p$ definiendo

$$P_i:\mathbb{E}\to\mathbb{E},\qquad P_i(x)=x_i \text{ si } x=x_1+\ldots+x_p \text{ con } x_j\in\mathbb{F}_j.$$

Cada P_i proyecta[4] sobre $\mathbb{F}_i$ a lo largo de $\mathbb{F}_1\oplus\cdots\oplus\mathbb{F}_{i-1}\oplus\mathbb{F}_{i+1}\oplus\ldots\oplus\mathbb{F}_p$. Es sencillo comprobar que

$$P_i\circ P_j=0 \text{ si } i\neq j\qquad\text{y}\qquad \mathrm{id}_{\mathbb{E}}=P_1+\ldots+P_p. \tag{6.3}$$

[4] Surge la duda de si P va de E en $\mathbb{F}_i$ o de $\mathbb{E}$ en $\mathbb{E}$. Optamos por la segunda opción sin olvidar que $\mathrm{im}(P_i)=\mathbb{F}_i\subset\mathbb{E}$.

La identidad es entonces suma de p endomorfismos P_i. Advertimos que las condiciones (6.3) para endomorfismos P_j, *que no se sabe previamente que sean proyecciones*, implican que necesariamente han de serlo porque

$$P_j = P_j \circ \mathrm{id} = P_j \circ (P_1 + \ldots + P_p) = \sum_{i=1}^{p} P_j \circ P_i = P_j \circ P_j.$$

Siendo $P_j = P_j \circ P_j$ y se vio en el problema 114 que si L cumple $L^2 = L$ entonces L es la proyección asociada a $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ con $\mathbb{F} = \mathrm{im}\,(L)$ y $\mathbb{G} = \ker(L)$.

Podemos asociar a la descomposición $\mathbb{E} = \mathbb{F}_1 \oplus \cdots \oplus \mathbb{F}_p$ la sucesión de proyecciones $(P_1, \ldots, P_p)$ cumpliendo (6.3). Y recíprocamente, a una sucesión de endomorfismos $(P_1, \ldots, P_p)$ cumpliendo (6.3), que por lo comentado han de ser proyecciones, una descomposición $\mathbb{E} = \mathrm{im}\,(P_1) \oplus \ldots \oplus \mathrm{im}\,(P_p)$.

Problema 250 *Probar que* $\mathbb{E} = \mathrm{im}\,(P_1) \oplus \ldots \oplus \mathrm{im}\,(P_p)$ *para* $(P_1, \ldots, P_p)$ *cumpliendo (6.3).*

Diremos que $(P_1, \ldots, P_p)$ es una **familia completa de proyecciones** si se cumple (6.3), aunque las condiciones (6.3) bastan para asegurar que los P_j son proyecciones; o sea, $P_j^2 = P_j$ para todo j. Hasta aquí no hemos hecho referencia a ningún endomorfismo L de $\mathbb{E}$ pero cuando se tiene una de estas familias y una sucesión $\lambda = (\lambda_1, \ldots, \lambda_p)$ en $\Bbbk$ podemos considerar un **endomorfismo asociado a la familia de proyecciones con coeficientes** $(\lambda_1, \ldots, \lambda_p)$ que se define por $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$. Obviamente, $\lambda = (1, 1, \ldots, 1)$ da $L = \mathrm{id}$. ¿Cuáles son los endomorfismos L expresables de esta forma?

Teorema 130 *Los endomorfismos expresables como* $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$*, con* $(P_1, \ldots, P_p)$ *familia completa de proyecciones, son justamente los endomorfismos diagonalizables. El que sea* $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$ *nos informa de que los valores propios son* $(\lambda_1, \ldots, \lambda_p)$ *y la multiplicidad geométrica o algebraica de* λ_j *(que es la misma) es la dimensión de* $\mathrm{im}\,(P_j)$.

Demostración. Si L es diagonalizable y $\mathbb{F}_1, \ldots, \mathbb{F}_p$ sus subespacios propios, con $\mathbb{F}_j$ el subespacio propio de λ_j, comparamos L con $M = \lambda_1 P_1 + \ldots + \lambda_p P_p$ y veremos que son iguales. Como $\mathbb{E} = \mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_p$ basta mostrar que si $x_j \in \mathbb{F}_j$ es $L(x_j) = M(x_j)$. Al ser x_j vector propio de λ_j, $L(x_j) = \lambda_j x_j$. Por otra parte

$$M(x_j) = (\lambda_1 P_1 + \ldots + \lambda_p P_p)(x_j) = \lambda_1 P_1(x_j) + \ldots + \lambda_p P_p(x_j) = \lambda_j P_j(x_j) = \lambda_j x_j.$$

Recíprocamente, si $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$ y tomamos una base $\mathcal{U}$ obtenida yuxtaponiendo bases $\mathcal{U}_j = (u_{j1}, \ldots u_{jm_j})$ de cada $\mathbb{F}_j$, se obtiene como antes que $L(u_{jk}) = \lambda_j u_{jk}$. Así se ve que la matriz de L es diagonal con $\lambda_1, \ldots, \lambda_p$ en ella y tantas veces repetido λ_j cuanta sea la dimensión m_j de $\mathbb{F}_j$. Ya tenemos que L es diagonalizable y, vista la matriz de L en la base $\mathcal{U}$, es inmediato que estos m_j son las multiplicidades geométricas y algebraicas de los λ_j. ♣

La utilidad de este teorema se pone de manifiesto en la sección siguiente.

6.3. La descomposición $L = D + N$

El lector debe repasar la sección *El polinomio minimal* del capítulo *Autovalores y autovectores.* Supondremos en adelante que L tiene polinomios mínimo y característico linealmente factorizables

$$M(X) = (X - \lambda_1)^{s_1} \cdots (X - \lambda_p)^{s_p}, \qquad C(X) = (-1)^n (X - \lambda_1)^{m_1} \cdots (X - \lambda_p)^{m_p}.$$

Ambos polinomios tienen las mismas raíces $\lambda_1, \ldots, \lambda_p$ y debe ser $1 \leq s_j \leq m_j$ para $j = 1, \ldots, p$ por el teorema 107.

El resultado esencial es que L determina polinomios $H_1(X), \ldots, H_p(X)$ de modo que las funciones $P_i = H_i(L) : \mathbb{E} \to \mathbb{E}$ forman una familia completa de proyecciones. Cuando L sea diagonalizable con valores propios $\lambda_1, \ldots, \lambda_p$, las P_i serán las proyecciones sobre los subespacios propios $\mathbb{E}(\lambda_i)$ luego, en virtud del teorema 130, tendremos $L = \sum_{i=1}^{p} \lambda_i P_i = \sum_{i=1}^{p} \lambda_i H_i(L)$. Si L no es diagonalizable, puede en todo caso definirse $D = \sum_{i=1}^{p} \lambda_i P_i$, que es diagonalizable (otra vez por el teorema 130) pero obviamente distinto de L. En este caso, se probará que $N = L - D$ es nilpotente y conmuta con D, quedando probado el teorema 127, pendiente la unicidad.

Aplicamos el teorema 128 a $M(X)$, siendo entonces

$$\phi_i(X) = \frac{M(X)}{(X-\lambda_{i1})^{s_i}} = (X-\lambda_1)^{s_1} \cdots \widehat{(X-\lambda_1)^{s_i}} \cdots (X-\lambda_p)^{s_p}.$$

Existen polinomios $Q_1(X), \ldots, Q_p(X)$ de grados estrictamente menores a $s_1, \ldots, s_p$ tales que

$$1 = \sum_{i=1}^{p} \phi_i(X) Q_i(X) = \sum_{i=1}^{p} (X-\lambda_1)^{s_1} \cdots \widehat{(X-\lambda_1)^{s_i}} \cdots (X-\lambda_p)^{s_p} Q_i(X). \tag{6.4}$$

Los polinomios $H_i(X) = (X-\lambda_1)^{s_1} \cdots \widehat{(X-\lambda_1)^{s_i}} \cdots (X-\lambda_p)^{s_p} Q_i(X)$ son los anunciados. Cumplen

$$\begin{aligned}
&\textbf{(a)}\ 1 = H_i(X) + \ldots + H_p(X)\\
&\textbf{(b)}\ H_i(X) H_j(X) = M(X) K_{ij}(X) \text{ para } i \neq j\\
&\textbf{(c)}\ (X-\lambda_i)^{s_i} H_i(X) = M(X) Q_i(X)
\end{aligned}$$

sin que necesitemos precisar los polinomios $K_{ij}(X)$. Al sustituir X por L y ser $M(L) = 0$ (¡clave!),

$$\begin{aligned}
&\textbf{(a)}\ 1 = H_i(L) + \ldots + H_p(L)\\
&\textbf{(b)}\ H_i(L) \circ H_j(L) = 0\\
&\textbf{(c)}\ (L-\lambda_i)^{s_i} \circ H_i(L) = 0
\end{aligned} \tag{6.5}$$

Obsérvese que los $H_i(L)$, que son endomorfismos de $\mathbb{E}$, cumplen en virtud de **(a)** y **(b)** en (6.5) que son una familia completa de proyecciones (ver (6.3)). Damos más detalles en el siguiente teorema.

Teorema 131 *Si para el polinomio minimal $M(X)$ de L se tiene la descomposición (6.4), hay una familia completa de proyecciones*

$$P_i : \mathbb{E} \to \mathbb{E}, \quad P_i = H_i(L) = \phi_i(L) \circ Q_i(L), \quad i = 1, \ldots, p.$$

Los subespacios $\mathbb{G}_i = \operatorname{im}(P_i)$ de la suma directa asociada son los núcleos de $(L-\lambda_i)^{s_i}$.

Demostración. Ya hemos visto que las $P_i = H_i(L)$ forman una familia completa de proyecciones. La condición $(L-\lambda_i)^{s_i} \circ H_i(L) = 0$ de **(c)** en (6.5) nos dice que $\mathbb{G}_i = \operatorname{im}(P_i) \subset \ker(L-\lambda_i)^{s_i}$. Sea ahora $x \in \ker(L-\lambda_i)^{s_i}$. Elijamos $j \neq i$. Todos los endomorfismos que sean polinomios en L conmutan entre sí, y lo son $(L-\lambda_k)^{s_k}$ y $Q_j(L)$. Con esto,

$$\begin{aligned}
P_j(x) &= (L-\lambda_1)^{s_1} \circ \cdots \circ \widehat{(L-\lambda_1)^{s_j}} \circ \cdots \circ (L-\lambda_1)^{s_i} \circ \cdots \circ (L-\lambda_p)^{s_p} \circ Q_j(L)(x)\\
&= (L-\lambda_1)^{s_1} \circ \cdots \circ \widehat{(L-\lambda_1)^{s_j}} \circ \cdots \circ \widehat{(L-\lambda_1)^{s_i}} \circ \cdots \circ (L-\lambda_p)^{s_p} \circ Q_j(L) \circ (L-\lambda_1)^{s_i}(x)
\end{aligned}$$

y como $(L-\lambda_1)^{s_i}(x) = 0$, debe ser $P_j(x) = 0$. Ahora, de $x = \sum_{j=1}^{p} P_j(x)$ obtenemos $x = P_i(x)$, luego $x \in \operatorname{im}(P_i)$ y $\ker(L-\lambda_i)^{s_i} \subset \operatorname{im}(P_i)$. ♣

Vamos a aplicar el teorema 131 al caso en el que L sea diagonalizable, que equivale a que sea $M(X) = (X-\lambda_1)\cdots(X-\lambda_p)$ con todos los $s_j = 1$. Entonces (6.1) nos da

$$\begin{aligned}
&\mathrm{id}_{\mathbb{E}} = \textstyle\sum_{i=1}^{h} \frac{1}{(\lambda_i-\lambda_1)\cdots\widehat{(\lambda_i-\lambda_i)}\cdots(\lambda_i-\lambda_p)} (L-\lambda_1) \circ \cdots \circ \widehat{(L-\lambda_i)} \circ \cdots \circ (L-\lambda_p),\\
&Q_i(L) = \left[(\lambda_i-\lambda_1)\cdots\widehat{(\lambda_i-\lambda_i)}\cdots(\lambda_i-\lambda_p)\right]^{-1}, \quad H_i(L) = \tfrac{(L-\lambda_1)\circ\cdots\circ\widehat{(L-\lambda_i)}\circ\cdots\circ(L-\lambda_p)}{(\lambda_i-\lambda_1)\cdots\widehat{(\lambda_i-\lambda_i)}\cdots(\lambda_i-\lambda_p)}.
\end{aligned} \tag{6.6}$$

El teorema 131 nos dice, al ser $s_i = 1$, que la imagen de P_i es el espacio propio de λ_i, luego $L(P_i(x)) = \lambda_i P_i(x)$ para todo $x \in \mathbb{E}$ y, a continuación,

$$L(x) = L(\mathrm{id}(x)) = L(P_1(x) + \ldots + P_p(x)) = \lambda_1 P_1(x) + \ldots + \lambda_p P_p(x);$$

o sea, $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$. Se calcula enseguida

$$L^2 = (\lambda_1 P_1 + \ldots + \lambda_p P_p) \circ (\lambda_1 P_1 + \ldots + \lambda_p P_p) = \sum_{i,j=1}^{p} \lambda_i \lambda_j P_i \circ P_j = \sum_{i=1}^{p} \lambda_i \lambda_i P_i \circ P_i = \sum_{i=1}^{p} (\lambda_i)^2 P_i,$$

ya que $P_i \circ P_j = 0$ si $i \neq j$ y ${}_jP_i \circ P_i = P_i$. Por inducción,

$$L^k = L \circ \cdots \circ L = \sum_{i=1}^{p} (\lambda_i)^k P_i, \tag{6.7}$$

que es un método para calcular potencias arbitrarias de L.

Problema 251 *Hacer el trabajo anterior para $L : \mathbb{R}^2 \to R^2$ con matriz*

$$a = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix},$$

calculando las λ_i, P_i y la forma concreta de (6.7). ♦

Solución. Es inmediato que $C(X) = M(X) = (X-2)(X-3)$ luego $\lambda_1 = 2$ y $\lambda_2 = 3$. La descomposición para $p = 2$ tras la fórmula general (6.6) es

$$1 = \frac{(X-3)}{(2-3)} + \frac{(X-2)}{(3-2)}.$$

De aquí se obtiene

$$P_1 = \frac{(a-3)}{(2-3)} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad P_2 = \frac{(a-2)}{(a-2)} = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix},$$

$$L^k = 2^k \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + 3^k \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2^k & 2^k - 3^k \\ 0 & 3^k \end{pmatrix}. \blacklozenge$$

Problema 252 *Probar que $L : \mathbb{R}^3 \to \mathbb{R}^3$ asociado a la matriz a*

$$\begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

es diagonalizable y aplicar las construcciones anteriores para expresar $L = \sum \lambda_i P_i$ siendo los λ_i los autovalores y las P_i las proyecciones en los subespacios propios. ¿Cuánto vale la matriz de L^{17} ? ♦

Solución. El polinomio característico es

$$C(X) = \begin{vmatrix} 1-X & 0 & -4 \\ 0 & 1-X & 0 \\ -1 & 0 & 1-X \end{vmatrix} = (1-X) \begin{vmatrix} 1-X & -4 \\ -1 & 1-X \end{vmatrix}$$
$$= (1-X)\left((1-X)^2 - 4\right) = (1-X)(X+1)(X-3).$$

Al tener tres valores propios $\lambda_1 = 1$, $\lambda_2 = -1$ y $\lambda_3 = 3$ diferentes, es diagonalizable y $M(X) = (X-1)(X+1)(X-3)$. Para descomponer en fracciones simples usamos la fórmula (6.4) y

$$1 = \frac{(X+1)(X-3)}{(1+1)(1-3)} + \frac{(X-1)(X-3)}{((-1)-1)((-1)-3)} + \frac{(X-1)(X+1)}{(3-1)(3+1)}$$
$$= \frac{(X+1)(X-3)}{-4} + \frac{(X-1)(X-3)}{8} + \frac{(X-1)(X+1)}{8}.$$

Por tanto,

$$\mathrm{id} = -\frac{1}{4}(L+1)(L-3) + \frac{1}{8}(L-1)(L-3) + \frac{1}{8}(L-1)(L+1),$$

siendo las proyecciones en los subespacios propios (P_i proyecta sobre $\mathbb{E}(\lambda_i)$)

$$P_1 = -\frac{1}{4}(L+1)(L-3) = -\frac{1}{4} \begin{pmatrix} 2 & 0 & -4 \\ 0 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix} \begin{pmatrix} -2 & 0 & -4 \\ 0 & -2 & 0 \\ -1 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$P_2 = \frac{1}{8}(L-1)(L-3) = \frac{1}{8}\begin{pmatrix} 0 & 0 & -4 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}\begin{pmatrix} -2 & 0 & -4 \\ 0 & -2 & 0 \\ -1 & 0 & -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix},$$

$$P_3 = \frac{1}{8}(L-1)(L+1) = \frac{1}{8}\begin{pmatrix} 0 & 0 & -4 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 & -4 \\ 0 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

Se puede comprobar que los cuadrados de las tres matrices de la derecha son ellas mismas (si no, no serían proyecciones). La descomposición pedida es

$$L \approx a = (1)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + (-1)\begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix} + (3)\begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

donde a se identifica con L. Para la pregunta final,

$$\begin{aligned} L^{17} \approx a^{17} &= (1)^{17}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + (-1)^{17}\begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix} + (3)^{17}\begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix} + 3^{17}\begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}. \end{aligned}$$

Por supuesto, sin ordenador cuesta ver que $3^{17} = 129\,140\,163$ pero ¿no es mucho peor calcular a^{17}? ♦

Problema 253 *Calcular a^k con $k \in \mathbb{N}$ para la matriz*

$$a = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

que aparece en la sección sobre la sucesión de Fibonacci. Nota: para abreviar los cálculos le decimos al lector que el polinomio característico es $C(X) = X^2 - X - 1$ con raíces

$$\lambda = \frac{1}{2} + \frac{1}{2}\sqrt{5}, \quad \mu = \frac{1}{2} - \frac{1}{2}\sqrt{5} \text{ cumpliendo } \lambda + \mu = 1,\ \lambda - \mu = \sqrt{5}.$$

Conviene poner la solución en función de λ y μ y, si se quiere, sustituir al final.

Problema 254 *Estudiar si $L : \mathbb{R}^3 \to \mathbb{R}^3$, que viene respectivamente asociado a las matrices*

$$a = \begin{pmatrix} \alpha & -1 & -1 \\ -1 & \alpha & -1 \\ -1 & 1 & \alpha \end{pmatrix}, \quad b = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ -1 & 1 & 2 \end{pmatrix}$$

es diagonalizable, y entonces expresarlo como $L = \sum \lambda_i P_i$ con los valores propios λ_i y las proyecciones P_i sobre los subespacios propios. ¿Cuánto vale $b^{10}(1,1,1)^\top$?

Vamos a descomponer ahora L con la llamada **descomposición espectral** válida aunque L no sea diagonalizable. Para L con polinomio minimal $M(X) = (X - \lambda_1)^{s_1} \cdots (X - \lambda_p)^{s_p}$ definimos

$$D, N : \mathbb{E} \to \mathbb{E}, \quad D = \lambda_1 P_1 + \ldots + \lambda_p P_p, \quad N = (L - \lambda_1) \circ P_1 + \ldots + (L - \lambda_p) \circ P_p. \tag{6.8}$$

Ahora podemos completar el teorema 127 al principio del capítulo y demostrarlo.

Teorema 132 *Con las definiciones de (6.8) se tiene que D es diagonalizable, N es nilpotente, $D \circ N = N \circ D$ y $L = D + N$. La descomposición $L = D + N$ con estas propiedades es única.*

Demostración. Es evidente que D es diagonalizable por el teorema 130 y que como P y N son polinomios en L (por construcción) han de conmutar. Además

$$D + N = \sum_{i=1}^{p} (\lambda_i + L - \lambda_i) \circ P_i = \sum_{i=1}^{p} L \circ P_i = L \circ \left(\sum_{i=1}^{p} P_i \right) = L \circ \mathrm{id} = L.$$

Mostremos que N es nilpotente. Empecemos calculando

$$\begin{aligned} N^2 = N\circ N &= [(L-\lambda_1)\circ P_1 + \ldots + (L-\lambda_p)\circ P_p]\circ[(L-\lambda_1)\circ P_1 + \ldots + (L-\lambda_p)\circ P_p] \\ &= \sum_{i,j=1}^{p} (L-\lambda_i)\circ P_i \circ (L-\lambda_j)\circ P_j = \sum_{i,j=1}^{p} (L-\lambda_i)\circ(L-\lambda_j)\circ P_i \circ P_j \\ &= \sum_{i=1}^{p} (L-\lambda_i)\circ(L-\lambda_i)\circ P_i^2 = \sum_{i=1}^{p} (L-\lambda_i)^2 \circ P_i. \end{aligned}$$

Se ha usado que los diversos factores son polinomios en L y por tanto conmutan, además de $P_i \circ P_j = 0$ si $i \neq j$ y que $P_i \circ P_i = P_i$. Queda para el lector probar por inducción con ideas muy parecidas que

$$N^k = \sum_{i=1}^{p} (L-\lambda_i)^k \circ P_i.$$

Si $k = \max\{s_1, \ldots, s_p\}$, se anula $(L-\lambda_i)^{s_i}$ sobre $\operatorname{im}(P_i)$ (teorema 131) y obtenemos $(L-\lambda_i)^k \circ P_i = 0$ para todo i y $N^k = 0$.

Probamos la unicidad. Sea $L = D' + N'$ otra descomposición con D' diagonal, N' nilpotente y $D' \circ N' = N' \circ D'$. Calculamos

$$D'\circ L - L\circ D' = D'\circ(D'+N') - (D'+N')\circ D' = (D')^2 + D'\circ N' - (D')^2 - N'\circ D' = 0.$$

Como consecuencia, al ser D y N polinomios de L, los cuatro endomorfismos D, N, D' y N'conmutan entre si. El binomio de Newton de $N- N'$ da

$$(N-N')^k = \sum_{i=0}^{k} \binom{k}{i} N^i \circ (N')^{k-i}.$$

Es obvio que si h y h' son los primeros exponentes tales que $N^h = (N')^{h'} = 0$ y $k > h + h'$ uno de los factores en $N^i \circ (N')^{k-i}$ es nulo. Esto muestra que $(N-N')^k = 0$ y $N - N'$ es nilpotente. Admitamos de momento que si D y D' son diagonalizables hay una base *común para ambos* donde D y D' tienen matriz diagonal. Evidentemente, en esa base $D - D'$ tiene matriz diagonal d. Pero $D - D' = N' - N$ implica que d es nilpotente, y solo 0 es a la vez nilpotente y diagonal. Así pues $D - D' = N' - N = 0$. Esto prueba la unicidad salvo una afirmación pendiente que se probará a continuación. ♣

Si L y M son diagonalizables, existen bases $\mathcal{U}$ y $\mathcal{V}$ en las que respectivamente L y M tienen matriz diagonal pero podría suceder, por ejemplo, que L no tuviera matriz diagonal en $\mathcal{V}$. Se puede conseguir la simultánea diagonalización en condiciones muy generales.

Teorema 133 *Sea $\mathcal{L}$ un conjunto de endomorfismos de $\mathbb{E}$ que son diagonalizables y conmutan entre ellos. Hay una base $\mathcal{W}$ en la que todos los endomorfismos de $\mathcal{L}$ tienen matriz diagonal.*

Demostración. Como regla general, para endomorfismos *arbitrarios* L y M se tiene que si $L\circ M = M\circ L$ y $\mathbb{F}$ es un subespacio de $\mathbb{E}$ estable por M, entonces $\operatorname{im}(M)$ y $\ker(M)$ son estables por L. En efecto para $x = M(y) \in \operatorname{im}(M)$ se tiene $L(x) = L(M(y)) = M(L(y)) \in \operatorname{im}(M)$. El otro caso es similar. Esto nos permite saber que si $L\circ M = M\circ L$, los subespacios propios de M (sin duda estables por M), lo son también por L, y viceversa.

La demostración se hace por inducción sobre $\dim(\mathbb{E}) = n$, siendo cierto el teorema si $n = 1$. Tomemos $n > 1$ y supongamos cierto el teorema para espacios de dimensión $m < n$. Si todos los elementos de $\mathcal{L}$ son del tipo $L(x) = \kappa x$ con $\kappa \in \Bbbk$, cualquier base sirve. Supongamos pues que al menos un $L \in \mathcal{L}$ no es una homotecia. La descomposición $\mathbb{E} = \mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_p$ en subespacios propios de L tiene $p \geq 2$ y todos los sumandos con $\dim(\mathbb{F}_j) = m_j < n$. Por el párrafo precedente, cualquier $M \in \mathcal{L}$ tiene a los $\mathbb{F}_k$ como subespacios estables. Por la hipótesis inductiva hay bases $\mathcal{W}_k$ en cada $\mathbb{F}_k$ en donde todos los $M \in \mathcal{L}$, incluido L que sin duda preserva los $\mathbb{F}_k$, cumplen que $M|_{\mathbb{F}_k} : \mathbb{F}_k \to \mathbb{F}_k$ tienen matriz diagonal. Yuxtaponiendo las bases $\mathcal{W}_k$ resulta una base $\mathcal{W}$ de $\mathbb{E}$ que cumple el teorema. ♣

Problema 255 *Descomponer $L = D + N$ para $L : \mathbb{R}^4 \to \mathbb{R}^4$ asociado a la matriz*

$$a = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 3 \end{pmatrix}. \blacklozenge$$

Solución. Se calcula fácilmente que $C(X) = (X-2)^3(X-3)$ luego hay dos valores propios $\lambda_1 = 2$ y $\lambda_2 = 3$. Habrá que conocer algunas de las potencias de $a-2$ y $a-3$ y sus productos, por tanto hay que guardar las cuentas que se hacen para calcular $M(X)$, que son

$$(a-3)(a-2) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 1 & 0 \\ -1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix},$$

$$(a-3)(a-2)^2 = \begin{pmatrix} 0 & -1 & 1 & 0 \\ -1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix},$$

$$(a-3)(a-2)^3 = \begin{pmatrix} -1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} = 0.$$

Por tanto $M(X) = C(X)$. Para tener las proyecciones $P_1 = H_1(L)$ y $P_2 = H_2(L)$ necesitamos conocer los polinomios $Q_i(X)$ en

$$1 = Q_1(X)(X-3) + Q_2(X)(X-2)^3, \text{ siendo } \deg Q_1(X) \leq 3, \deg Q_2(X) < 1.$$

Con la fórmula del problema 249 o del teorema 129 siendo $d = \lambda_2 - \lambda_1 = 1$, obtenemos

$$1 = -\left[(X-2)^2 + (X-2) + 1\right](X-3) + (X-2)^3.$$

Entonces las proyecciones $P_i = Q_i(L)\,\phi_i(L)$ son

$$P_1 = -\left[I + (a-2) + (a-2)^2\right](a-3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad P_2 = (a-2)^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Con todo esto, la parte diagonalizable D es

$$D = \lambda_1 P_1 + \lambda_2 P_2 = 2\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} + 3\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 3 \end{pmatrix}$$

y la parte nilpotente es

$$\begin{aligned} N &= (L-\lambda_1)\circ P_1 + (L-\lambda_2)\circ P_2 \\ &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix}. \end{aligned}$$

El lector puede pensar que hay un error en D, porque no es una matriz *diagonal* pero todo se aclara porque lo que dice el teorema 127 es que D es *diagonalizable* y, efectivamente, en una base *que no es* $\mathcal{E}$, D tiene matriz diagonal. Para mucho cálculos, no es el conocimiento de D y N lo que importa sino, también, las *descomposiciones* $D = \lambda_1 P_1 + \lambda_2 P_2$ y $N = (L-\lambda_1)\circ P_1 + (L-\lambda_2)\circ P_2$.

Si quisiéramos calcular L^{100} observaríamos primero que

$$N^2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 \end{pmatrix}$$

y $N^3 = 0$. Con el binomio de Newton,

$$L^{100} = (D+N)^{100} = \sum_{i=0}^{100} \binom{100}{i} D^{100-i} N^i = D^{100} + 100 D^{99} N + (50 \cdot 99)\, D^{98} N^2,$$

y queda pendiente conocer D^k. Podría hacerse con las técnicas del capítulo anterior, pero es mejor usar que $D^k = \lambda_1^k P_1 + \lambda_2^k P_2$ y obtener por ejemplo

$$D^{99} = 2^{99} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} + 3^{99} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

La fórmula de L^{100} es sin duda complicada, pero puede conseguirse. ♦

Problema 256 *Sea L descomponible como $L = D + N$ por el teorema 132 o 127 (son el mismo) y que $H : \mathbb{E} \to \mathbb{E}$ es un isomorfismo. Si $M = H^{-1} \circ L \circ H$, tenemos que $M = H^{-1} \circ D \circ H + H^{-1} \circ N \circ H$, siendo $D' = H^{-1} \circ D \circ H$ diagonalizable, $N' = H^{-1} \circ M \circ H$ nilpotente, y $D' \circ N' = N' \circ D'$, luego $M = D' + N'$ es la descomposición de esos teoremas. Probar estas afirmaciones.*

Lo que acabamos de probar es una fuente de problemas de puro cálculo. Se toma L obviamente descomponible como $L = D + N$ y H, se define $M = H^{-1} \circ L \circ H$, y se pide calcular la descomposición $M = D' + N'$. Tomamos, identificando los endomorfismos con matrices 3×3,

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix},\ D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix},\ N = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ H = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

Problema 257 *Descomponer $M = D' + N'$ (se parte del desconocimiento de L, D, N y H).*

A quien piense que hay una respuesta inmediata

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

le advertimos que N no es nilpotente y $D \circ N - N \circ D \neq 0$ luego *la respuesta es falsa*. Para abreviar, el lector puede partir de que $C_M(X) = M_M(X) = X(X-1)^2$ con raíces $\lambda_1 = 0$ y $\lambda_2 = 1$.

Damos otro ejemplo muy parecido, advirtiendo que los cálculos son laboriosos. Tomamos, identificando los endomorfismos con matrices 4×4,

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},\ D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},\ N = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},\ H = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}.$$

Problema 258 *Descomponer $M = D' + N'$ (se parte del desconocimiento de L, D, N y H).*

Ayudamos diciendo que $C_M(X) = X^4 - 2X^3 + X^2 = X^2(X-1)^2$ con raíces $\lambda_1 = 0$ y $\lambda_2 = 1$, y que

$$M - 0 = \begin{pmatrix} 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 \end{pmatrix}, \quad M - 1 = \begin{pmatrix} -1 & -1 & 0 & -1 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix},$$

siendo $M(M-1)$, $M^2(M-1)$ y $M(M-1)^2$ distintos de 0, luego $C(X) = M(X) = X^2(X-1)^2$.

6.4. Un breve contacto con las ecuaciones diferenciales lineales

6.4.1. Qué son estas ecuaciones y cómo son sus soluciones

La utilidad de la descomposición $L = D + N$ se entiende al ver que es una herramienta efectiva de solución de ecuaciones diferenciales lineales. En *Ecuaciones diferenciales matriciales* del capítulo *Autovalores y autovectores* explicamos lo que era una **ecuación diferencial lineal con coeficientes constantes** $\dot{x} = ax$ y el que $x : \mathbb{R} \to \mathbb{R}^n$ fuera una **solución de la ecuación**. Allí resolvimos el caso en el que $a \in \mathbb{R}^{n\times n}$ era una matriz *diagonalizable*. El trabajo suponía calcular una matriz invertible c tal que $d = c^{-1}ac$ fuese diagonal y estudiar como ecuación auxiliar $\dot{y} = dy$. Veremos aquí otro procedimiento que evita el cálculo de c y que tiene la ventaja, no tanto de ahorrar trabajo, sino de poder afrontar el caso en el que a no sea diagonalizable. No obstante, se necesitará en todo caso que el polinomio mínimo sea linealmente factorizable. (Esto equivale a que $C(X)$ lo sea).

En el caso $n = 1$ no había problema porque $a \in \mathbb{R}$ y $\dot{x} = ax$ tenía solución general $x(t) = x_0 e^{at}$ y, como $e^0 = 1$, se tenía $x(0) = x_0$ y x_0 era la **condición inicial** en el instante $t = 0$. Lo que importa, aunque no se demostrará, es que hay un teorema que generaliza el caso $n = 1$ *con un gran paralelismo.* Vamos a pedir al lector que admita durante un breve periodo que a cada matriz cuadrada $b \in \mathbb{R}^{n\times n}$ se le puede asignar una nueva matriz llamada la **exponencial de la matriz** b, que se denota por e^b o $\exp(b)$ que pronto explicaremos. La solución general de $\dot{x}(t) = ax(t)$ para n arbitrario con una matriz arbitraria a es

$$x(t) = e^{ta}x_0 = \exp(ta)\, x_0, \quad t \in I,\ x_0 \in \mathbb{R}^n.$$

Dicho con más detalle, la solución más general se obtiene multiplicando un $x_0 \in \mathbb{R}^n$ por una matriz e^{ta} y así se obtiene $x(t)$ función del tiempo t, que es la solución de la ecuación, y hay una solución para cada elección de x_0. Aunque exp no se ha definido aún, diremos que $\exp(0) = e^0$ (la exponencial de la matriz 0) es la matriz unidad I. Dicho esto, $x(0) = e^0 \cdot x_0 = Ix_0$ de modo que x_0 tiene la interpretación similar a la ya conocida si $n = 1$ de representar el estado inicial del sistema físico del que es modelo la ecuación diferencial. Nuestro trabajo se va a limitar a definir la exponencial y mostrar como calcularla, lo que permite resolver $\dot{x} = ax$ si el polinomio mínimo o característico de a es linealmente factorizable.

6.4.2. La exponencial de una matriz y su cálculo

Si $a \in \mathbb{R}$, el *número* e^a, la **exponencial de** a, se define como la suma de la serie

$$e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!} = \lim_{k\to\infty} \left(1 + \frac{a}{1!} + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots + \frac{a^{k-1}}{(k-1)!} + \frac{a^k}{k!}\right) = \lim_{k\to\infty} \left(\sum_{n=0}^{k} \frac{a^n}{n!}\right).$$

Podría ser que la sucesión $(s_k) = \left(\sum_{n=0}^{k} \frac{a^n}{n!}\right)$ no tuviera límite (y entonces no existiría e^a) pero de hecho sí existe y es un número > 0. Se ve claramente que $e^0 = 1$ y se puede probar que $e^{a+b} = e^a e^b$ (la exponencial de la suma es el producto de las exponenciales). Aún así, números como $e^{3/2}$ o $e^1 = e$ (el **célebre número de Euler** e), son solo manejables a través de una aproximación decimal, como sucede con $\pi \approx 3{,}141592$. Digamos que $e \approx 2{,}71828$.[5]

[5] Para acordarse de las primeras cifras de e, recordar "He studied a treatise on Calculus" y para π "How I like a drink, ¡alcoholic of course!, after two heavy lectures". Basta contar las letras de cada palabra y saber inglés. El autor está seguro que dentro de muchos años el lector habrá olvidado mucha ciencia y fatigas de las matemáticas de Primer Curso pero recordará estos pequeños datos.

La exponencial de las matrices cuadradas a se define formalmente como en la ecuación para $a \in \mathbb{R}$ que se puede copiar sin variar nada. Solo hay que tener en cuenta que (s_k), la sucesión de sumas parciales, es ahora una sucesión de matrices y no de números, y que una sucesión de matrices (c_n) converge a una matriz c si hay convergencia componente a componente. Pues bien, como en el caso $a \in \mathbb{R}$, se cumple que $e^a \in \mathbb{R}^{n\times n}$ siempre existe cualquiera que sea la matriz a. Se tiene también como evidente que $e^0 = I$, la matriz unidad, pero exige un cálculo complejo mostrar que $e^{a+b} = e^a e^b$ *si las matrices conmutan* (o sea, $ab = ba$). Es cierto también que la condición $e^a > 0$ cuando $a \in \mathbb{R}$ se generaliza a que e^a sea una matriz invertible. *Queda pendiente el cálculo efectivo de* e^a *(que será* e^{ta} *para resolver ecuaciones) y aquí interviene el Álgebra Lineal.*

Digamos primero que si a es diagonal es fácil ver, aplicando directamente la definición, que

$$\exp\begin{pmatrix} d_1^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n^n \end{pmatrix} = \begin{pmatrix} \exp\left(d_1^1\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp\left(d_n^n\right) \end{pmatrix}$$

y, por ejemplo,

$$\exp\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} e^2 & 0 \\ 0 & e^4 \end{pmatrix}.$$

Si a es diagonalizable, hay una matriz invertible c tal que $c^{-1}ac = d$ con d diagonal. Como $a^k = \left(cdc^{-1}\right)^k = cd^kc^{-1}$ se obtiene enseguida que $\exp(a) = c\exp(d)\,c^{-1}$. Vimos en el capítulo anterior que c es matriz de cambio de base y cómo puede obtenerse, así que, si d es diagonalizable sabemos cómo calcular $\exp(a)$.[6] Si en vez de a ponemos ta con $t \in \mathbb{R}$, se tiene $c^{-1}(ta)\,c = td$ y $\exp(ta) = c\exp(td)\,c^{-1}$. La solución $x(t)$ de $\dot{x}(t) = ax(t)$ con la condición inicial $x(0) = x_0$ es entonces

$$x(t) = \exp(ta)\cdot x_0 = \left(c\exp(td)\,c^{-1}\right)\cdot x_0. \tag{6.9}$$

Hasta aquí no se obtiene nada realmente nuevo respecto a lo visto en el capítulo anterior porque $\exp(td)$ se obtiene como una nueva matriz diagonal calculando término a término las exponenciales de los números en la diagonal de d. Mientras a sea diagonalizable, el lector puede elegir resolver $\dot{x} = ax$ como en el capítulo anterior o como aquí se lo hemos contado.

Sin embargo, la situación puede ser bastante más complicada. Lo vamos a estudiar con L, endomorfismo de $\mathbb{E}$, pero si se quiere se puede pensar que L es una matriz y que al descomponer $L = D+N$, D y N son endomorfismos o matrices. Tengamos en cuenta que, aun pensando en matrices, D es *diagonalizable*, pero no necesariamente *diagonal*, y esto complica un tanto el cálculo de e^D. Si $L = D + N$, tenemos sin duda que $e^L = e^D e^N$ y que, al ser N nilpotente, la serie de e^N se convierte en una suma finita, de hecho un polinomio en N. Aunque en casos sencillos con $n = 2, 3$, se puede abordar el cálculo directo, conviene tener una fórmula para e^L que ahorre operar al máximo.

Teorema 134 *A partir de la descomposición*

$$L = [\lambda_1 P_1 + (L-\lambda_1)\circ P_1] + \ldots + [\lambda_p P_p + (L-\lambda_p)\circ P_p]$$

y, abreviando $N_i = (L-\lambda_i)\circ P_i$, *se tiene que*

$$e^L = e^{\lambda_1}e^{N_1}\circ P_1 + \cdots + e^{\lambda_j}e^{N_j}\circ P_j = \sum_{i=1}^{p} e^{\lambda_i}e^{N_i}\circ P_i.$$

Demostración. Se va a utilizar muchas veces sin mayor comentario que como las e^{λ_i} son números o matrices escalares, y las matrices P_i y e^{N_j} son polinomios en L, *todas ellas conmutan entre sí*. Esto, unido a que $P_i^2 = P_i$ y $P_i\circ P_j = 0$ si $i \neq j$, va a llevar a una gran simplificación de la fórmula inicial de e^L. Para empezar, como los sumandos $\lambda_i P_i + (L-\lambda_i)\circ P_i$ conmutan entre sí, la fórmula $e^{A+B} = e^A\circ e^B$ si $A\circ B = B\circ A$ nos da que

$$e^L = \exp(\lambda_1 P_1 + (L-\lambda_1)\circ P_1)\circ\cdots\circ\exp(\lambda_p P_p + (L-\lambda_p)\circ P_p).$$

Sobrentendiendo el subíndice i, calculamos

$$e^{\lambda P} = \frac{(\lambda P)^0}{0!} + \frac{(\lambda P)^1}{1!} + \ldots + \frac{(\lambda P)^k}{k!} + \ldots = 1 - P + \frac{\lambda^0}{0!}P + \frac{\lambda^1}{1!}P + \ldots + \frac{\lambda^k}{k!}P + \ldots = 1 - P + e^{\lambda}P,$$

[6] Hay otro procedimiento con lo visto en este capítulo que ya comentaremos.

$$\begin{aligned} e^{(L-\lambda)\circ P} &= \frac{((L-\lambda)\circ P)^0}{0!} + \frac{((L-\lambda)\circ P)^1}{1!} + \ldots + \frac{((L-\lambda)\circ P)^k}{k!} + \ldots \\ &= 1 - P + \frac{(L-\lambda)^0}{0!}\circ P + \frac{(L-\lambda)^1}{1!}\circ P + \ldots + \frac{(L-\lambda)^k}{k!}\circ P + \ldots \\ &= 1 - P + e^N \circ P, \end{aligned}$$

$$\begin{aligned} \exp(\lambda P + (L-\lambda)\circ P) &= \exp(\lambda P)\circ\exp((L-\lambda)\circ P) = \big(1 - P + e^\lambda P\big)\circ\big(1 - P + e^N \circ P\big) \\ &= \big(1 - P + e^\lambda P\big) - \big(P - P^2 + e^\lambda P^2\big) + \big(e^N\circ P - P\circ e^N \circ P + e^\lambda P \circ e^N \circ P\big) \\ &= \big(1 - P + e^\lambda P\big) - \big(e^\lambda P^2\big) + \big(e^N\circ P - e^N\circ P + e^\lambda \circ e^N \circ P\big) \\ &= 1 - P + e^\lambda \circ e^N \circ P, \end{aligned}$$

Sustituyendo, llegamos a

$$e^L = \big[1 - P_1 + e^{\lambda_1} e^{N_1}\circ P_1\big]\circ\cdots\circ\big[1 - P_p + e^{\lambda_p} e^{N_p}\circ P_p\big].$$

Vamos a probar por inducción sobre j la fórmula

$$\big[1 - P_1 + e^{\lambda_1} e^{N_1}\circ P_1\big]\circ\cdots\circ\big[1 - P_j + e^{\lambda_j} e^{N_j}\circ P_j\big] = 1 - \sum_{i=1}^{j} P_i + \sum_{i=1}^{j} e^{\lambda_i} e^{N_i}\circ P_i,$$

que es obvia si $j = 1$. Supongamos cierta la fórmula para $j-1$. Entonces,

$$\begin{aligned} &\big[1 - P_1 + e^{\lambda_1} e^{N_1} \circ P_1\big] \circ \cdots \circ \big[1 - P_{j-1} + e^{\lambda_{j-1}} e^{N_{j-1}} \circ P_j\big] \circ \big[1 - P_j + e^{\lambda_j} e^{N_j} \circ P_j\big] \\ &= \left(1 - \sum_{i=1}^{j-1} P_i + \sum_{i=1}^{j-1} e^{\lambda_i} e^{N_i} \circ P_i\right) \circ \big[1 - P_j + e^{\lambda_j} e^{N_j} \circ P_j\big] \\ &= \left(1 - \sum_{i=1}^{j-1} P_i + \sum_{i=1}^{j-1} e^{\lambda_i} e^{N_i} \circ P_i\right) \\ &- \left(1 \circ P_j - \sum_{i=1}^{j-1} P_i \circ P_j + \sum_{i=1}^{j-1} e^{\lambda_i} e^{N_i} \circ P_i \circ P_j\right) \\ &+ \left(1 \circ e^{\lambda_j} e^{N_j} \circ P_j - \sum_{i=1}^{j-1} P_i \circ e^{\lambda_j} e^{N_j} \circ P_j + \sum_{i=1}^{j-1} e^{\lambda_i} e^{N_i} \circ P_i \circ e^{\lambda_j} e^{N_j} \circ P_j\right) \\ &= \left(1 - \sum_{i=1}^{j-1} P_i + \sum_{i=1}^{j-1} e^{\lambda_i} e^{N_i} \circ P_i\right) - P_j + e^{\lambda_j} e^{N_j} \circ P_j. \end{aligned}$$

Haciendo $j = p$ y utilizando que $1 = P_1 + \ldots P_p$, llegamos a la fórmula enunciada. ♣

Si, aunque L no sea diagonalizable, el valor propio λ_i tiene multiplicidad 1, el sumando $e^{\lambda_i} e^{N_i}\circ P_i = e^{\lambda_i} P_i$. Esto se ve de muchas maneras. Por ejemplo, porque $N_i = (L - \lambda_i)\circ P_i = 0$ al proyectar P_i en el subespacio propio de λ_i. Para el cálculo de e^L, aparte de los P_i, tendremos el trabajo adicional de los e^{N_j} con $N_j \neq 0$.

Damos ejemplos. En el problema 252 se han calculado para a, que es diagonalizable, sus valores propios $\lambda_1 = 1$, $\lambda_2 = -1$ y $\lambda_3 = 3$ y sus proyecciones P_1, P_2 y P_3. En concreto,

$$a = \begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad P_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad P_2 = \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}, \quad P_3 = \begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

Aquí $N = 0$ porque $L = a$ es diagonalizable al tener tres valores propios diferentes. Resulta

$$e^a = e^1 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + e^{-1} \begin{pmatrix} \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix} + e^3 \begin{pmatrix} \frac{1}{2} & 0 & -1 \\ 0 & 0 & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{pmatrix}.$$

Observemos que a es diagonalizable pero no diagonal, y que el procedimiento de cálculo de e^a ha evitado determinar una base diagonalizadora al precio de tener que construir las P_i. El lector decidirá qué vía prefiere.

Otro ejemplo es el cálculo de

$$e^N = \exp\left(\begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix},$$

porque $N^2 = 0$. En estos casos con $D = 0$, el cálculo directo suele ser lo más sencillo. Presentamos un problema aunque trucado. Sean

$$q = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad h = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix}, \quad a = qhq^{-1} = \begin{pmatrix} 3 & -1 & 1 & 0 \\ 1 & 1 & -1 & -1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix}.$$

Problema 259 *Nos dan la matriz a (sin conocer la factorización $a = qhq^{-1}$). Se pide e^a.* ♦

Solución. Calculamos de modo directo o haciendo trampa (porque $C_a(X) = C_h(X)$) que $C(X) = (X-2)^2(X-3)^2$. Hay dos valores propios $\lambda_1 = 2$ y $\lambda_2 = 3$. Necesitaremos

$$a - 2 = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad (a-2)^2 = \begin{pmatrix} 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix}^2,$$

$$a - 3 = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & -2 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (a-3)^2 = \begin{pmatrix} -1 & 2 & 1 & 1 \\ -2 & 3 & 2 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Con estas matrices se ve que si en el polinomio $(X-2)^i(X-3)^j$ se sustituye X por a no sale 0 salvo si $i = j = 2$, luego $C(X) = M(X) = (X-2)^2(X-3)^2$ Para las proyecciones se necesita la descomposición planteada en el problema 247 de $(X-\alpha)^2(X-\beta)^2$, que es

$$\begin{aligned} 1 &= \left(\frac{2}{d^3}(X-\alpha) + \frac{1}{d^2}\right)(X-\beta)^2 + \left(-\frac{2}{d^3}(X-\beta) + \frac{1}{d^2}\right)(X-\alpha)^2 \\ &= (2(X-2)+1)(X-3)^2 + (-2(X-3)+1)(X-2)^2, \end{aligned}$$

Según el teorema 134 la fórmula es $e^L = e^{\lambda_1} e^{N_1} \circ P_1 + e^{\lambda_2} e^{N_2} \circ P_2$. Tenemos que P_1 y P_2 son

$$(2(a-2)+1)(a-3)^2 = \begin{pmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (-2(a-3)+1)(a-2)^2 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$e^{N_1} = 1 + (a-2) = \begin{pmatrix} 2 & -1 & 1 & 0 \\ 1 & 0 & -1 & -1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}, \qquad e^{N_2} = 1 + (a-3) = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

$$e^{N_1} \circ P_1 = \begin{pmatrix} 2 & -1 & -2 & -2 \\ 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad e^{N_2} \circ P_2 = \begin{pmatrix} 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

$$e^a = e^2 \begin{pmatrix} 2 & -1 & -2 & -2 \\ 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + e^3 \begin{pmatrix} 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2e^2 & -e^2 & 2e^3 - 2e^2 & e^3 - 2e^2 \\ e^2 & 0 & -e^2 & -e^2 \\ 0 & 0 & e^3 & 0 \\ 0 & 0 & e^3 & e^3 \end{pmatrix}. \quad ♦$$

El lector puede proponerse cuantos problemas quiera de este tipo, si bien le recomendamos que siga un camino similar al elegido para el problema precedente. La primera ventaja es que el polinomio minimal

será sencillo y la segunda que podrá comprobar si tiene bien el resultado. En efecto, $a = qhq^{-1}$ implica que $e^a = qe^h q^{-1}$ y e^h es muy fácil de calcular porque

$$h = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix} = h_1 + h_2 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

con h_1 diagonal, h_2 nilpotente y $h_1 h_2 = h_2 h_1$. Por tanto,

$$e^h = \begin{pmatrix} e^2 & 0 & 0 & 0 \\ 0 & e^2 & 0 & 0 \\ 0 & 0 & e^3 & 0 \\ 0 & 0 & 0 & e^3 \end{pmatrix} \left(\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \frac{1}{1!} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \right) \right) = \begin{pmatrix} e^2 & 0 & 0 & 0 \\ e^2 & e^2 & 0 & 0 \\ 0 & 0 & e^3 & 0 \\ 0 & 0 & e^3 & e^3 \end{pmatrix}.$$

Según esto, llegamos por otra vía al valor de e^a confirmando el cálculo anterior,

$$\begin{aligned} e^a &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^2 & 0 & 0 & 0 \\ e^2 & e^2 & 0 & 0 \\ 0 & 0 & e^3 & 0 \\ 0 & 0 & e^3 & e^3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 2e^2 & -e^2 & -2e^2 + 2e^3 & -2e^2 + e^3 \\ e^2 & 0 & -e^2 & -e^2 \\ 0 & 0 & e^3 & 0 \\ 0 & 0 & e^3 & e^3 \end{pmatrix}. \end{aligned}$$

Problema 260 *Calcular la exponencial de*

$$a = \begin{pmatrix} 1+h & 1 & 1 & \cdots \\ 1 & 1+h & 1 & \cdots \\ 1 & 1 & 1+h & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \in \mathbb{R}^{n \times n}$$

(todo 1 *excepto* $1+h$ *en la diagonal). Si se prefiere, suponer* $n = 4$ *o incluso* $n = 3$.

Problema 261 *Le damos como dato al lector que*

$$a = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

tiene $M(X) = X(X-1)(X-2)$. *Calcular* e^a.

Problema 262 *Para* $h \in \mathbb{R}$ *arbitrario no nulo, calcular* e^a *siendo*

$$a = \begin{pmatrix} 1 & 0 & 0 \\ 1 & h & 1 \\ h-1 & 0 & h \end{pmatrix}.$$

Capítulo 7

La forma de Jordan

Este capítulo nos parece de una dificultad muy superior al resto, pero lo cierto es que muchos primeros textos de Álgebra Lineal incluyen con más o menos detalle la forma de Jordan. *No es parte del curso que damos*, conformándonos con la sección *Un primer contacto con la forma de Jordan* del capítulo *Autovalores y autovectores*. Al liberar al estudiante nos sentimos libres a su vez para hacer aquí una exposición bastante extensa que, por necesidad o curiosidad, pueda el lector consultar en el futuro. Quizás se sienta tentado a leer *Planteamiento y pasos iniciales* y saber lo que aquí queda en depósito.

7.1. Planteamiento y pasos iniciales

Sea $\mathbb{E}$ de dimensión n y $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo con polinomios característico y minimal

$$C(X) = (\lambda_1 - X)^{m_1} \cdots (\lambda_p - X)^{m_p}, \quad M(X) = (X - \lambda_1)^{s_1} \cdots (X - \lambda_p)^{s_p}, \tag{7.1}$$

factorizables con factores de grado 1 (si $C(X)$ lo cumple, lo cumple $M(X)$ y viceversa). *Esto se supondrá en todo este capítulo si no se dice lo contrario e, implícitamente, que las raíces λ_j de $C(X)$ se han ordenado de una vez por todas.* Se sabe que el que $C(X)$ sea linealmente factorizable no es suficiente para que L sea diagonalizable (i.e., tenga una base de vectores propios y matriz diagonal en ella). Vamos a mostrar que si se generaliza el concepto de vector propio, el que $C(X)$ sea linealmente factorizable es *necesario y suficiente* para que L admita una base de vectores propios generalizados (aún no definidos).[1] Como se puede imaginar, en estas bases la matriz a de L va a ser particularmente sencilla y nuestro objeto es tratar este problema. Esto tiene interés desde un punto de vista meramente *computacional* pues facilita el cálculo de las potencias de a.[2] Pero hay otra cuestión importantísima, y es que el número de matrices a que pueden aparecer una vez fijados los valores propios de L, es *finito*. Hay pues un *procedimiento de clasificación* de estos endomorfismos y, conocidos todos los tipos, podemos *hacernos idea de la variedad y complejidad* que en ellos puede haber. La condición de que $C(X)$ y $M(X)$ sean linealmente factorizables no es restrictiva si $\mathbb{k} = \mathbb{C}$ (por el teorema fundamental del Álgebra) siendo además frecuente si $\mathbb{k} \neq \mathbb{C}$, luego podemos estudiar muchos casos interesantes.

Para L endomorfismo de $\mathbb{E}$ abreviaremos $L_\lambda = L - \lambda \operatorname{id}_{\mathbb{E}} = L - \lambda$. Consideramos las potencias L_λ^q para $q = 0, 1, 2, \ldots$ de L_λ, siendo obviamente $L_\lambda^0 = \operatorname{id}_{\mathbb{E}}$ y $L_\lambda^1 = L_\lambda$. y será frecuente para ayudar en la comprensión de definiciones poner $L_\lambda^0(v)$ y $L_\lambda^1(v)$ en vez de v y $L_\lambda(v)$. Sea $v \neq 0$ y la sucesión de vectores $Z_\lambda(v) = \left(L_\lambda^0(v), L_\lambda^1(v), L_\lambda^2(v), \ldots, L_\lambda^{q-1}(v)\right)$ siendo todos los $L_\lambda^j(v) \neq 0$ pero $L_\lambda^q(v) = 0$. Decimos que $Z_\lambda(v)$ es el **ciclo generado por** v, siendo q la **longitud del ciclo** y $L_\lambda^0(v)$ y $L_\lambda^{q-1}(v)$ los **elementos inicial** y **final de** $Z_\lambda(v)$.[3] El concepto es semejante pero no igual al de ciclo tratado para demostrar el teorema de Cayley-Hamilton. y se puede probar de modo muy semejante a cómo allí se hizo que un ciclo es una sucesión independiente. La importancia de estos ciclos radica en que se puede

[1] Se vio en *Triangulación* en el capítulo *Autovalores y autovectores* que el que $C(X)$ sea linealmente factorizable es necesario y suficiente para que L tenga una matriz triangulable, pero lo que pretendemos ahora es bastante mejor.

[2] Y por ello la exponencial y la resolución de ecuaciones diferenciales lineales.

[3] La mayoría de los autores ordenan los ciclos en orden inverso $\left(L_\lambda^{q-1}(v), \ldots, L_\lambda^2(v), L_\lambda(v), L_\lambda^0(v)\right)$, siendo para ellos $L_\lambda^{q-1}(v)$ y $L_\lambda^0(v) = v$ los elementos inicial y final. Esto supone (se verá más adelante) que los bloques de Jordan sean, con esa regla, matrices triangulares superiores y para nosotros triangulares inferiores.

poner con $v = v_0$,

$$(v_0, v_1, v_2, \ldots, v_{q-1}) = \left(L_\lambda^0(v), L_\lambda^1(v), L_\lambda^2(v), \ldots, L_\lambda^{q-1}(v)\right) = (v_0, L_\lambda(v_0), L_\lambda(v_1), \ldots, L_\lambda(v_{q-2})),$$

y $L_\lambda^q(v_0) = L_\lambda(v_{q-1}) = 0$. Al ser $L_\lambda(x) = (L-\lambda)x = y$ deducimos que $L(x) = \lambda x + y$. Entonces, las v_i se relacionan por

$$L(v_0) = \lambda v_0 + v_1,\ L(v_1) = \lambda v_1 + v_2,\ \ldots\ ,\ L(v_{q-2}) = \lambda v_{q-2} + v_{q-1},\ L(v_{q-1}) = \lambda v_{q-1}.$$

Vemos por tanto que si $Z_\lambda(v)$ forma parte de una base, al aplicar L a uno de sus elementos $L_\lambda^i(v) = v_i$, su valor solo dependerá de v_i y v_{i+1} salvo para v_{q-1}, lo que garantiza la aparición de ceros en la matriz. Si como caso extremo $\mathbb{E}$ tiene una base formada por un *único* $Z_\lambda(v)$ (no siempre será así pero en este caso $q = n = \dim(\mathbb{E})$) la matriz de L en esa base es

$$J_\lambda(q) = \begin{pmatrix} \lambda & & & & & \\ 1 & \lambda & & & & \\ & 1 & \lambda & & & \\ & & & \ddots & & \\ & & & & \lambda & \\ & & & & 1 & \lambda \end{pmatrix} \in \mathbb{k}^{q\times q}$$

con todo λ en la diagonal, todo 1 bajo la diagonal y 0 los demás términos 0. ¿Y si $\mathbb{E}$ tiene una base formada por dos ciclos $Z_\lambda(v)$ y $Z_\lambda(w)$ con r como longitud de $Z_\lambda(w)$?. El lector comprobará como en el caso anterior que la matriz de L es

$$J_\lambda(p,q) = \begin{pmatrix} J_\lambda(q) & \\ & J_\lambda(r) \end{pmatrix} \in \mathbb{k}^{(p+r)\times(p+r)} \text{ con ceros fuera de las } J_\bullet(\bullet).$$

Al ser $J_\lambda(q)$ y $J_\lambda(p,q)$ triangulares con todo λ en la diagonal, queda claro que estas matrices son posibles solo si L tiene un único valor propio λ. No obstante, aunque L tenga solamente el valor propio λ, podría suceder (y sucede) que $\mathbb{E}$ tenga una base formada yuxtaponiendo o alineando k ciclos en la forma

$$(Z_\lambda(v_1), \ldots, Z_\lambda(v_r)) = \left(\left(L_\lambda^0(v_1), L_\lambda^1(v_1), \ldots, L_\lambda^{q_1-1}(v_1)\right), \ldots, \left(L_\lambda^0(v_k), L_\lambda^1(v_k), \ldots, L_\lambda^{q_k-1}(v_k)\right)\right),$$

en cuyo caso la matriz sería una generalización de $J_\lambda(p,q)$ para la que introducimos la notación

$$J_\lambda(q_1, q_2, \ldots, q_k) = \begin{pmatrix} J_\lambda(q_1) & & & \\ & J_\lambda(q_2) & & \\ & & \ddots & \\ & & & J_\lambda(q_k) \end{pmatrix} \in \mathbb{k}^{(q_1+q_2+\ldots+q_k)\times(q_1+q_2+\ldots+q_k)}.$$

Las matrices $J_\lambda(q)$ se llaman **bloque simple de Jordan (del valor λ)** y las matrices $J_\lambda(q_1, q_2 \ldots, q_k)$ **bloques de Jordan (del valor λ)**. Una matriz formada con bloques de Jordan, quizás con diferentes λ, en posición diagonal es una **matriz de Jordan.**

Lo esencial del **teorema de Jordan** es que se pueden tomar ciclos (con diversos v) de modo que al yuxtaponerlos formen base, las llamadas **bases de Jordan**. En ellas, L tendrá una matriz de la forma

$$J = \begin{pmatrix} J_{\lambda_1}\left(q_{(1)1}, \ldots, q_{(1)k_1}\right) & & & & \\ & \ddots & & & \\ & & J_{\lambda_h}\left(q_{(h)1}, \ldots, q_{(h)k_h}\right) & & \\ & & & \ddots & \\ & & & & J_{\lambda_p}\left(q_{(p)1}, \ldots, q_{(p)k_p}\right) \end{pmatrix}$$

formada por bloques de Jordan. Si fijamos desde el principio una numeración $(\lambda_1, \ldots, \lambda_p)$ para los valores propios, exigimos que en cada $J_{\lambda_h}\left(q_{(h)1}, \ldots, q_{(h)k_h}\right)$ sea $q_{(h)1} \geq q_{(h)2} \geq \ldots \geq q_{(h)k_h}$, la matriz J es única (no así la base de Jordan). La unicidad de J se basa en que todo lo que en ella interviene (los $\lambda_h \in \mathbb{k}$ y los $q_{(h)j} \in \mathbb{N}$) son entes directamente calculables a partir de L.

7.2. La descomposición primaria

Recordemos la factorización (7.1) para $C(X)$ y $M(X)$. Como $L \circ (L-\lambda_h)^{s_h} = (L-\lambda_h)^{s_h} \circ L$ tenemos para $x \in \ker (L-\lambda_h)^{s_h}$ que $((L-\lambda_h)^{s_h} \circ L)(x) = (L \circ (L-\lambda_h)^{s_h})(x) = L(0) = 0$, luego $L(\ker (L-\lambda_h)^{s_h}) \subset \ker (L-\lambda_h)^{s_h}$ y, fijado h de una vez por todas, hay una función inducida por restricción de L de $\ker (L-\lambda_h)^{s_h}$ en sí mismo, que llamaremos R. Importa saber como son los polinomios minimal y característico de R.

Teorema 135 *Con las notaciones expuestas tenemos que se descompone en suma directa*

$$\mathbb{E} = \ker (L-\lambda_1)^{s_1} \oplus \ldots \oplus \ker (L-\lambda_p)^{s_p}.$$

Los polinomios minimal y característico $M_h(X)$ y $C_h(X)$ de la restricción R de L a $\ker (L-\lambda_h)^{s_h}$ son $(L-\lambda_h)^{s_h}$ y $(\lambda_h - X)^{m_h}$, siendo además $\dim \ker (L-\lambda_h)^{s_h} = m_h$.

Demostración. Ya se probó la descomposición en el teorema 131.

Sea $P(X) = (X-\lambda_h)^{s_h}$, que cumple por definición de $\ker (L-\lambda_h)^{s_h}$ que $P(R) = 0$. El polinomio minimal $M_h(X)$ de R ha de dividir a $P(X)$ (teorema 107), luego será de la forma $M_h(X) = (X-\lambda_h)^k$ con $k \leq s_h$. Si fuese $k < s_h$ consideraríamos

$$Q(X) = (X-\lambda_1)^{s_1} \cdots (X-s_{h-1})^{s_{h-1}} (X-\lambda_h)^k \left(X-\lambda_{s_{h+1}}\right) \cdots (X-\lambda_p)^{s_p},$$

y $x \in \mathbb{E}$ arbitrario. Descompondríamos $x = x_1 + \ldots + x_p$ con $x_j \in \ker (L-\lambda_j)^{s_j}$ y

$$Q(L)(x) = Q(L)(x_1) + \ldots + Q(L)(x_h) + \ldots + Q(L)(x_p).$$

Para $j \neq h$ es $Q(L)(x_j) = 0$ porque podemos permutar los $(L-\lambda_i)$ de modo que $(L-\lambda_j)^{s_j}$ quede en primer lugar a la derecha junto a x_j, con lo que $(L-\lambda_j)^{s_j}(x_j) = 0$ y $Q(L)(x_j) = 0$. Con esto nos reducimos a $Q(L)(x) = Q(L)(x_h)$, pero, de modo similar, ponemos a la derecha $(L-\lambda_h)^k$ y ahora $(X-\lambda_h)^k(x_j) = 0$ por ser $(X-\lambda_h)^k$ el polinomio minimal de la restricción a $\ker (L-\lambda_h)^{s_h}$. Como x es arbitrario y $Q(L)(x) = 0$, tenemos contradicción, ya que $\deg Q(X) < \deg M(X)$.

El polinomio característico ha de tener las mismas raíces que el minimal (teorema 107), luego si $M_h(X) = (X-\lambda_h)^{s_h}$, deberá ser $C_h(X) = (\lambda_h - X)^{d_h}$ con d_h por determinar. Con carácter general, si $\mathbb{E} = \mathbb{F}_1 \oplus \ldots \oplus \mathbb{F}_k$ y cada $\mathbb{F}_i$ es estable por L, se tiene que $C_L(X) = C_1(X) \cdots C_k(X)$, siendo $C_i(X)$ el polinomio característico de la restricción de L a un endomorfismo de $\mathbb{F}_i$. Aplicado esto a nuestro caso, $C_L(X) = (\lambda_1 - X)^{d_1} \cdots (\lambda_p - X)^{d_p}$. Como la factorización de un polinomio es única, comparando con (7.1), $d_i = m_i$. Después de saber que $C_h(X) = (\lambda_h - X)^{m_h}$, usamos que $\dim \ker (L-\lambda_h)^{s_h} = \deg C_h(X) = m_h$ y hemos acabado. ♣

Teorema 136 *Para $i = 1, \ldots, p$ es $\ker (L-\lambda_1)^{s_1} = \ker (L-\lambda_1)^{m_1}$ luego podemos descomponer también*

$$\mathbb{E} = \ker (L-\lambda_1)^{m_1} \oplus \ldots \oplus \ker (L-\lambda_p)^{m_p}.$$

Demostración. La segunda afirmación es obvia tras la primera por el teorema 135. Obviamente $\ker (L-\lambda_h)^{s_h} \subset \ker (L-\lambda_h)^{m_h}$ para $h = 1, \ldots, p$. Probaremos que $\ker (L-\lambda_i)^{m_i} \cap \ker (L-\lambda_j)^{m_j} = 0$ si $i \neq j$. Habría entonces una suma directa $\ker (L-\lambda_1)^{m_1} \oplus \ldots \oplus \ker (L-\lambda_p)^{m_p} = \mathbb{S}$. Lo que cuenta no es que pueda ser o no $\mathbb{S} = \mathbb{E}$ sino que $\dim(\mathbb{S}) = \sum_{h=1}^p \dim \ker (L-\lambda_h)^{m_h} \leq n$. Entonces

$$n = \sum_{h=1}^p m_h = \sum_{h=1}^p \dim \ker (L-\lambda_h)^{s_h} \leq \sum_{h=1}^p \dim \ker (L-\lambda_h)^{m_h} = \dim \mathbb{S} \leq n$$

porque $\dim \ker (L-\lambda_h)^{s_h} = m_h \leq \dim \ker (L-\lambda_h)^{m_h}$. para algún h la desigualdad fuese estricta se obtendría $n < n$. Así pues $\ker (L-\lambda_h)^{s_h} \subset \ker (L-\lambda_h)^{m_h}$ pero con la misma dimensión m_h, luego son iguales.

Veamos pues para $i \neq j$ que $\ker (L-\lambda_i)^{m_i} \cap \ker (L-\lambda_j)^{m_j} = 0$. Más generalmente, tomemos x tal que $(L-\lambda_i)^p(x) = (L-\lambda_j)^q(x) = 0$ con $p \geq s_i$ y $q \geq s_j$. Entonces,

$$((L-\lambda_i) - (L-\lambda_j))^{p+q}(x) = \sum_{k=0}^{p+q} \binom{p+q}{k} (-1)^k (L-\lambda_i)^{(p+q)-k} \circ (L-\lambda_j)^k(x) = 0$$

ya que cada sumando es nulo. En efecto, o bien $(p+q)-k \geq p$ o $k \geq q$. Si es $k \geq q$, $(L-\lambda_j)^k(x)=0$, y si es $(p+q)-k \geq p$ permutamos (es posible porque los factores son polinomios en L) y

$$(L-\lambda_i)^{(p+q)-k} \circ (L-\lambda_j)^k(x) = (L-\lambda_j)^k \circ (L-\lambda_i)^{(p+q)-k}(x) = (L-\lambda_j)^k(0) = 0.$$

Pero por otro lado, $((L-\lambda_i)-(L-\lambda_j))^{p+q}(x) = (\lambda_j-\lambda_i)^{p+q}(x)$ y, $\lambda_j-\lambda_i \neq 0$ implica que $x=0$. ♣

Las descomposiciones en suma directa de los teorema 135 y 136 (es la misma expresada de dos modos) se llaman la **descomposición primaria** (de L). Adelantamos que los ciclos $Z_\lambda(v)$ que van a formar la base de Jordan se eligen dentro de los diversos $\ker(L-\lambda_j)^{s_j} = \ker(L-\lambda_j)^{m_j}$. Vamos a ver que lo fundamental es estudiar la siguiente situación que es un caso particular de nuestra búsqueda de la base de Jordan: Sea $\mathbb{V}$ un espacio vectorial y $N \neq 0$ un endomorfismo nilpotente de $\mathbb{V}$; o sea, tal que existe un primer $s \geq 1$ tal que $N^s = 0$. *Afirmamos que hay una base de Jordan para* N. Si tomamos como $\mathbb{V}$ y N los diversos $\ker(L-\lambda_i)^{s_i} = \ker(L-\lambda_i)^{m_i}$ y $L-\lambda_i = L_{\lambda_i}$, tendremos bases de Jordan $\mathcal{J}_i$ y la base $\mathcal{J}$ que resulta al yuxtaponerlas es una base de Jordan para todo L. Los detalles se dan más adelante tras estudiar en abstracto el caso nilpotente.

7.3. El caso nilpotente

En esta sección $N : \mathbb{V} \to \mathbb{V}$ es un endomorfismo nilpotente no nulo y $s \in \mathbb{N}$ es el primer exponente tal que $N^s = 0$, llamado **índice de nilpotencia de** N. Es muy fácil ver que $\lambda = 0$ es el único valor propio posible porque $N(v) = \lambda v$ con $v \neq 0$ da $0 = N^s(v) = \lambda^s v$. Definimos $\mathbb{K}^i = \ker\left(N^i\right)$ para $i = 0, 1, \ldots, s$. Obsérvese que $\mathbb{K}^0 = \ker(\mathrm{id}) = 0$, $\mathbb{K}^1 = \ker(N)$ y $\mathbb{K}^s = \ker(N^s) = \ker(0) = \mathbb{V}$. Damos varias propiedades sencillas pero básicas sobre los $\mathbb{K}^i$.

Teorema 137 *Los subespacios* $\mathbb{K}^0 = 0, \mathbb{K}^1, \ldots, \mathbb{K}^{s-1}, \mathbb{K}^s = \mathbb{V}$ *verifican*

1. *Forman una sucesión estrictamente creciente para* $\subset$; *es decir,*

$$\mathbb{K}^0 = \ker\left(N^0\right) = 0 \subset \mathbb{K}^1 = \ker\left(N^1\right) \subset \mathbb{K}^2 = \ker\left(N^2\right) \subset \ldots \subset \mathbb{K}^s = \ker\left(N^s\right) = \mathbb{V}$$

   y $\mathbb{K}^{i-1} \neq \mathbb{K}^i$ *para* $i = 1, \ldots, s$.

2. N *transporta en sentido inverso*

$$\mathbb{K}^0 = 0 \xleftarrow{N} \mathbb{K}^1 \xleftarrow{N} \mathbb{K}^2 \xleftarrow{N} \ldots \xleftarrow{N} \mathbb{K}^{i-1} \xleftarrow{N} \mathbb{K}^i \xleftarrow{N} \cdots \xleftarrow{N} \mathbb{K}^s = \mathbb{V},$$

   pero si $x \in \mathbb{K}^i - \mathbb{K}^{i-1}$ $(2 \leq i \leq s)$ *no puede suceder que* $N(x)$ *esté en* $\mathbb{K}^{i-2}$. *En particular, si* $\mathbb{T}^i$ *es un suplementario de* $\mathbb{K}^{i-1}$ *en* $\mathbb{K}^i$; *o sea, si* $\mathbb{K}^i = \mathbb{K}^{i-1} \oplus \mathbb{T}^i$, *se cumple que la restricción de* N *a* $\mathbb{T}^i$ *es inyectiva.*

3. *Los elementos* v *en* $\mathbb{K}^i - \mathbb{K}^{i-1}$ $(2 \leq i \leq s)$ *sirven como elementos iniciales de ciclos* $Z(v)$ *de longitud* i.

Demostración. Sin duda $N^i(x) = 0$ implica que $N^{i+1}(x) = N\left(N^i(x)\right) = N(0) = 0$ así que $\mathbb{K}^i \subset \mathbb{K}^{i+1}$. Si fuese $\mathbb{K}^{i-1} = \mathbb{K}^i$ para cierto $i \leq s$ tomaríamos $x \in \mathbb{V}$ arbitrario y se tendría $0 = N^s(x) = N^i\left(N^{s-i}(x)\right)$ luego $\left(N^{s-i}(x)\right) \in \mathbb{K}^i$, pero $\mathbb{K}^i = \mathbb{K}^{i-1}$, así que $0 = N^{i-1}\left(N^{s-i}(x)\right) = N^{s-1}(x)$. Al ser x arbitrario, el índice de N no sería s sino $s-1$. Esta contradicción prueba **1.**

Si $x \in \mathbb{K}^i$, la condición $N^i(x) = 0$ se escribe como $N^{i-1}(N(x)) = 0$, luego $N(x) \in \mathbb{K}^{i-1}$. Si $x \in \mathbb{K}^i - \mathbb{K}^{i-1}$ que verificase $N(x) \in \mathbb{K}^{i-2}$, tendríamos $0 = N^{i-2}(N(x)) = N^{i-1}(x)$ luego $x \in \mathbb{K}^{i-1}$ que es contradictorio.[4] Si $x \neq 0$ está en $\mathbb{T}^i$, debe cumplirse $x \in \mathbb{K}^i - \mathbb{K}^{i-1}$ y si fuese $N(x) = 0$ se tendría en particular $N(x) \in \mathbb{K}^{i-2}$ y $x \in \mathbb{K}^{i-1}$. Esta contradicción prueba **2.**

Queda **3**. Si $v \in \mathbb{K}^i - \mathbb{K}^{i-1}$, por **1**, $N(v) \in \mathbb{K}^{i-1}$, y por **2**, $N(v) \notin \mathbb{K}^{i-2}$ así que $N(v) \in \mathbb{K}^{i-1} - \mathbb{K}^{i-2}$. Repitiendo el razonamiento con $N(v)$ vemos que $N^2(v) \in \mathbb{K}^{i-2} - \mathbb{K}^{i-3}$ hasta llegar paso a paso a

[4] Estos razonamientos, aunque rápidos, resultan muy resbaladizos pues parece que nos están explicando una cosa y su contraria. Probablemente hay una tendencia a pensar que cuanto mayor es i en $\mathbb{K}^i$, más cerca se está "de que $x \in \mathbb{K}^i$ sea nulo" y es justo al contrario: puede que necesitemos aplicar una potencia tan alta como la i para llevar x a cero. Igualmente N lleva $\mathbb{K}^i$ en $\mathbb{K}^{i-1}$ y no a $\mathbb{K}^{i+1}$ porque cuanto más alto sea j en $N^j(x)$ menor es la potencia de N necesaria para llevar $N^j(x)$ a cero.

$N^{i-1}(v) \in \mathbb{K}^1 - \mathbb{K}^0 = \ker N$. Esto último implica que $N^{i-1}(v) \neq 0$ y $N^i(v) = 0$, luego $Z(v)$ es un ciclo de longitud i. ♣

Definimos $n^i = \operatorname{nul}\left(N^i\right) = \dim\left(\ker\left(N^i\right)\right)$. El que los $\mathbb{K}^i$ formen sucesión estrictamente creciente implica $n^0 < n^1 < \ldots < n^{s-1} < n^s = \dim(\mathbb{V})$. Hay otras desigualdades que derivan del teorema 137.

Teorema 138 *La sucesión* $\left(n^s - n^{s-1}, n^{s-1} - n^{s-2}, \ldots, n^2 - n^1, n^1 - n^0\right)$ *es creciente.*

Demostración. Probamos que $n^{i+1} - n^i \leq n^i - n^{i-1}$. Hay un artificio ingenioso. Como $\mathbb{K}^{i-1} \subset \mathbb{K}^i$ podemos encontrar un subespacio $\mathbb{T}^i$ tal que $\mathbb{K}^i = \mathbb{K}^{i-1} \oplus \mathbb{T}^i$. Vimos en **2** del teorema 137 que la restricción de N a $\mathbb{T}^i$ es inyectiva y $N\left(\mathbb{T}^i\right) \subset \mathbb{K}^{i-1}$ cumple $N\left(\mathbb{T}^i\right) \cap \mathbb{K}^{i-2} = 0$. Así pues, $\dim N\left(\mathbb{T}^i\right) + \dim \mathbb{K}^{i-2} \leq \dim\left(\mathbb{K}^{i-1}\right)$ y

$$n^i - n^{i-1} = \dim\left(\mathbb{T}^i\right) = \dim\left(N\left(\mathbb{T}^i\right)\right) \leq \dim\left(\mathbb{K}^{i-1}\right) - \dim\left(\mathbb{K}^{i-2}\right) = n^{i-1} - n^{i-2}.$$

Un comentario: la "demostración" del decrecimiento basada en restar $n^i < n^{i+1}$ y $n^{i-1} < n^i$ es *errónea*. La suma de dos desigualdades con la misma orientación da otra desigualdad pero es falsa la desigualdad al restarlas pues $1 < 3$ y $1 < 2$ no implica $1 - 1 < 2 - 3$. ♣

Tomemos, por ahora de modo arbitrario, suplementarios $\mathbb{T}^i$ de $\mathbb{K}^{i-1}$ en $\mathbb{K}^i$; o sea, $\mathbb{K}^i = \mathbb{K}^{i-1} \oplus \mathbb{T}^i$ para $i = s, s-1, \ldots, 2, 1$. Por el teorema 138, $\dim\left(\mathbb{T}^i\right) = t_i = n^i - n^{i-1}$ y la sucesión $(t_s, t_{s-1}, \ldots, t_2, t_1)$ es creciente. (Conviene contemplar las sucesiones $\left(\mathbb{T}^i\right)$ y (t_i) *en orden inverso.*) La base de Jordan va a ser construida yuxtaponiendo bases $\mathcal{B}_s, \mathcal{B}_{s-1}, \ldots, \mathcal{B}_2, \mathcal{B}_1$ de ciertos $\mathbb{T}_s, \mathbb{T}_{s-1}, \ldots, \mathbb{T}_2, \mathbb{T}_1$ que son suplementarios, verificando

$$\mathbb{V} = \mathbb{K}^s = \mathbb{K}^{s-1} \oplus \mathbb{T}^s,\ \mathbb{K}^{s-1} = \mathbb{K}^{s-2} \oplus \mathbb{T}^{s-1},\ \ldots\ \mathbb{K}^i = \mathbb{K}^{i-1} \oplus \mathbb{T}^i,\ \ldots\ , \mathbb{K}^2 = \mathbb{K}^1 \oplus \mathbb{T}^2,\ \mathbb{K}^1 = \mathbb{K}^0 \oplus \mathbb{T}^1.$$

En general, yuxtaponiendo sucesiones de vectores independientes, como son los $\mathcal{B}_i$, no hay garantía de que lo sea $\mathcal{B} = (\mathcal{B}_s, \mathcal{B}_{s-1}, \ldots, \mathcal{B}_2, \mathcal{B}_1)$, sin embargo en este caso no hay problema.

Teorema 139 *En las condiciones descritas* $\mathcal{B}$ *es base de* $\mathbb{V}$.

Demostración. Se demuestra con otro artificio curioso. Con sucesivas sustituciones,

$$\begin{aligned}\mathbb{K}^s &= \mathbb{K}^{s-1} \oplus \mathbb{T}^s = \mathbb{K}^{s-2} \oplus \mathbb{T}^{s-1} \oplus \mathbb{T}^s = \mathbb{K}^{s-3} \oplus \mathbb{T}^{s-2} \oplus \mathbb{T}^{s-1} \oplus \mathbb{T}^s \\ &= \ldots = \mathbb{K}^{s-j} \oplus \mathbb{T}^{s-j+1} \oplus \mathbb{T}^{s-j+2} \oplus \ldots \oplus \mathbb{T}^{s-1} \oplus \mathbb{T}^s = \mathbb{K}^0 \oplus \mathbb{T}^1 \oplus \mathbb{T}^2 \oplus \ldots \oplus \mathbb{T}^{s-1} \oplus \mathbb{T}^s\end{aligned}$$

y como $\mathbb{K}^0 = \ker N^0 = \ker(\mathrm{id}) = 0$ y $\mathbb{K}^s = \ker N^s = \ker(0) = \mathbb{V}$, queda $\mathbb{V} = \mathbb{T}^1 \oplus \mathbb{T}^2 \oplus \ldots \oplus \mathbb{T}^{s-1} \oplus \mathbb{T}^s$. Cada $\mathcal{B}_i$ es base de un sumando de una suma directa, y su yuxtaposición da una base de $\mathbb{V}$. ♣

La base de Jordan $\mathcal{J}$ va a construirse con bases $\mathcal{J}_i$ de bien elegidos suplementarios $\mathbb{T}_i$. Advertencia importante: *en lugar de numerar secuencialmente usaremos dobles índices*, porque $\left(v_1^i, \ldots, v_{t_i}^i\right)$ será base de $\mathbb{T}^i$ (pero no solo eso).

Teorema 140 *Existe una base de Jordan* $\mathcal{J}$ *cuyos elementos* v_j^i *dispuestos en una tabla de altura* s*, el índice de* N*, y anchura* $t_1 = \dim\ker N$*, tal como*

v_1^s	$\cdots$	$v_{t_s}^s$										
v_1^{s-1}	$\cdots$	$v_{t_s}^{s-1}$	$\cdots$	$v_{t_{s-1}}^{s-1}$								
v_1^{s-2}	$\cdots$	$v_{t_s}^{s-2}$	$\cdots$	$v_{t_{s-1}}^{s-2}$	$\cdots$	$v_{t_{s-2}}^{s-2}$						
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$						
v_1^i	$\cdots$	$v_{t_s}^i$	$\cdots$	$v_{t_{s-1}}^i$	$\cdots$	$v_{t_{s-2}}^i$	$\cdots$	$\cdots$	$v_{t_i}^i$			
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$			
v_1^1	$\cdots$	$v_{t_s}^1$	$\cdots$	$v_{t_{s-1}}^1$	$\cdots$	$v_{t_{s-2}}^1$	$\cdots$	$\cdots$	$v_{t_i}^1$	$\cdots$	$\cdots$	$v_{t_1}^1$

, (7.2)

verifican que (¡ojo! las filas se numeran de abajo arriba*)*

1. *La sucesión* $\left(v_1^i, \ldots, v_{t_i}^i\right) = \mathcal{J}_i$ *de la fila* i *es base de un suplementario* $\mathbb{T}_i$ *tal que* $\mathbb{K}^i = \mathbb{K}^{i-1} \oplus \mathbb{T}^i$.
2. *Si* v_j^i *es tal que* v_j^{i+1} *está definido,*[5] $N\left(v_j^{i+1}\right) = v_j^i$.

[5] Se entiende que en la tabla no están ocupados todos los huecos y que los huecos ocupados forman una escalera que baja de izquierda a derecha.

Demostración. Vamos a construir $\mathcal{J}$ por filas, de arriba abajo. Para $\left(v_1^s,\ldots,v_{t_s}^s\right)=\mathcal{J}_s$ tomamos un suplementario cualquiera $\mathbb{T}_s$ tal que $\mathbb{K}^s=\mathbb{K}^{s-1}\oplus\mathbb{T}^s$ y una base $\left(v_1^s,\ldots,v_{t_s}^s\right)$ de él. Para $\mathcal{J}_{s-1}$ empezamos observando que, por **2** en el teorema 137, los vectores $\left(N\left(v_1^s\right),\ldots,N\left(v_{t_s}^s\right)\right)$ son independientes, porque N restringida a $\mathbb{T}^s$ es inyectiva. Denotaremos esta sucesión alternativamente como $\left(v_1^{s-1},\ldots,v_{t_s}^{s-1}\right)$. Se tiene $N\left(\mathbb{T}^s\right)\cap\mathbb{K}^{s-2}=0$ (**2** en el teorema 137) pero puede no ser $N\left(\mathbb{T}^s\right)\oplus\mathbb{K}^{s-2}=\mathbb{K}^{s-1}$. No obstante, se puede ampliar $\left(v_1^{s-1},\ldots,v_{t_s}^{s-1}\right)$ a $\left(v_1^{s-1},\ldots,v_{t_s}^{s-1},v_{t_s+1}^{s-1},\ldots,v_{t_{s-1}}^{s-1}\right)=\mathcal{J}_{s-1}$ (por definición) de modo que si $\mathbb{T}^{s-1}=\lg\left(\mathcal{J}_{s-1}\right)$ se cumpla que $\mathbb{K}^{s-1}=\mathbb{K}^{s-2}\oplus\mathbb{T}^{s-1}$. Claramente se cumple **2** porque si v_j^{s-1} es tal que v_j^s esta definido, es porque $v_j^{s-1}=N\left(v_j^s\right)$.

Supongamos construidas bases $\mathcal{J}_s,\mathcal{J}_{s-1},\ldots,\mathcal{J}_{i+1}$ y suplementarios $\mathbb{T}^s,\ldots,\mathbb{T}^{i+1}$ con las propiedades enunciadas, y vamos a construir $\mathcal{J}_i$ y $\mathbb{T}^i$. Aplicamos N a $\mathcal{J}^{i+1}=\left(v_1^{i+1},\ldots,v_{t_{s+1}}^{i+1}\right)$ base de $\mathbb{T}^{i+1}$ y renombramos $\left(N\left(v_1^{i+1}\right),\ldots,N\left(v_{t_{i+1}}^{i+1}\right)\right)=\left(v_1^i,\ldots,v_{t_{i+1}}^i\right)$. Se tiene $N\left(\mathbb{T}^{i+1}\right)\cap\mathbb{K}^{i-1}=0$ (**2** en el teorema 137) pero puede no ser $N\left(\mathbb{T}^{i+1}\right)\oplus\mathbb{K}^{i-1}=\mathbb{K}^i$. No obstante, se puede ampliar $\left(v_1^i,\ldots,v_{t_{i+1}}^i\right)$ a $\left(v_1^i,\ldots,v_{t_{i+1}}^i,v_{t_{i+1}+1}^i,\ldots,v_{t_i}^i\right)=\mathcal{J}_i$ (por definición) de modo que si $\mathbb{T}^i=\lg\left(\mathcal{J}_i\right)$ se cumpla que $\mathbb{K}^i=\mathbb{K}^{i-1}\oplus\mathbb{T}^i$. Claramente se cumple **2** porque si v_j^i es tal que v_j^{i+1} esta definido, es porque $v_j^i=N\left(v_j^{i+1}\right)$.

Por el teorema 139, la yuxtaposición $\mathcal{J}$ de $\mathcal{J}_s,\ldots,\mathcal{J}_1$ es una base. Además es base de Jordan. Para verlo, hay que detectar los ciclos, y es aquí donde se ve el porqué hemos elegido expresar $\mathcal{J}$ en forma de cuadro y en escalera. Los ciclos son las *columnas de la tabla (¡no las filas!)* $\mathcal{J}_s,\ldots,\mathcal{J}_1$. En efecto, tomamos v_j^1 en la primera fila y el v_j^i sobre él en la tabla a altura máxima. Se obtiene

$$v_j^1=N\left(v_j^2\right),\ v_j^2=N\left(v_j^3\right),\ v_j^3=N\left(v_j^4\right),\ \ldots\ ,v_j^{i-1}=N\left(v_j^i\right),$$

y sustituyendo,

$$\begin{array}{l}v_j^1=N\left(v_j^2\right)=N^2\left(v_j^3\right)=N^3\left(v_j^4\right)=\ldots=N^{i-1}\left(v_j^i\right),\\ v_j^2=N\left(v_j^3\right)=N^2\left(v_j^4\right)=\ldots=N^{i-2}\left(v_j^i\right),\\ \vdots\\ v_j^{i-1}=N\left(v_j^i\right).\end{array}$$

Por tanto, $Z\left(v_j^i\right)=\left(v_j^i,N\left(v_j^i\right),\ldots,N^{i-2}\left(v_j^i\right),N^{i-2}\left(v_j^i\right)\right)$ es la columna j del cuadro leída de arriba abajo. ♣

7.4. La tabla $*$ de una base de Jordan (caso nilpotente)

La forma de presentar en tabla la base de Jordan como en el teorema 140, que llamaremos en lo sucesivo la **tabla** $*$ o **tabla de estrellas**, tiene al menos dos importantes consecuencias. **(a)** Si solo nos importa conocer la matriz de Jordan de N pero no la base de Jordan concreta, hay posibilidad de hacerlo con un ahorro considerable de trabajo a base de calcular tan solo rangos y nulidades de las potencias N^i; y **(b)** Si bien hay muchas bases de Jordan para N, la matriz de Jordan $J_0\left(q_1,\ldots,q_k\right)$ es única si acordamos colocar las cajas de modo que $q_1\geq q_2\geq\ldots\geq q_k$.

Abordamos **(a)**. Vemos la tabla como una escalera que baja de izquierda a derecha. Como $t^s\leq t^{s-1}<\ldots\leq t^2\leq t^1$, los t^i nos dan la anchura de los peldaños, y como puede ser $t^i=t^{i-1}$, la altura de los peldaños puede variar. Primero dos ejemplos

v_1^3	
$v_1^2=N\left(v_1^3\right)$	v_2^2
$v_1^1=N\left(v_1^2\right)=N^2\left(v_1^3\right)$	$v_2^1=N\left(v_2^2\right)$

,

v_1^4	v_2^4		
$v_1^3=N\left(v_1^4\right)$	$v_2^3=N\left(v_2^4\right)$		
$v_1^2=N^2\left(v_1^4\right)$	$v_2^2=N^2\left(v_2^4\right)$		
$v_1^1=N^3\left(v_1^4\right)$	$v_2^1=N^3\left(v_2^4\right)$	v_3^1	v_4^1

,

que, si solo ponemos los huecos no vacíos, se simplifican a

Tabla 1

$*$	
$*$	$*$
$*$	$*$

, Tabla 2

$*$	$*$		
$*$	$*$		
$*$	$*$		
$*$	$*$	$*$	$*$

.

Ahora veremos, y es la clave, que si solo nos interesa la matriz de Jordan, pero no la base de Jordan que la crea, la forma esquemática "con estrellas" es suficiente. Si tratamos el primer ejemplo, tenemos que $n = 5$ (porque hay 5 estrellas y la base de Jordan, como cualquier otra base, tiene 5 elementos), y dos ciclos, de longitudes 3 y 2 (porque la primera columna tiene altura 3 y la otra altura 2), luego la matriz de Jordan será $J_0(3,2)$. En el otro ejemplo, es $n = 10$, habiendo 2 ciclos de longitud 4 y 2 de longitud 1. La matriz de Jordan será

$$J_0(4,4,2,2) = \begin{pmatrix} 0 & & & & & & & & & \\ 1 & 0 & & & & & & & & \\ & 1 & 0 & & & & & & & \\ & & 1 & 0 & & & & & & \\ & & & & 0 & & & & & \\ & & & & 1 & 0 & & & & \\ & & & & & 1 & 0 & & & \\ & & & & & & 1 & 0 & & \\ & & & & & & & & 0 & \\ & & & & & & & & & 0 \end{pmatrix}$$

y se entiende que los coeficientes que no se precisan son 0.

En el caso general se cuentan las estrellas en las columnas de la tabla. Si hay t^s columnas de altura s, hay t^s ciclos de longitud s. Si no hay peldaños de altura i no hay ciclos de longitud i en la base de Jordan. (En el ejemplo de la tabla 1 no hay ciclos de longitud 1 pero sí de longitud 2, y sucede al revés en la tabla 2.) Si hubiese peldaños de todas las alturas $1, 2, \ldots, s$, habría ciclos de todas las longitudes $1, \ldots, s$. Si tenemos $t_{i+1} < t_i$ el peldaño a altura i tiene anchura $t_i - t_{i-1}$ y el número de ciclos de longitud i es $t_i - t_{i-1}$. Como vemos, la tabla de estrellas permite saber cuántos ciclos hay y de qué longitudes, lo que determina J.

¿Cómo construir la tabla de estrellas? Se tiene nada más conocer $(t_s, t_{s-1}, \ldots, t_2, t_1)$. Recordemos que

$$t^i = \dim \mathbb{T}^i = \dim \mathbb{K}^i - \dim \mathbb{K}^{i-1} = n^i - n^{i-1} = \operatorname{nul} N^i - \operatorname{nul} N^{i-1}$$

luego bastará calcular las potencias $N^0 = \mathrm{id}$, $N^1 = N$, N^2, $\ldots$ $N^s = 0$ y sus nulidades $n^0 = 0$, n^1, $\ldots$ $n^s = n = \dim \mathbb{V}$. Posiblemente sea más fácil calcular los rangos $r^i = \operatorname{rg}\left(N^i\right)$ y como por el teorema del rango-nulidad es $n^i = n - r^i$, obtenemos

$$t^i = n^i - n^{i-1} = \left(n - r^i\right) - \left(n - r^{i-1}\right) = r^{i-1} - r^i.$$

Teorema 141 *Sean $\left(n^0, n^1, \ldots, n^{s-1}, n^s\right)$ y $\left(r^0, r^1, \ldots, r^{s-1}, r^s\right)$ las sucesiones de nulidades y rangos de las potencias N^i. Los números t^i que dan el número de estrellas en la fila i (se numera de arriba abajo con $s, s-1, \ldots, 1$) de la tabla de la base de Jordan de N están dados por*

$$\begin{cases} t^s = n^s - n^{s-1} \\ t^{s-1} = n^{s-1} - n^{s-2} \\ \vdots \\ t^1 = n^1 - n^0 \end{cases} \quad \textit{o bien} \quad \begin{cases} t^s = r^{s-1} - r^s \\ t^{s-1} = r^{s-2} - r^{s-3} \\ \vdots \\ t^1 = r^0 - r^1 \end{cases},$$

teniendo en cuenta que $n^s = r^0 = \dim(\mathbb{V})$ y $n^0 = r^s = 0$.

Como consecuencia, la matriz de Jordan $J_0(q_1, \ldots, q_k)$ con $q_1 \geq \ldots \geq q_k$ viene determinada exclusivamente por N.

Demostración. Solo queda probar la unicidad. Los q^i en $J_0(q_1, \ldots, q_k)$ dependen de la distribución de estrellas en la tabla, y esta distribución de cómo sean los t^i. Pero los t^i dependen de los n^i y estos solo de N. Así pues la matriz de Jordan, con la distribución de cajas referida, viene unívocamente determinada por N. ♣

Problema 263 *Identificamos $N : \mathbb{R}^5 \to \mathbb{R}^5$ con la matriz N, siendo*

$$N = \begin{pmatrix} -1 & 0 & -1 & 1 & 0 \\ -4 & -1 & -3 & 2 & 1 \\ -2 & -1 & -2 & 1 & 1 \\ -3 & -1 & -3 & 2 & 1 \\ -8 & -2 & -7 & 5 & 2 \end{pmatrix}, \quad N^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 \\ -1 & 0 & -1 & 1 & 0 \end{pmatrix}, \quad N^3 = 0.$$

Determinar la matriz de Jordan J y una base de Jordan $\mathcal{J}$. ♦

Solución. Tenemos $n = 5$ y $s = 3$. Los rangos de N y N^2 son $r^1 = 3$ y $r^2 = 1$. Las t^i y tabla de estrellas son

$$\left\{\begin{array}{l} t^3 = r^2 - r^3 = 1 - 0 = 1 \\ t^2 = r^1 - r^2 = 3 - 1 = 2 \\ t^1 = r^0 - r^1 = 5 - 3 = 2 \end{array}\right. , \qquad \begin{array}{|c|c|}\hline * & \\ \hline * & * \\ \hline * & * \\ \hline \end{array} = \begin{array}{|c|c|}\hline v_1^3 & \\ \hline v_1^2 & v_2^2 \\ \hline v_1^1 & v_2^1 \\ \hline \end{array}$$

La base de Jordan está compuesta por dos ciclo $Z\left(v_1^3\right)$ y $Z\left(v_2^2\right)$ de longitudes 3 y 2 y debe ser

$$J = J_0\left(3,2\right) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Elegimos v_1^3 tal que si $\mathbb{T}^3 = \lg\left(v_1^3\right)$ se tenga $\mathbb{K}^3 = \mathbb{K}^2 \oplus \mathbb{T}^3$. Esto se consigue tomando $v_1^3 \in \mathbb{K}^3 - \mathbb{K}^2$ y tomamos como buena elección $v_1^3 = e_1 = (1,0,0,0,0)^\top$. Decimos "buena elección" porque $\left(v_1^3, N\left(v_1^3\right), N^2\left(v_1^3\right)\right)$ serán los tres primeros vectores de $\mathcal{J}$ y, como ya están calculados N y N^2, al ser $v_1^3 = e_1$, $N\left(v_1^3\right)$ y $N^2\left(v_1^3\right)$ serán las primeras columnas de N y N^2. De momento guardamos

$$Z\left(v_1^3\right) = \left(v_1^3, N\left(v_1^3\right), N^2\left(v_1^3\right)\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -4 \\ -2 \\ -3 \\ -8 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \\ -1 \\ -1 \end{pmatrix}.$$

Para el ciclo $Z\left(v_2^2\right)$ que falta, ha de elegirse $v_2^2 \in \mathbb{K}^2 - \mathbb{K}^1$ tal que $\left(Z\left(v_1^3\right), Z\left(v_2^2\right)\right)$ sea independiente. Probamos con $v_2^2 = e_5 = (0,0,0,0,1)^\top$, que da $N\left(v_2^2\right) = (0,1,1,1,2)^\top$. Yuxtaponiendo vectores,

$$c = \begin{vmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 & 1 \\ 0 & -2 & -1 & 0 & 1 \\ 0 & -3 & -1 & 0 & 1 \\ 0 & -8 & -1 & 1 & 2 \end{vmatrix} = - \begin{vmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & -4 & 0 & 1 & 0 \\ 0 & -2 & -1 & 1 & 0 \\ 0 & -3 & -1 & 1 & 0 \\ 0 & -8 & -1 & 2 & 1 \end{vmatrix} = \begin{vmatrix} -4 & 0 & 1 \\ -2 & -1 & 1 \\ -3 & -1 & 1 \end{vmatrix} = -1 \neq 0,$$

luego $\mathcal{J} = \left(Z\left(v_1^3\right), Z\left(v_2^2\right)\right)$ es base de Jordan. Como $c = \mathrm{mat}_{\mathcal{J}}^{\mathcal{E}}\left(\mathrm{id}\right)$, $N = \mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}\left(N\right)$ y $J = \mathrm{mat}_{J}^{\mathcal{J}}\left(N\right)$ debe cumplirse $c^{-1}Nc = J$, cosa que el ordenador comprueba para nuestra tranquilidad. Aunque esta comprobación no suele hacerse, recomendamos si se hace, que se evite calcular c^{-1} y se verifique que $cJ = ac$, que es equivalente. ♦

Problema 264 *Identificando $N(x) = ax$ o bx, endomorfismo de $\mathbb{R}^2$ o $\mathbb{R}^3$ con las natrices*

$$a = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 1 & -1 \\ -1 & -2 & 3 \\ -1 & -1 & 1 \end{pmatrix},$$

calcular matrices y bases de Jordan.

Problema 265 *Calcular una base de Jordan para*

$$a = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Problema 266 *Determinar para $N : \Bbbk^6 \to \Bbbk^6$ con $N(x) = ax$ y*

$$a = \begin{pmatrix} 0 & \beta & \beta & \beta & 0 & 0 \\ 0 & 0 & \beta & \beta & 0 & 0 \\ 0 & 0 & 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \beta \neq 0,$$

su matriz de Jordan y una base de Jordan.

Problema 267 *Sea la matriz* $a \in \Bbbk^{n\times n}$ *cuyas dos primeras filas son* $(1,1,\ldots,1,1)$, $(-1,-1,\ldots,-1,-1)$ *y las otras filas cero. Probar que* $N:\Bbbk^n \to \Bbbk^n$, $N(x)=ax$ *es nilpotente, y calcular su matriz y una base de Jordan.* ♦

Solución. Para calcular que hay un s tal que $N^s=0$, es mejor observar directamente que, sobre la base estándar, $N(e_i)=e_1-e_2$, $i=1,\ldots,n$, luego $N^2(e_i)=N(e_1-e_2)=(e_1-e_2)-(e_1-e_2)=0$. Claramente, $N^2=0$. La sucesión de rangos es $r^0=n$, $r^1=1$, $r^2=0$, luego las t^i, la "tabla de estrellas", y la matriz de Jordan, son

$$\begin{cases} t^2=r^1-r^2=1-0=1 \\ t^1=r^0-r^1=n-1 \end{cases}, \qquad \begin{array}{|c|c|c|c|} \hline * & & & \\ \hline * & * & \cdots & * \\ \hline \end{array}, \qquad J_0(2,1,\ldots,1)$$

con $n-1$ estrellas en la fila 1.

La base de Jordan será $\mathcal{J}=\left(Z\left(v_1^2\right),v_2^1,v_3^1,\ldots,v_n^1\right)$ con el ciclo $Z\left(v_1^2\right)$ de longitud 2 y $v_1^2\in\mathbb{K}^2-\mathbb{K}^1$. Tomamos $v_1^2=e_1=(1,0,\ldots,0)^\top$ con lo que $v_1^1=N\left(v_1^2\right)=(1,-1,0,\ldots,0)$. Los restantes $v_2^1,\ldots,v_{n-1}^1$ pueden elegirse arbitrariamente en $\mathbb{K}^1=\ker(N)$ pero formando sucesión independiente e independiente de $v_1^1=N\left(v_1^2\right)=(1,-1,0,\ldots,0)$. Podemos tomar

$$v_2^1=(1,0,-1,0,\ldots,0),\ v_3^1=(1,0,0,-1,0,\ldots,0),\ \ldots,v_i^1=(1,0,0,\ldots,-1,0,\ldots,0),\ \ldots .\blacklozenge$$

Problema 268 *Sea* $f:\mathbb{E}\to\Bbbk$ *una forma lineal, un vector* $u\neq 0$ *tal que* $f(u)=0$ *y otro* v *tal que* $f(v)=1$. *Probar que la función* $N:\mathbb{E}\to\mathbb{E}$, $N(x)=f(x)u$ *es nilpotente y construir una base de Jordan de la forma* $(v,u,u_1,\ldots,u_{n-2})$, *explicando cómo se eligen los* u_i.

Las sucesiones de rangos $\left(r^0,r^1,\ldots,r^s\right)$ o nulidades $\left(n^0,n^1,\ldots,n^s\right)$ no pueden ser arbitrarias.

Problema 269 *Dígase si puede existir una matriz* $a\in\Bbbk^{n\times n}$ *nilpotente tal que si* $r^i=\operatorname{rg}\left(a^i\right)$, *esta sucesión de rangos sea* $\left(r^0,r^1,r^2,r^3,r^4,r^5\right)=(n,10,8,5,3,0)$. ♦

Solución. No han precisado n pero no hace falta. Supongamos que a exista. Las ecuaciones del teorema (141) deben ser

$$\begin{cases} t^5=r^4-r^5=3-0=3 \\ t^4=r^3-r^4=5-3=2 \\ t^3=r^2-r^3=8-5=3 \\ t^2=r^1-r^2=10-8=2 \\ t^1=r^0-r^1=n-10 \end{cases}, \quad \text{con tabla} \quad \begin{array}{|c|c|c|c|} \hline * & * & * & \\ \hline * & * & & \\ \hline * & * & * & \\ \hline * & * & & \\ \hline * & * & ? & \cdots \\ \hline \end{array}$$

que rompen, en dos ocasiones al menos, la forma escalonada; o si se prefiere decir de otra forma, las desigualdades $t^s\leq t^{s-1}\leq\ldots\leq t^1$. No existe esta matriz, sea como sea n. ♦

Problema 270 *En el problema anterior cambiamos* $\left(r^0,r^1,r^2,r^3,r^4,r^5,r^6\right)$ *a* $(14,10,7,5,3,h,0)$. *¿Hay valores de* $h=r^5$ *para los que pueda existir* a? *Si fuera posible, ¿cuántas posibilidades habría?*

Casi todos los problemas toman $\mathbb{V}=\mathbb{R}^n$ y N se identifica con una matriz nilpotente. Sin embargo hay casos un poco diferentes. Tomemos $\mathbb{V}=\mathbb{R}_4[X]$, los polinomios de grado ≤ 4. Sea $N=D$ la derivada, que es sin duda nilpotente, porque $N^5=D^5$ anula cualquier polinomio. Entonces...

Problema 271 *Dar una base de Jordan y su matriz de Jordan correspondiente para* N. *Nota: es más fácil de lo que parece, y se simplifica todo un poco más si se toma como base en lugar de la estándar* $\mathcal{E}$ *la base* $\mathcal{B}$ *de los polinomios* $B_i(X)=X^i/i!$.

Problema 272 *Sea* $\mathbb{V}=\mathbb{R}_4[X]$, D *la derivada, y* $N=D^2+2D$. *Calcular la matriz y una base de Jordan para* N. *Nota: Aconsejamos usar la base* $\mathcal{B}$ *de los polinomios* $B_i(X)=X^i/i!$ *en vez de la estándar* $\mathcal{E}$.

7.5. Matrices y bases de Jordan en general

Hasta ahora hemos probado para un endomorfismo nilpotente N de $\mathbb{E}$ la existencia de bases y matrices de Jordan y cómo calcularlas. Tratamos el caso con L arbitraria. Si L solo tiene un valor propio λ, luego $C(X) = (X-\lambda)^n$ y $M(X) = (X-\lambda)^s$, basta poner $L = \lambda + (L-\lambda) = \lambda\,\mathrm{id} + L_\lambda$ con $L_\lambda = L - \lambda$. Claramente $L_\lambda = N$ es nilpotente de índice s porque $N^s = M(N) = 0$ y tenemos, por la sección anterior, una base de $\mathcal{J}$ en la que $N = L_\lambda$ tendrá matriz $J_0(q_1,\dots,q_k)$. Una transformación del tipo $\lambda\,\mathrm{id}$ tiene en *cualquier base* matriz diagonal λI, y entonces

$$\mathrm{mat}_{\mathcal{J}}^{\mathcal{J}}(L) = \mathrm{mat}_{\mathcal{J}}^{\mathcal{J}}(\lambda\,\mathrm{id}) + \mathrm{mat}_{\mathcal{J}}^{\mathcal{J}}(N) = \lambda I_n + J_0(q_1,\dots,q_k) = J_\lambda(q_1,\dots,q_k).$$

En resumen, la base de Jordan de L_λ sirve como base de Jordan de L y la matriz de Jordan de L es la de L_λ cambiando la diagonal de ceros por la diagonal de lambdas. ¿Qué sucede si hay más de un valor propio? Lleva su tiempo responder y lo haremos después del teorema 144. Primero un teorema general y abstracto pero sencillo.

Teorema 142 *Sea $\mathbb{E} = \mathbb{E}_1 \oplus \ldots \oplus \mathbb{E}_p$ y $L : \mathbb{E} \to \mathbb{E}$ tal que todos los subespacios $\mathbb{E}_h$ sean estables por L. Si cada $\mathcal{B}_h$ es una base de $\mathbb{E}_h$ y $a_{(h)}$ es la matriz de la restricción $L_h : \mathbb{E}_h \to \mathbb{E}_h$, se tiene que la matriz a de L en la base $\mathcal{B}$ obtenida por yuxtaposición de $\mathcal{B}_1,\dots,\mathcal{B}_h$ es*

$$a = \begin{pmatrix} a_{(1)} & & & & \\ & \ddots & & & \\ & & a_{(h)} & & \\ & & & \ddots & \\ & & & & a_{(p)} \end{pmatrix} = a_{(1)} \oplus \ldots \oplus a_{(h)} \oplus \ldots \oplus a_{(p)},$$

con ceros fuera de los $a_{(h)}$. (Se denotará $a = a_{(1)} \oplus \ldots \oplus a_{(h)} \oplus \ldots \oplus a_{(p)}$.)

Demostración. Es rutinaria y queda para el lector. ♣

Sean $(\lambda_1,\dots,\lambda_p)$ son los valores propios, ordenados de una vez para siempre. Tenemos la descomposición (teorema 135)

$$\mathbb{E} = \ker(L-\lambda_1)^{s_1} \oplus \ldots \oplus \ker(L-\lambda_p)^{s_p} = \mathbb{V}_1 \oplus \ldots \oplus \mathbb{V}_p, \qquad \mathbb{V}_h = \ker(L-\lambda_h)^{s_h}. \tag{7.3}$$

Antes de seguir adelante hay que tener claro dónde se mueven los índices. Hemos dicho que hay p valores propios y los índices h y k se moverán en $\{1,\dots,p\}$, mientras que i y j serán ≥ 0, denotando muchas veces potencias de L_{λ_h}; es decir, $L_{\lambda_h}^i = (L_{\lambda_h})^i = (L-\lambda_h)^i$. Los subespacios $\mathbb{V}_h$ son estables por L y por cualquier homotecia $x \to \mu x$ con $\mu \in \Bbbk$, luego lo serán asimismo por $L_\mu = L - \mu$. Si $\mu = \lambda_h$, uno de los valores propios de L, denotaremos por $\bar{L}_{\lambda_h} : \mathbb{V}_h \to \mathbb{V}_h$ a la restricción de L_{λ_h}. En principio, los subespacios

$$\mathbb{K}_h^i = \ker L_{\lambda_h}^i = \left\{x \in \mathbb{E} \mid (L-\lambda_h)^i(x) = 0\right\} \text{ y } \bar{\mathbb{K}}_h^i = \ker \bar{L}_{\lambda_h}^i = \left\{x \in \mathbb{V}_h \mid (L-\lambda_h)^i(x) = 0\right\}$$

podrían ser diferentes y sus dimensiones n_h^k y $\bar{n}_h^k$ también, *pero no es así*. Primero un teorema técnico.

Teorema 143 *Si $h \neq k$, la función L_{λ_h} es inyectiva si se la restringe a $\mathbb{V}_k = \ker L_{\lambda_h}^{s_k}$.*

Demostración. Si $x \in \mathbb{V}^k$ tenemos por definición que $(L-\lambda_k)^{s_k}(x) = 0$. Supongamos que sea $(L-\lambda_h)(x) = 0$. Entonces, como $L-\lambda_h$ y $L-\lambda_k$ conmutan, el binomio de Newton es aplicable y

$$((L-\lambda_k) - (L-\lambda_h))^{s_k} = \sum_{q=0}^{s_k} \binom{s_k}{q} (-1)^q (L-\lambda_k)^q \circ (L-\lambda_h)^{s_k-q}(x).$$

Todos los sumandos son nulos porque si $q < s_k$, $(L-\lambda_h)^{s_k-q}(x) = 0$ ya que $(L-\lambda_h)(x) = 0$, y si $q = s^k$ el sumando es $(-1)^{s_k}(L-\lambda_k)^{s_k}(x) = 0$. Esto nos lleva a

$$0 = ((L-\lambda_k) - (L-\lambda_h))^{s_k} = (\lambda_h - \lambda_k)^{s_k} x$$

y como $\lambda_h - \lambda_k \neq 0$, debe ser $x = 0$. ♣

Teorema 144 *Para todo i es $\mathbb{K}_h^i = \ker L_{\lambda_h}^i = \bar{\mathbb{K}}_h^i = \ker \bar{L}_{\lambda_h}^i$.*

Demostración. El contenido $\bar{\mathbb{K}}_h^i \subset \mathbb{K}_h^i$ es obvio. Para $x \in \mathbb{K}_h^i$ descomponemos $x = x_1 + \ldots + x_p$, siendo $x_k \in \mathbb{V}_k$. Entonces,

$$0 = (L-\lambda_h)^i(x) = (L-\lambda_h)^i(x_1) + \ldots + (L-\lambda_h)^i(x_p).$$

Pero al ser los sumandos de la suma directa estables por $(L-\lambda_h)^i$ (o sea, $(L-\lambda_h)^i(x_k) \in \mathbb{V}_k$), deducimos que todo $(L-\lambda_h)^i(x_k) = 0$ y, a continuación, $x_k = 0$ porque $L-\lambda_h$ es inyectiva (teorema 143). Por consiguiente, $x = x_h \in \mathbb{V}_h$, que unido a $(L-\lambda_h)^i(x) = 0$ nos da $x \in \bar{\mathbb{K}}_h^i$. ♣

Preparamos el **teorema de Jordan** un poco más abajo. La base de Jordan $\mathcal{J}$, que no es única, se va a construir yuxtaponiendo bases $\mathcal{J}_h$ de los $\mathbb{V}_h$ de (7.3). En concreto hemos visto que cada $\mathbb{V}_h$ es estable por L_{λ_h} y llamado $\bar{L}_{\lambda_h} : \mathbb{V}_h \to \mathbb{V}_h$ a la restricción, que es nilpotente porque $\mathbb{V}_h = \ker(L-\lambda_h)^{s_h}$. Existe entonces por el teorema 140 una base de Jordan $\mathcal{J}_h$ de $\bar{L}_{\lambda_h}$. La restricción de L a $\mathbb{V}_h$ es $L_{(h)} = \lambda_h \operatorname{id}_{\mathbb{V}_h} + \bar{L}_{\lambda_h}$ (no confundir $L_{(h)}$ y $\bar{L}_{\lambda_h}$) y como tiene un único valor propio λ_h, tal como hemos visto al inicio de la sección, la base $\mathcal{J}_h$ será también base de Jordan de $L_{(h)}$. Cada $\mathcal{J}_h$ está formada por ciclos disjuntos y las $\mathcal{J}_h$ están en sumandos diferentes de una suma directa, luego la yuxtaposición $\mathcal{J}$ está formado por ciclos disjuntos, y esta es precisamente la definición de base de Jordan. Para conocer la matriz de L en $\mathcal{J}$, utilizaremos el teorema 142 para $\mathbb{E}_h = \mathbb{V}_h$. Como cada restricción $L_{(h)}$ tiene en $\mathcal{J}_h$ un bloque de Jordan $J_{\lambda_h}\left(q_{(h)1}, \ldots, q_{(h)k_h}\right)$, la matriz J de L en $\mathcal{J}$ será (se entiende que hay ceros fuera de las J_{λ_h})

$$J = \begin{pmatrix} J_{\lambda_1}\left(q_{(1)1}, \ldots, q_{(1)k_1}\right) & & & & \\ & \ddots & & & \\ & & J_{\lambda_h}\left(q_{(h)1}, \ldots, q_{(h)k_h}\right) & & \\ & & & \ddots & \\ & & & & J_{\lambda_p}\left(q_{(p)1}, \ldots, q_{(p)k_p}\right) \end{pmatrix}, \tag{7.4}$$

Teorema 145 (de Jordan) *Para L con polinomios característico y minimal como en (7.1) existe una base de Jordan $\mathcal{J}$ en la que L tiene matriz J como en (7.4).*

El cálculo de matrices y bases de Jordan supone aplicar lo conocido para matrices nilpotentes a las diversas $L_{\lambda_h} = L - \lambda_h$ y el trabajo puede ser considerable. Conviene calcular la tabla de estrellas (7.2) de cada λ_h. Hemos distinguido minuciosamente entre $L_{\lambda_h} : \mathbb{E} \to \mathbb{E}$ y la restricción $\bar{L}_{\lambda_h} : \mathbb{V}_h \to \mathbb{V}_h$ porque para la tabla de estrellas necesitamos los números $\bar{t}_h^i = \bar{n}_h^i - \bar{n}_h^{i-1}$. Si hubiesen sido distintos los $n_h^i = \dim \mathbb{K}_h^i$ de los $\bar{n}_h^i = \dim \bar{\mathbb{K}}_h^i$ nos enfrentaríamos a un trabajo adicional, pero el teorema 144 nos dice que $n_h^i = \bar{n}_h^i$ y por consiguiente $t_h^i = \bar{t}_h^i$. Respecto a los rangos si sucede que $r_h^i = \operatorname{rg}\left(L_{\lambda_h}^i\right) \neq \bar{r}_h^i = \operatorname{rg}\left(\bar{L}_{\lambda_h}^i\right)$, pero no supone problema para el cálculo de los $t_h^i = \bar{t}_h^i$ ya que, por el teorema del rango nulidad, $r_h^i = n - n_h^i$ con lo que

$$r_h^{i-1} - r_h^i = \left(n - n_h^{i-1}\right) - \left(n - n_h^i\right) = n_h^i - n_h^{i-1} = t_h^i.$$

El resumen es que para las tablas de estrellas se pueden seguir utilizando

$$\begin{cases} t_h^s = n_h^s - n_h^{s-1} \\ \vdots \\ t_h^i = n_h^i - n_h^{i-1} \\ \vdots \\ t_h^1 = n_h^1 - n_h^0 \end{cases} \quad \text{o bien} \quad \begin{cases} t_h^s = r_h^{s-1} - r_h^s \\ \vdots \\ t_h^i = r_h^{i-1} - r_h^i \\ \vdots \\ t_h^1 = r_h^0 - r_h^1 \end{cases} \tag{7.5}$$

con los rangos y nulidades de L_{λ_h} en vez de los de $\bar{L}_{\lambda_h}$, que es mucho más cómodo.

Hay una cuestión pendiente para los cálculos: ¿cuántos $n_\bullet^\bullet$ o $r_\bullet^\bullet$ hay que calcular? En el caso nilpotente $L = N$, al ir calculando las diversas N^i llega un momento en que $N^s = 0$ y solo hay que calcular $r^0, r^1, \ldots, r^s$ o $n^0, n^1, \ldots, n^s$. Por supuesto, $\bar{L}_{\lambda_h}$ también verifica $\left(\bar{L}_{\lambda_h}\right)^{s_h} = 0$, pero hemos dicho que se pretende evitar trabajar con $\bar{L}_{\lambda_h} : \mathbb{V}_h \to \mathbb{V}_h$ haciéndolo con $L_{\lambda_h} : \mathbb{E} \to \mathbb{E}$ en su lugar. Sucede que si hay más de dos valores propios, L_{λ_h} no es nilpotente (por algo distinguíamos $\bar{L}_{\lambda_h}$ y L_{λ_h}) y no sabemos

cuándo parar el cálculo de los $(L_{\lambda_h})^i$ y r_h^i. Puede argumentarse que si se conoce el polinomio minimal $M(X)$, debemos parar en $i = s_h$, pero lo usual es conocer $C(X)$ y $s_h \leq m_h$. Si fuese $s_h < m_h$ los cálculos de $(L_{\lambda_h})^i$ y r_h^i para $i > s_h$ serían superfluos. Vamos a decir cómo determinar exactamente s_h y de paso disponer de un procedimiento de cálculo alternativo de $M(X)$ a partir de $C(X)$.

Teorema 146 *El exponente s_h en $M(X)$ es el primer i tal que* $\operatorname{rg}(L_{\lambda_h})^i = \operatorname{rg}(L_{\lambda_h})^{i+1}$.

Demostración. Consideremos $L_{\lambda_h} : \mathbb{V}_1 \oplus \ldots \oplus \mathbb{V}_p \longrightarrow \mathbb{V}_1 \oplus \ldots \oplus \mathbb{V}_p$. Cada $\mathbb{V}_h$ es estable por L_{λ_h}, siendo la restricción a $\mathbb{V}_k$ con $k \neq h$ un isomorfismo (teorema 143) y la restricción a $\mathbb{V}_h$ lo que hemos llamado $\bar{L}_{\lambda_h}$. Por definición, el rango es la dimensión de la imagen, luego

$$\begin{aligned} r_h^i &= \operatorname{rg}\left(L_{\lambda_h}^i\right) = \dim \operatorname{im}\left(L_{\lambda_h}^i\right) = \dim L_{\lambda_h}^i(\mathbb{V}_1) + \ldots + \dim L_{\lambda_h}^i(\mathbb{V}_h) + \ldots + \dim L_{\lambda_h}^i(\mathbb{V}_p) \\ &= \dim \mathbb{V}_1 + \ldots + \dim \mathbb{V}_{h-1} + \dim \bar{L}_{\lambda_h}^i(\mathbb{V}_h) + \dim \mathbb{V}_{h+1} + \ldots + \dim \mathbb{V}_p \\ &= \dim \mathbb{V}_1 + \ldots + \dim \mathbb{V}_{h-1} + \bar{r}_h^i + \dim \mathbb{V}_{h+1} + \ldots + \dim \mathbb{V}_p. \end{aligned}$$

El exponente s_h se caracteriza porque para $j \geq s_h$ son nulos todos los $\bar{n}_h^j$ y por tanto todos los $\bar{r}_h^j = \dim \mathbb{V}_h - \bar{n}_h^j$ toman el mismo valor $\dim \mathbb{V}_h$ para $j \geq s_h$. Por la ecuación de más arriba $r_h^i - \bar{r}_h^i = \sum_{j \neq h} \dim \mathbb{V}_j$ vemos que si los $\bar{r}_h^j$ toman el mismo valor para $j \geq s_h$, lo mismo les pasa a los r_h^j.

¿Puede haber repeticiones en $r_h^0, r_h^1, \ldots, r_h^i$? No, y de hecho hay *decrecimiento estricto.* En efecto, según **1** en el teorema 137, en los núcleos $\ker\left(L_{\lambda_h}^i\right)$ para $i \leq s_h$, hay *crecimiento estricto*, que obviamente se dará también en los $\bar{n}_h^i$. Al ser $\bar{r}_h^i = s_h - \bar{n}_h^i$, tendremos en las $\bar{r}_h^i$ decrecimiento estricto. Finalmente, como $r_h^i - \bar{r}_h^i = \sum_{j \neq h} \dim \mathbb{V}_j$, habrá también en las r_h^i con $i = 0, 1, \ldots, s_h$ decrecimiento estricto. ♣

Problema 273 *Calcular el polinomio minimal de la matriz*

$$a = \begin{pmatrix} 0 & -5 & -4 & -2 \\ 0 & 2 & 0 & 0 \\ 1 & 2 & 4 & 2 \\ 0 & \frac{1}{2} & 0 & 1 \end{pmatrix}. \blacklozenge$$

Solución. Calculamos que $C(X) = (X-2)^3(X-1)$ y $M(X) = (X-2)^s(X-1)$ con $1 \leq s \leq 3$. Se podría verificar si $(a-2)^2(a-1) = 0$ y sería $M(X) = (X-2)^2(X-1)$. Sin embargo vamos a seguir el procedimiento del teorema 146. Se calcula $(a-2)^i$ para $i = 1, 2, 3$, que son

$$\begin{pmatrix} -2 & -5 & -4 & -2 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 2 \\ 0 & \frac{1}{2} & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & -1 \end{pmatrix},$$

de rangos fácilmente calculables $r^1 = 2$, $r^2 = r^3 = 1$, y por el teorema 146 es $s = 2$. ♦

Damos un ejemplo, algo excesivo para el cálculo manual, pero que da idea del trabajo a realizar.

Problema 274 *Calcular la matriz y una base de Jordan para $L : \mathbb{R}^6 \to \mathbb{R}^6$, $L(x) = ax$, siendo*

$$a = \begin{pmatrix} 1 & -2 & -1 & -4 & -3 & -5 \\ 1 & 3 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix}. \blacklozenge$$

Solución. Calculando por cajas,

$$C(X) = \left| \begin{array}{ccc|ccc} 1-X & -2 & -1 & & & \\ 1 & 3-X & 1 & & & \\ 0 & 1 & 2-X & & & \\ \hline & & & 4-X & 0 & 0 \\ & & & 1 & 4-X & 0 \\ & & & 0 & 0 & 6-X \end{array} \right|.$$

El segundo determinante es $-(X-4)^2(X-6)$. El primero es

$$(2-X)\begin{vmatrix} 1-X & -2 \\ 1 & 3-X \end{vmatrix} - \begin{vmatrix} 1-X & -1 \\ 1 & 1 \end{vmatrix} = -(X-2)^3$$

y en definitiva, $C(X) = (X-2)^3(X-4)^2(X-6)$ con raíces $\lambda_1 = 2$, $\lambda_2 = 4$ y $\lambda_3 = 6$. Las matrices $a-2$, $a-4$ y $a-6$ son respectivamente. Las potencias $1, 2, 3, 4$ de $a-2 = L_{\lambda_1}$ son

$$a-2 = \begin{pmatrix} -1 & -2 & -1 & -4 & -3 & -5 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}, \quad (a-2)^2 = \begin{pmatrix} -1 & -1 & -1 & -9 & -5 & -17 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 \end{pmatrix},$$

$$(a-2)^3 = \begin{pmatrix} 0 & 0 & 0 & -20 & -8 & -64 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 12 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 64 \end{pmatrix}, \quad (a-2)^4 = \begin{pmatrix} 0 & 0 & 0 & -48 & -16 & -256 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 32 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 256 \end{pmatrix}.$$

cuyos rangos son $\left(r_1^1, r_1^2, r_1^3, r_1^4\right) = (5, 4, 3, 3)$. Por el teorema 146 tenemos que $s_1 = 3$. La justificación del valor de los rangos no es difícil. El rango de $a-2$ no puede ser 6 porque no es invertible y, por cajas, se calcula el determinante de la submatriz formada por las 5 últimas filas y columnas, que es $(-1)\cdot 2^2 \cdot 4 \neq 0$. Sumando a la fila 3 de $(a-2)^2$ la fila 1, obtenemos otra matriz cuyo rango es también $\operatorname{rg}(a-2)^2$ cuyo rango 4 es de fácil cálculo. El rango 3 de $(a-2)^3$ y $(a-2)^4$ es todavía más fácil de ver.

Con las fórmulas $t_1^i = r^{i-1} - r^i$ obtenemos que $t_1^3 = t_1^2 = t_1^1 = 1$ y la tabla de estrellas y la parte de la matriz de Jordan correspondiente a $\lambda_1 = 2$ son

$$\begin{array}{|c|}\hline * \\ \hline * \\ \hline * \\ \hline \end{array}, \quad J_2(3) = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

El cálculo de s_2 y s_3 es más fácil. Como $m_3 = 1$ y $1 \leq s_3 \leq m_3$, debe de ser $s_3 = 1$. Análogamente $1 \leq s_2 \leq m_2 = 2$. Dado que $\lambda_2 = 4$, podríamos calcular, será $s_2 = 2$ si los valores constantes de r_2^i son a partir de $i = 2$ y será $s_2 = 1$ si $r_2^1 > r_2^2$. Tenemos

$$(a-4)^1 = \begin{pmatrix} -3 & -2 & -1 & -4 & -3 & -5 \\ 1 & -1 & 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad (a-4)^2 = \begin{pmatrix} 7 & 7 & 3 & 7 & 7 & 3 \\ -4 & 0 & -4 & -4 & -4 & -4 \\ 1 & -3 & 5 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix},$$

y sus rangos respectivos son $r_2^1 = 5$ y $r_2^2 = 4$. Lo afirmado de los rangos se ve porque en $(a-4)^1$, tachando la fila 4 y la columna 5, queda una submatriz con determinante -16; y en $(a-4)^2$, tachando las filas 4 y 5 y las columnas 4 y 5, queda otra submatriz con determinante -16.

Conocidos los rangos, la tabla de estrellas y la parte de la matriz de Jordan, son

$$\begin{array}{|c|}\hline * \\ \hline * \\ \hline \end{array}, \quad J_4(2) = \begin{pmatrix} 4 & 0 \\ 1 & 4 \end{pmatrix}$$

para $\lambda_2 = 4$. Lo mismo para $\lambda_3 = 6$ es $J_6(1) = (6)$. La matriz de Jordan de L es

$$J = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 \end{pmatrix}.$$

Es bastante fácil conocer las bases de Jordan porque cada uno de los tres $\mathbb{V}_h$ tiene como base un solo ciclo. Serán tres ciclos $Z\left(v_1^3(1)\right)$, $Z\left(v_1^2(2)\right)$ y $Z\left(v_1^1(3)\right)$ de longitudes 3, 2 y 1. Se elige $v_1^3(1)$ en $\ker(a-2)^3 - \ker(a-2)^2$. Más arriba hemos calculado las $(a-2)^i$ y es obvio que $v_1^3(1) = (1,0,0,0,0,0)^\top$ es una buena elección. Con esto,

$$Z\left(v_1^3(1)\right) = \left((a-2)^i v_1^3(1)\right)_{i=0,1,2} = \left(\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}-1\\1\\0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\\0\\0\\0\end{pmatrix}\right).$$

Para calcular $v_1^2(2)$ en $\ker(a-4)^2 - \ker(a-4)^1$ obtenemos a ojo que $(-1,0,0,1,0,0)^\top$ y $(-1,0,0,0,1,0)^\top$ forman base de $\ker(a-4)^2$. (Justificación: están en $\ker(a-4)^2$, son independientes y $\dim\ker(a-4)^2 = 6-4 = 2$.) El primer vector no está en $\ker(a-4)^1$, y por eso tomamos $v_1^2(2) = (-1,0,0,1,0,0)^\top$. El segundo ciclo es

$$Z\left(v_1^2(2)\right) = \left((a-4)^i v_1^3(1)\right)_{i=0,1} = \left(\begin{pmatrix}-1\\0\\0\\1\\0\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\0\\0\\1\\0\end{pmatrix}\right)$$

Para el tercer ciclo nos basta un vector propio $v_1^1(3)$ de $a-6$ y con buen ojo se toma $(-1,0,0,0,0,1)^\top$. El ordenador comprueba que

$$\begin{pmatrix}1&-1&-1&-1&-1&-1\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}^{-1}\begin{pmatrix}1&-2&-1&-4&-3&-5\\1&3&1&1&1&1\\0&1&2&0&0&0\\0&0&0&4&0&0\\0&0&0&1&4&0\\0&0&0&0&0&6\end{pmatrix}\begin{pmatrix}1&-1&-1&-1&-1&-1\\0&1&0&0&0&0\\0&0&1&0&0&0\\0&0&0&1&0&0\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}$$

es la matriz de Jordan de más arriba. ♦

La receta para calcular a partir de a su matriz de Jordan es esta:

1. Factorizar $C(X) = (\lambda_1 - X)^{m_1} \cdots (\lambda_p - X)^{m_p}$
2. Para cada valor propio λ_h calcular la sucesión de rangos $r_h^0 > r_h^1 > r_h^2 > \ldots > r_h^i > \ldots$ hasta que aparezca $r_h^s = r_h^{s+1}$, en cuyo caso $s = s_h$ es el exponente de $(\lambda_h - X)$ en el polinomio minimal. Si se prefiere, se puede trabajar con las nulidades $n_h^0 < n_h^1 < n_h^2 < \ldots < r_h^i < \ldots$ hasta que aparezca $n_h^s = n_h^{s+1}$, en cuyo caso $s = s_h$ es el exponente de $(\lambda_h - X)$ en el polinomio minimal.
3. Escribir las ecuaciones del teorema 141 con los r_h^i o los n_h^i y, por medio de ellas, la tabla de estrellas. Contamos la sucesión de estrellas por columna, que deben formar una sucesión decreciente $q_{(h)1} \geq q_{(h)2} \geq \ldots \geq q_{(h)k_h}$. La parte de la matriz de Jordan que corresponde a λ_h es $J_{\lambda_h}\left(q_{(h)1}, q_{(h)2}, \ldots, q_{(h)k_h}\right)$. Yuxtaponiendo estas matrices en diagonal para los λ_h, obtenemos la matriz de Jordan.
4. La base de Jordan ha de calcularse por el procedimiento constructivo del teorema de Jordan (teorema 145). Siempre ayudará tener presente la tabla de estrellas para tener una idea de los ciclos que debemos construir.

Problema 275 *Calcular la matriz y una base de Jordan para*

$$a = \begin{pmatrix}1&0&1&1\\0&1&0&1\\0&0&1&0\\1&0&0&1\end{pmatrix}$$

sabiendo que los valores propios son $\lambda_1 = 0$, $\lambda_2 = 1$ *y* $\lambda_3 = 2$.

Problema 276 *Calcular las formas de Jordan de*

$$a=\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & 0 & -1 \end{pmatrix},\quad b=\begin{pmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{pmatrix},\quad c=\begin{pmatrix} -3 & -2 & 2 \\ 8 & 5 & -4 \\ 4 & 2 & -1 \end{pmatrix},\quad d=\begin{pmatrix} 4 & 1 & -1 \\ -1 & 1 & 3 \\ -1 & -1 & 4 \end{pmatrix}.$$

Para aligerar el trabajo decimos que sus polinomios característicos son, salvo signo,

$$C_a(X)=X^3,\quad C_b(X)=(X-3)(X-2)^2,\quad C_c(X)=(X+1)(X-1)^2,\quad C_d(X)=(X-3)^3.$$

Problema 277 *Calcular la forma de Jordan y una base de Jordan de*

$$a=\begin{pmatrix} 0 & 1 & 0 & -1 \\ -2 & 3 & 0 & -1 \\ -2 & 1 & 2 & -1 \\ 2 & -1 & 0 & 3 \end{pmatrix}.$$

Problema 278 *Para β arbitrario dar la matriz de Jordan de*

$$a=\begin{pmatrix} \beta & 1 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 1 & \beta \end{pmatrix}.$$

Problema 279 *Calcular la matriz y una base de Jordan para*

$$a=\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Problema 280 *En $\mathbb{E}=\mathbb{R}^{16}$ se tiene un endomorfismo L con tres valores propios $(\lambda_1,\lambda_2,\lambda_3)$ algunos de cuyos rangos $r_h^i=\operatorname{rg}\left(L_{\lambda_h}^i\right)$ vienen dados por*

$$\begin{cases} \left(r_1^1,r_1^2,r_1^3\right)=(14,R,11) \\ \left(r_2^1,r_2^2\right)=(13,10) \\ \left(r_3^1\right)=12 \end{cases}$$

y otros se desconocen. Determinar las posibles matrices de Jordan de L.

Problema 281 *En $\mathbb{E}=\mathbb{R}^{15}$ se tiene un endomorfismo L con tres valores propios $(\lambda_1,\lambda_2,\lambda_3)$ algunos de cuyos rangos $r_h^i=\operatorname{rg}\left(L_{\lambda_h}^i\right)$ vienen dados por*

$$\begin{cases} \left(r_1^1,r_1^2,r_1^3\right)=(13,11,10) \\ \left(r_2^1,r_2^2,r_2^3,r_2^4\right)=(13,11,10,9) \\ \left(r_3^1,r_3^2,r_3^3,r_3^4\right)=(14,13,12,11) \end{cases}$$

y otros se desconocen. Determinar las posibles matrices de Jordan de L.

7.6. Otras cuestiones sobre la forma de Jordan

¿Cómo se relaciona el polinomio mínimo $M(X)$ con la matriz de Jordan?

Teorema 147 *Supongamos que en la matriz de Jordan J de L los $q_{(h)}$ de cada λ_h están en orden decreciente $q_{(h)1}\geq\ldots\geq q_{(h)k_{(h)}}$. Entonces $M(X)=(X-\lambda_1)^{q_{(1)1}}(X-\lambda_2)^{q_{(2)1}}\cdots(X-\lambda_p)^{q_{(p)1}}$. Con palabras: al escribir $M(X)=(X-\lambda_1)^{s_1}\cdots(X-\lambda_h)^{s_p}$, el exponente s_h de $(X-\lambda_h)$ es la dimensión del bloque simple de Jordan de mayor dimensión de los que pueda tener $J_{\lambda_h}\left(q_{(h)1},\ldots,q_{(h)k_{(h)}}\right)$.*

Demostración. Tomemos una base de Jordan $\mathcal{J}$ con la que obtenemos J. El ciclo más largo de λ_h tiene longitud $q_{(h)1}$ puesto que hemos convenido que $q_{(h)1} \geq \ldots \geq q_{(h)k_{(h)i}}$. Por tanto, si $r \geq q_{(h)1}$, debe anularse $L_{\lambda_h}^r$ aplicado a cualquier término de estos ciclos. Al estar la base de Jordan $\mathcal{J}$ formada por los vectores de estos ciclos, $P(X) = (X-\lambda_1)^{q_{(1)1}}(X-\lambda_2)^{q_{(2)1}}\cdots(X-\lambda_p)^{q_{(p)1}}$ cumple que $P(L)$ se anula sobre los vectores de $\mathcal{J}$ y $P(L) = 0$. El polinomio minimal $M(X) = (X-\lambda_1)^{s_1}\cdots(X-\lambda_p)^{s_p}$ debe entonces dividir al $P(X)$ y se tendrá $s_h \leq q_{(h)1}$ para $h = 1, \ldots, p$.

Veamos que si para algún k fuese $s_k < q_{(k)1}$ habría contradicción. Sea $v_1^{q(k)1} = u$ el primer elemento del ciclo más largo de $\mathbb{V}_k$. Por la definición de ciclo, si $s_k < q_{(k)1}$ debe ser $L_{\lambda_k}^{s_k}(u) = w \neq 0$. Permutando los monomios a nuestra conveniencia,

$$M(L)(u) = \left(L_{\lambda_1}^{s_1} \circ \cdots \circ \widehat{L_{\lambda_k}^{s_k}} \circ \cdots \circ L_{\lambda_p}^{s_p}\right)\left(L_{\lambda_k}^{s_k}(u)\right) = \left(L_{\lambda_1}^{s_1} \circ \cdots \circ \widehat{L_{\lambda_k}^{s_k}} \circ \cdots \circ L_{\lambda_p}^{s_p}\right)(w).$$

Ahora bien, el teorema 143 dice que si $h \neq k$, la función L_{λ_h} es inyectiva restringida a $\mathbb{V}_k$ y $w \in \mathbb{V}_k$. Por tanto $u \neq 0$ pero $M(L)(u) \neq 0$. Contradicción, y $M(X)$ tiene la forma descrita. ♣

Damos un ejemplo: si

$$J = \begin{pmatrix} 6 & 0 & 0 & 0 & 0 \\ 1 & 6 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 1 & 6 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{pmatrix}, \text{ entonces } M(X) = (X-6)^2(X-7) \text{ pero } C(X) = (X-6)^4(X-7).$$

Problema 282 *Probar que la condición necesaria y suficiente para que sea $C(X) = M(X)$ es que para cada valor propio λ_h en una base de Jordan cualquiera solo haya un ciclo correspondiente a λ_h.*

Problema 283 *Digamos que L tiene $C(X) = (3-X)^9$, $r^1 = \operatorname{rg}(L-3) = 5$ y $r^2 = \operatorname{rg}(L-3)^2 = 2$. Determinar las posibles formas de Jordan de L y sus polinomios mínimos. Si hay más de una, ¿son estos polinomios diferentes?*

Problema 284 *Nos dan $L : \mathbb{E} \to \mathbb{E}$ con $\dim(\mathbb{E}) = 12$ y polinomios característico y minimal $C(X) = X^4(X-1)^4\left(X+\sqrt{2}\right)^4$ y $M(X) = X^2(X-1)\left(X+\sqrt{2}\right)^3$. Determinar las posibles formas de Jordan.*

Los endomorfismos semejantes diagonalizables tienen la misma forma diagonal. Otro tanto sucede con las formas de Jordan.

Teorema 148 *Sean $L, M, H : \mathbb{E} \to \mathbb{E}$ endomorfismos relacionados por $L = H \circ M \circ H^{-1}$ con H invertible. Si $\mathcal{V} = (v_1, \ldots, v_n)$ es una base de Jordan para M, entonces $\mathcal{U} = (H(v_1), \ldots, H(v_n)) = (u_1, \ldots, u_n)$ es base de Jordan de L. Además, las matrices de M y L en $\mathcal{V}$ y $\mathcal{U}$ respectivamente, son la misma.*

Demostración. La demostración es pesada pero, al seguir la idea natural, es simple en el fondo. Una base de Jordan está formada por ciclos, de modo que basta probar que si $Z_\lambda^M(v)$ es un ciclo para M con valor propio λ, la sucesión "transportada" por H,

$$H\left(Z_\lambda^M(v)\right) = \left(H\left(M_\lambda^0(v)\right), H\left(M_\lambda^1(v)\right), \ldots, H\left(M_\lambda^{q-1}(v)\right)\right)$$

es un ciclo para L. Lo es porque para todo i se tiene $L_\lambda^i(H(v)) = H\left(M_\lambda^i(v)\right)$, como se verá, y por tanto

$$\begin{aligned} H\left(Z_\lambda^M(v)\right) &= \left(H\left(M_\lambda^0(v)\right), H\left(M_\lambda^1(v)\right), \ldots, H\left(M_\lambda^{q-1}(v)\right)\right) \\ &= \left(L_\lambda^0(H(v)), L_\lambda^1(H(v)), \ldots, L_\lambda^{q-1}(H(v))\right) = Z_\lambda^L(H(v)). \end{aligned}$$

Teníamos pendiente que $L_\lambda^i(H(v)) = H\left(M_\lambda^i(v)\right)$. Esto es así puesto que

$$L_\lambda^i = (L-\lambda)^i = \left(H \circ M \circ H^{-1} - H \circ (\lambda\,\mathrm{id}) \circ H^{-1}\right)^i = H \circ (M-\lambda)^i \circ H^{-1} = H \circ (M_\lambda)^i \circ H^{-1}.$$

La última parte que parece que va a ser muy difícil es muy sencilla. *Con carácter general*, si se tiene $L = H \circ M \circ H^{-1}$, la matriz b de M en $\mathcal{V}$ es la misma que la matriz a de L en $\mathcal{U} = H(\mathcal{V})$. En efecto, $L(u_j)$ se puede calcular de dos maneras. Por una parte, haciendo intervenir la matriz b de M,

$$L(u_j) = L(H(v_j)) = \left[H \circ M \circ H^{-1}\right](H(v_j)) = H(M(v_j)) = H\left(\sum_{i=1}^n b_j^i v_i\right) = \sum_{i=1}^n b_j^i H(v_i) = \sum_{i=1}^n b_j^i u_i.$$

Y por otra, la definición de a da $L(u_j) = \sum_{i=1}^n a_j^i u_i$. Por tanto, $a = b$. ♣

Parece razonable pensar que la recíproca es cierta, y lo es.

Teorema 149 *Si existen bases $\mathcal{U}$ y $\mathcal{V}$ de Jordan para L y M en donde la matriz de Jordan respectiva de L y M es la misma, entonces L y M son semejantes.*

Demostración. Queda ver que si L y M tienen la misma matriz de Jordan b aunque sea para bases diferentes, entonces son semejantes. Esto es parte de un resultado mucho más general: si L y M tienen la misma matriz a en bases $\mathcal{U}$ y $\mathcal{V}$ (del tipo que sea), entonces L y M son semejantes. Vamos a probarlo.

La hipótesis es que hay una matriz a tal que $L(u_j) = \sum_{i=1}^n a_j^i u_i$ y $M(v_j) = \sum_{i=1}^n a_j^i v_i$ para $j = 1, \ldots, n$. Sea $P : \mathbb{E} \to \mathbb{E}$ dado por $P(u_j) = v_j$ para $j = 1, \ldots, n$. Entonces $P^{-1} \circ M \circ P = L$ equivale a $M \circ P = P \circ L$, cosa que se cumple porque

$$(M \circ P)(u_j) = M(v_j) = \sum_{i=1}^n a_j^i v_i, \quad (P \circ L)(u_j) = P\left(\sum_{i=1}^n a_j^i u_i\right) = \sum_{i=1}^n a_j^i P(u_i) = \sum_{i=1}^n a_j^i v_i,$$

Y al ser $M \circ P = P \circ L$ sobre una base, lo son sobre $\mathbb{E}$. ♣

Dado L, ¿es única su matriz de Jordan J? Hay dos maneras de entender la pregunta. La primera es que si se repasa la demostración del teorema 145 que construye J, que es matriz de L para una base $\mathcal{J}$ muy lejos de ser única, se podría pensar que distintas $\mathcal{J}$ podrían dar distintas J. Sabemos que no es así porque los coeficientes de J vienen dados por la tabla de estrellas y la forma de la tabla depende para cada λ_h solo de los rangos de $L^i_{\lambda_h}$. Hay no obstante otra pregunta con un matiz más general. Supongamos que L admite una base $\mathcal{B}$ para la que tiene una matriz b formada por cajas simples de Jordan en la diagonal. Permutando convenientemente estas cajas simples, ¿puede obtenerse la matriz de Jordan J de L? La respuesta es sí, luego si definimos matriz de Jordan como una matriz formada por bloques simples de Jordan (en orden arbitrario), se verifica que, salvo en lo referente al orden de los bloques, la matriz de Jordan de L es única.

Teorema 150 *Suponemos que $b = \mathrm{mat}_{\mathcal{B}}^{\mathcal{B}}(L)$ en la base $\mathcal{B}$ esta formada por bloques simples de Jordan en posición diagonal. Para el valor propio λ, sea N_λ^q el número de bloques simples de Jordan en b de la forma $J_\lambda(q)$ (puede ser $N_\lambda^q = 0$.) Entonces, las sucesiones $\left(N_\lambda^1, N_\lambda^2, \ldots, N_\lambda^n\right)$ están determinadas por λ y la matriz de Jordan de L es única en cuanto al número de veces que aparece $J_\lambda(q)$.*

Demostración. Utilizaremos que si a tiene la forma semidiagonal

$$a = \begin{pmatrix} a_{(1)} & & \\ & \ddots & \\ & & a_{(p)} \end{pmatrix} \text{ se cumple que } a^t = \begin{pmatrix} a_{(1)}^t & & \\ & \ddots & \\ & & a_{(p)}^t \end{pmatrix} \text{ y } \operatorname{rg} a^t = \operatorname{rg} a_{(1)}^t + \ldots + \operatorname{rg} a_{(p)}^t.$$

Y también que el bloque simple de Jordan $J_\lambda(q)$ es invertible si $\lambda \neq 0$ y que $\operatorname{rg}\left(J_0(q)^t\right) = \operatorname{máx}(0, q-t)$. (Al aumentar el exponente t los rangos bajan de una en una unidad hasta que es cero si $t = q$.)

Fijemos en adelante un λ. Claramente $L - \lambda$ tiene matriz $b - \lambda$, formada por bloques simples de Jordan de modo que el bloque $J_{\lambda_i}(q)$ en b se sustituye por $J_{\lambda_i - \lambda}(q)$ en $b - \lambda$. Tenemos entonces que $b - \lambda$ está formada por $N_\lambda^1, N_\lambda^2, \ldots, N_\lambda^n$ bloques de forma respectiva $J_0(1), J_0(2), \ldots, J_0(n)$, y el resto bloques $J_{\lambda_i - \lambda}(q)$, Si las dimensiones de estos últimos bloques $J_{\lambda_i - \lambda}(q)$ suman s, que de hecho es n menos la multiplicidad m_λ de λ en $C(X)$, tendremos que

$$\operatorname{rg}(b-\lambda)^t = \sum_{i=1}^k \operatorname{rg} J_0(q_i)^t + \sum_{\lambda_i \neq \lambda} J_{\lambda_i - \lambda}(q)^t = \sum_{i=1}^k \operatorname{rg} J_0(q_i)^t + s = \sum_{i=1}^k \operatorname{rg} J_0(q_i)^t + (n - m_\lambda).$$

Es posible que haya bloques simples del mismo tamaño repetidos; o sea, que tengamos H bloques $m \times m$, $J_0(q_{i_1})^t = J_0(q_{i_2})^t = \ldots = J_0^t(q_{i_H}) = J_0(m)^t$ con $q_{i_1} = q_{i_2} = \ldots = q_{i_H} = m$, en cuyo caso la suma de los rangos de todos ellos es

$$\operatorname{rg} J_0(q_{i_1})^t + \ldots + = \operatorname{rg} J_0(q_{i_H}) = H \operatorname{rg} J_0(m)^t = H \operatorname{máx}(0, m-t).$$

Nos queda entonces

$$R_\lambda^t = \operatorname{rg}(b-\lambda)^t - (n-m_\lambda) = \sum_{q=1}^{n} N_\lambda^q \operatorname{máx}(0, q-t).$$

Descartando los sumandos que con certeza son nulos,

$$R_\lambda^t = 1\cdot N_\lambda^{t+1} + 2\cdot N_\lambda^{t+2} + \ldots + (n-t+1)\cdot N_\lambda^{n-1} + (n-t)\cdot N_\lambda^n.$$

Estas ecuaciones lineales ligan $R_\lambda^1, \ldots, R_\lambda^{n-1}$ con $N_\lambda^2, \ldots, N_\lambda^n$ y se expresan en forma matricial

$$\begin{pmatrix} R_\lambda^1 \\ R_\lambda^2 \\ \vdots \\ R_\lambda^{n-2} \\ R_\lambda^{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & \cdots & n-2 & n-1 \\ 0 & 1 & 2 & \cdots & n-3 & n-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 2 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} N_\lambda^2 \\ N_\lambda^3 \\ \vdots \\ N_\lambda^{n-1} \\ N_\lambda^n \end{pmatrix}. \tag{7.6}$$

Obsérvese que la matriz $(n-1)\times(n-1)$ es invertible, luego $\left(R_\lambda^1, \ldots, R_\lambda^{n-1}\right)$ determina unívocamente $\left(N_\lambda^2, \ldots, N_\lambda^n\right)$, y también N_λ^1 porque $1\cdot N_\lambda^1 + 2\cdot N_\lambda^2 + \ldots + n\cdot N_\lambda^n = n - s = m_\lambda$. Tenemos con todo esto que L determina unívocamente los rangos R_λ^t, que a su vez determinan $\left(N_\lambda^1, N_\lambda^2, \ldots, N_\lambda^n\right)$.

Para L se pueden tener dos bases de Jordan $\mathcal{J}_1$ y $\mathcal{J}_2$ y sus correspondientes matrices b_1 y b_2, pero, por definición de base de Jordan, están constituidas por ciclos, luego b_1 y b_2 están formadas por bloques simples de Jordan en la diagonal. Como, por el teorema precedente, las veces en que aparece cada bloque $J_\lambda(q)$ solo depende de L, luego ha de ser el mismo en b_1 y b_2, tenemos que b_1 y b_2 son iguales excepto por la posición de estos bloques. ♣

La matriz de (7.6), llamémosla a considerándola $n\times n$, tiene una inversa curiosa. Tenemos que

$$a^{-1} = \begin{pmatrix} 1 & 2 & 3 & \cdots & n-2 & n-1 & n \\ 0 & 1 & 2 & \cdots & n-3 & n-2 & n-1 \\ 0 & 0 & 1 & \cdots & n-4 & n-3 & n-2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 2 & 3 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 2 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -2 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Descrita verbalmente, tiene 1 en la diagonal, -2 en la que está encima, 1 en la siguiente, y el resto 0. Por ejemplo,

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

Descrita a por filas, las operaciones son

$$\begin{pmatrix} a^1 \\ a^2 \\ \vdots \\ a^{n-1} \\ a^n \end{pmatrix} \xrightarrow{1} \begin{pmatrix} a^1 - a^2 \\ a^2 - a^3 \\ \vdots \\ a^{n-1} - a^n \\ a^n \end{pmatrix} = b = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 0 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \xrightarrow{2} \begin{pmatrix} b^1 - b^2 \\ b^2 - b^3 \\ \vdots \\ b^{n-1} - b^n \\ b^n \end{pmatrix} = I$$

En $\xrightarrow{1}$ hay $n-1$ operaciones: se sustituye a^1 por $b^1 = a^1 - a^2$, a^2 por $b^2 = a^2 - a^3$, … , a^{n-1} por $b^n = a^{n-1} - a^n$, y se deja $b^n = a^n$. La matriz b queda con todo 1 en la diagonal y por encima y 0 bajo la diagonal. En $\xrightarrow{2}$ se sustituye b^1 por $c^1 = b^1 - b^2$, b^2 por $c^2 = b^2 - b^3$, … b^{n-1} por $c^{n-1} = b^{n-1} - b^n$, y se deja $c^n = b^n$. Resulta $c = I$ la matriz unidad. Estas mismas operaciones en I nos dan

$$\begin{pmatrix} e^1 \\ e^2 \\ \vdots \\ e^{n-1} \\ e^n \end{pmatrix} \xrightarrow{1} \begin{pmatrix} e^1 - e^2 \\ e^2 - e^3 \\ \vdots \\ e^{n-1} - e^n \\ e^n \end{pmatrix} = u = \begin{pmatrix} 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \xrightarrow{2} \begin{pmatrix} u^1 - u^2 \\ u^2 - u^3 \\ \vdots \\ u^{n-1} - u^n \\ u^n \end{pmatrix} = v$$

y esta v es la que anunciamos como a^{-1}. La fórmula (7.6) nos permite escribir

$$\begin{pmatrix} N_\lambda^2 \\ N_\lambda^3 \\ \vdots \\ N_\lambda^{n-1} \\ N_\lambda^n \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -2 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_\lambda^1 \\ R_\lambda^2 \\ \vdots \\ R_\lambda^{n-2} \\ R_\lambda^{n-1} \end{pmatrix}, \tag{7.7}$$

que es una forma compacta de conocer J a partir de todos los rangos de las potencias de $L - \lambda_i$. Para ilustrarlo en un caso sencillo nos vamos al problema 272 donde $\mathbb{E} = \mathbb{R}_4[X]$, y L (que allí se llama N) es $L = D^2 + 2D$, siendo D la derivada. En este caso, L es nilpotente con solo $\lambda = 0$, lo que simplifica algo las fórmulas porque

$$R_\lambda^t = \operatorname{rg}(L - \lambda)^t - (n - m_\lambda) = \operatorname{rg}(L)^t + (5 - 5) = \operatorname{rg}(L)^t$$

y hay (o ha habido) que calcular los rangos $\operatorname{rg}(L) = 3$, $\operatorname{rg}\left(L^2\right) = 2$, $\operatorname{rg}\left(L^3\right) = 1$ y $\operatorname{rg}\left(L^4\right) = 0$, La fórmula (7.7) se traduce en

$$\begin{pmatrix} N^2 \\ N^3 \\ N^4 \\ N^5 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

a lo que añadimos $N^1 = n - \sum_{i=2}^5 iN^i = 5 - 4 \cdot 1 = 1$. La matriz J tiene un bloque simple 4×4 y otro 1×1; o sea

$$J = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

que es lo que se ha tenido que obtener en el problema 272.

Concluimos con un problema de tipo teórico pero fácil si hay claridad de conceptos.

Problema 285 *Probar que una matriz a y su traspuesta $a^\top$ son semejantes.*

Capítulo 8

Formas bilineales simétricas

8.1. Funciones bilineales

Mientras no digamos lo contrario, $\mathbb{E}$ será un espacio vectorial sobre el cuerpo $\Bbbk$, aunque el lector puede pensar en $\Bbbk = \mathbb{R}$ para un primer contacto y ejemplos. Diremos que una función $\beta : \mathbb{E} \times \mathbb{E} \to \Bbbk$ es **bilineal** (o que es una **forma bilineal**) si, fijada una variable, es lineal en la otra. Con más detalle: para cada $x \in \mathbb{E}$ la función $y \to \beta(x, y)$ de $\mathbb{E}$ en $\Bbbk$ es lineal y para cada $y \in \mathbb{E}$ la función $x \to \beta(x, y)$ de $\mathbb{E}$ en $\mathbb{R}$ es lineal. El ejemplo más sencillo, y a la vez revelador, es el que toma $\mathbb{E} = \Bbbk^n$ y

$$\beta : \Bbbk^n \times \Bbbk^n \to \Bbbk, \qquad \beta\left(\left(x^1, \dots, x^n\right)^\top, \left(y^1, \dots, y^n\right)^\top\right) = x^1 y^1 + \cdots + x^n y^n = \sum_{i=1}^{n} x^i y^i.$$

Para $\Bbbk = \mathbb{R}$ y $n = 3$ se tiene por ejemplo

$$\beta\left((1, -1, 4)^\top, (-1, -1, 2)^\top\right) = 1 \cdot (-1) + (-1)^2 + 4 \cdot 2 = 8, \quad \beta\left((1, -1, 0)^\top, (a, a, a)^\top\right) = a - a = 0.$$

Si fijamos la primera variable x de *cualquier modo*, la función

$$f\left(\left(y^1, \dots, y^n\right)^\top\right) = x^1 y^1 + \cdots + x^n y^n$$

es lineal. En el ejemplo ve que para $x = (1, -1, 4)$ sale $f\left(y^1, y^2, y^3\right) = y^1 - y^2 + 4y^3$, que es lineal. Se puede comprobar lo dicho, pero hay una manera más sencilla de verificarlo y mucho más general.

Teorema 151 *Sea b una matriz cuadrada de $\Bbbk^{n \times n}$ cuyos coeficientes representamos con dos subíndices*[1] *en la forma b_{ij}. Tenemos una función bilineal*

$$\beta : \Bbbk^n \times \Bbbk^n \to \Bbbk, \quad \beta(x, y) = x^\top b y = \left(x^1, \dots, x^n\right) \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \ddots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} = \sum_{i,j=1}^{n} b_{ij} x^i y^j.$$

Demostración. Es consecuencia inmediata de las propiedades del producto de matrices. Por ejemplo, $xb(y + z) = xby + xbz$. La fórmula con el sumatorio es también inmediata

$$\left(x^\top b y\right)_1^1 = \sum_{i=1}^{n} \left(x^\top\right)_i^1 (by)_1^i = \sum_{i=1}^{n} x^i (by)^i = \sum_{i=1}^{n} x^i \left(\sum_{j=1}^{n} b_{ij} y^j\right) = \sum_{i,j=1}^{n} b_{ij} x^i y^j.$$

Los "extraños" índices 1 se deben a que $x^\top b y \in \Bbbk^{1 \times 1}$. ♣

[1] El lector puede preguntarse por qué decidimos ahora utilizar dos subíndices para los coeficientes de una matriz y no subíndice y superíndice como hasta ahora. La razón, como se verá en el teorema siguiente, es que, al introducir bases y coordenadas, se mantiene la convención de Einstein (sumar en el índice que aparece una vez arriba y otra abajo en distintos factores). Y ya dijimos que esta convención ayuda para trabajar de modo rápido y seguro.

Por ejemplo, en $\mathbb{R}^3$ tenemos una función bilineal

$$\beta(x,y) = \left(x^1, x^2, x^3\right) \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \left(x^1, x^1+x^2, x^2+x^3\right) \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix}$$
$$= x^1y^1 + \left(x^1+x^2\right)y^2 + \left(x^2+x^3\right)y^3 = x^1y^1 + x^1y^2 + x^2y^2 + x^2y^3 + x^3y^3.$$

Se puede calcular como caso particular

$$\beta\left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right) = (1,2,1) \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} = (1,2,1) \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix} = -6.$$

Es más cómodo calcular con la forma matricial que con los sumatorios, pero hay que saber manejar tanto una como otra.

Hay muchísimos ejemplos de funciones bilineales aun sin ser $\mathbb{E}$ de dimensión finita. Por ejemplo, $\mathbb{E}$ puede ser el espacio de los polinomios o de las funciones continuas con valores en $\mathbb{R}$ definidas en un intervalo, digamos $[0,1]$, y

$$\beta(f(t), g(t)) = \int_0^1 f(t)\, g(t)\, dt \text{ o bien } \beta(f(t), g(t)) = \left(\int_0^1 f(t)\, dt\right)\left(\int_0^1 g(t)\, dt\right).$$

Otro ejemplo, para $\mathbb{E}$ un espacio vectorial de funciones $f : X \to \Bbbk$, podría ser $\beta(f,g) = f(p)\, g(p)$ siendo p, q puntos prefijados de X. El segundo y tercer ejemplos siguen el mismo patrón que consiste en tomar formas lineales $\phi, \psi : E \to \Bbbk$ y definir $\beta(x,y) = \phi(x)\, \psi(y)$. En esos ejemplos son respectivamente.

$$\phi(f) = \psi(f) = \int_0^1 f(t)\, dt, \qquad \phi(f) = f(p),\ \psi(f) = f(q).$$

La forma β obtenida a partir de ϕ, ψ por este procedimiento se denota por $\phi \otimes \psi$, se lee "ϕ tensorial ψ", y se llama el **producto tensorial (de las formas lineales)** ϕ y ψ.

Problema 286 *Probar que*

1. *La suma de formas bilineales y su producto por elementos de $\Bbbk$ da nuevas formas bilineales, luego el conjunto $\mathcal{B}(\mathbb{E})$ de las formas bilineales es con estas dos operaciones un espacio vectorial.*
2. *Probar que $\beta = \phi \otimes \psi$ es bilineal y, más generalmente, que si tenemos formas lineales $\phi_1, \ldots, \phi_k$ y $\psi_1, \ldots, \psi_k$ entonces $\beta = \phi_1 \otimes \psi_1 + \ldots + \phi_k \otimes \psi_k$ es bilineal.*

Este problema nos dice que es fácil conseguir ejemplos de formas bilineales. Por ejemplo, si $\mathbb{E} = \mathbb{R}[X]$, el espacio de los polinomios reales,

$$\beta(P(X), Q(X)) = P(1)\, Q'(2) + P(4) \int_{-1}^1 Q(X)\, dX$$

es una función bilineal. (¿Cuáles son las ϕ y ψ?)

8.1.1. Matrices de una función bilineal

Del mismo modo que al tratar espacios y funciones lineales advertimos que nos íbamos a limitar sobre todo al caso de espacios de dimensión finita, aquí vamos a hacer lo mismo con funciones bilineales. Si suponemos que $\mathcal{U} = (u_1, \ldots, u_n)$ es una base de $\mathbb{E}$, vamos a ver que se puede manejar β "en coordenadas"; más concretamente, que se le puede asignar a β una matriz $b \in \Bbbk^{n\times n}$ (dependiente de la base). Antes de ir a las definiciones advertimos que hay una posible fuente de confusiones con los endomorfismos $L : \mathbb{E} \to \mathbb{E}$ a los que también se les puede asignar una matriz $a = \mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) \in \Bbbk^{n\times n}$. *Las matrices cuadradas representan ambos objetos, pero β y L tienen naturalezas diferentes.* Definimos la **matriz de la forma bilineal (para la base $\mathcal{U}$)** como

$$b = (b_{ij}),\ \text{siendo } b_{ij} = \beta(u_i, u_j) \text{ y denotándose también } b = \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta) \in \Bbbk^{n\times n}.$$

Los b_{ij} se llaman los **coeficientes de la forma bilineal** (en la base $\mathcal{U}$) y, si no hay más que una base, se denotan a veces por β_{ij} (del mismo modo que los coeficientes del endomorfismo L pueden ser L_j^i). El lector habrá observado la diferente posición de índices en $\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)$ respecto a $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$.

Un ejemplo. Sea $\mathbb{E}=\mathbb{R}_2[X]$ y $\beta(P(X),Q(X))=P(2)Q(3)$. Para la base estándar $\mathcal{E}=\left(1,X,X^2\right)$,

$$\beta\left(X^p,X^q\right)=2^p3^q,\ \text{luego}\ \mathrm{mat}_{\mathcal{E}\mathcal{E}}(\beta)=\begin{pmatrix}2^03^0 & 2^03^1 & 2^03^2\\ 2^13^0 & 2^13^1 & 2^13^2\\ 2^23^0 & 2^23^1 & 2^23^2\end{pmatrix}=\begin{pmatrix}1 & 3 & 9\\ 2 & 6 & 18\\ 4 & 12 & 36\end{pmatrix}. \tag{8.1}$$

Teorema 152 *Sea $\beta:\mathbb{E}\times\mathbb{E}\to\Bbbk$ una función bilineal y $\mathcal{U}=(u_1,\ldots,u_m)$ una base numerada de $\mathbb{E}$. La matriz $b=\mathrm{mat}_{\mathcal{E}\mathcal{E}}(\beta)$ determina unívocamente β, con sumandos o con matrices, por*

$$\beta(x,y)=\sum_{1\le i,j\le m}x^iy^jb_{ij},\quad \textit{o bien}\quad \beta(x,y)=x^\top by=\left(x^1,\ldots,x^n\right)\begin{pmatrix}\beta_{11} & \cdots & \beta_{1n}\\ \vdots & \ddots & \ddots\\ \beta_{n1} & \cdots & \beta_{nn}\end{pmatrix}\begin{pmatrix}y^1\\ \vdots\\ y^n\end{pmatrix},$$

siendo $x=\sum_{i=1}^m x^iu_i$, $y=\sum_{j=1}^m y^ju_j$.

Demostración. La demostración es muy sencilla porque solo hay que usar la linealidad de β en cada variable. En efecto,

$$\begin{aligned}\beta(x,y)&=\beta\left(\sum_{i=1}^m x^iu_i,\sum_{j=1}^m y^ju_j\right)=\beta\left(\sum_{i=1}^m x^iu_i,y\right)\overset{1}{=}\sum_{i=1}^m x^i\beta(u_i,y)=\sum_{i=1}^m x^i\beta\left(u_i,\sum_{j=1}^m y^ju_j\right)\\ &\overset{2}{=}\sum_{i=1}^m x^i\left(\sum_{j=1}^m y^j\beta(u_i,u_j)\right)\overset{3}{=}\sum_{1\le i,j\le m}x^iy^j\beta(u_i,u_j)\overset{4}{=}\sum_{1\le i,j\le m}x^iy^jb_{ij}.\end{aligned}$$

En $\overset{1}{=}$ se usa la linealidad en la primera variable y en $\overset{2}{=}$ en la segunda variable. Recordemos que si L es lineal, con lenguaje de sumatorios, $L\left(\sum_{k=1}^m z^ku_k\right)=\sum_{k=1}^m z^kL(u_k)$. En $\overset{3}{=}$ se usa la distributividad generalizada y en $\overset{4}{=}$ hay una simple sustitución $\beta(u_i,u_j)=b_{ij}$. El trabajo para ver que la fórmula con sumatorios equivale a la fórmula con matrices es similar al realizado en el teorema 151 ♣

Con este teorema se ve enseguida que

$$\beta(x,y)=\mathrm{mat}^{\mathcal{U}}(x)^\top\,\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\,\mathrm{mat}^{\mathcal{U}}(y) \tag{8.2}$$

que se recuerda por analogía a la ecuación con fracciones $1=U\frac{1}{UU}U$, pero sin olvidar $\top$ en $\mathrm{mat}^{\mathcal{U}}(x)^\top$. En el ejemplo de más arriba

$$\beta\left(1-X+2X^2,X-X^2\right)=(1,-1,2)\begin{pmatrix}1 & 3 & 9\\ 2 & 6 & 18\\ 4 & 12 & 36\end{pmatrix}\begin{pmatrix}0\\ 1\\ -1\end{pmatrix}=-42,$$

también calculable como $\beta\left(1-X+2X^2,X-X^2\right)=\left[1-X+2X^2\right]_{X=2}\cdot\left[X-X^2\right]_{X=3}=-42$.

Problema 287 *En $\mathbb{E}$ de dimensión 3 con una base $\mathcal{U}=(u_1,u_2,u_3)$ tenemos una función bilineal β dada por $\beta(u_i,u_j)=\max\{i,j\}$. Escribir la matriz de β y calcular $\beta\left((1,2,1)^\top,(-1,-1,0)^\top\right)$. ¿Existen vectores x de la forma $(\lambda,\lambda,\lambda)^\top$ con $\lambda\neq0$ tales que $\beta(x,x)=0$?*

El ejemplo más importante de forma bilineal es el de la **forma estándar** para de $\mathbb{E}=\Bbbk^n$, que se puede describir de varias formas

$$\varepsilon(x,y)=x^1y^1+\ldots+x^ny^n=\sum_{i=1}^n x^iy^i,\qquad \varepsilon(x,y)=x^\top I_ny=x^\top y,\qquad \mathrm{mat}_{\mathcal{E}\mathcal{E}}(\varepsilon)=I_n.$$

En el caso $\Bbbk=\mathbb{R}$ se llama[2] a ε el **producto euclidiano estándar**. Ahora ε es una forma más entre otros muchas, pero merecerá trato distinguido en el capítulo siguiente, pues la idea de producto euclidiano es la clave para tratar las ideas de distancia y ángulo.

[2] La RAE solo admite *euclidiano* y no *euclídeo,* pero *euclídeo* está muy extendido, quizás por la influencia del *Euclidean* inglés.

Problema 288 *Comprobar que en la base estándar $\mathcal{E}$ de $\Bbbk^n$ las componentes son $\varepsilon_{ij} = \delta_{ij}$, las deltas de Kronecker δ_{ij} (que son 1 o 0 según sea $i = j$ o $i \neq j$). Comprobar también que en el caso $\Bbbk = \mathbb{R}$ siempre se tiene $\varepsilon(x,x) \neq 0$ para $x \neq 0$ (de hecho, $\varepsilon(x,x) > 0$) pero que para $\Bbbk = \mathbb{C}$ existe $x \neq 0$ cumpliendo que $\varepsilon(x,x) = 0$.*

Problema 289 *Sea $\mathbb{E} = \mathbb{R}_1[X]$, el espacio de polinomios de grado ≤ 1 y*

$$\beta(P(X), Q(X)) = \int_0^1 P(X)\, Q(X)\, dX.$$

Calcular las matrices de β para la base estándar $\mathcal{E} = (1, X)$ y para $\mathcal{V} = (1 + X, 1 - X)$.

En $\mathbb{R}_n[X]$ el espacio de los polinomios de grado $\leq n$ se pide la matriz de $\beta(P(X), Q(X)) = P(1)\, Q(1)$ para la base estándar $\mathcal{E}$ y para $\mathcal{V} = \left(1, 1 + X, 1 + X^2, \ldots, 1 + X + X^2 + \ldots + X^n\right)$

Problema 290 *Definimos $\beta : \Bbbk^2 \times \Bbbk^2 \to \Bbbk$ por $\beta(a_1, a_2) = \det(a)$ (juntamos los dos vectores columna y calculamos el determinante de la matriz cuadrada de $\Bbbk^{2\times 2}$) Calcular la matriz de β en la base estándar. Si hacemos lo análogo con la traza; o sea, $\alpha(a_1, a_2) = \operatorname{tr}(a)$, ¿se obtiene una función bilineal?*

Cambiamos de espacio vectorial a $\mathbb{E} = \Bbbk^{2\times 2}$. Definimos $\gamma(a,b) = \operatorname{tr}(a)\det(b) + \operatorname{tr}(b)\det(a)$ ¿se obtiene una función bilineal? ¿Y si se cambia + por −?

Si se toma como $\mathbb{E}$ un espacio de matrices $\Bbbk^{n\times n}$ aparecen formas bilineales curiosas. Trabajar con sus matrices es más difícil porque ahora $\dim \mathbb{E} = n^2$. Recordamos que la traza $\operatorname{tr} : \Bbbk^{n\times n} \to \Bbbk$, $\operatorname{tr}(a) = \sum_{i=1}^n a_i^i$, es lineal.

Problema 291 *Sea la función $\beta : (\Bbbk^{n\times n}) \times (\Bbbk^{n\times n}) \to \Bbbk$ dada por $\beta(a,b) = \operatorname{tr}\left(a^\top b\right)$. Probar que es bilineal y que $\beta(a,b) = \beta(b,a) = \sum_{i,j=1}^m a_i^j b_i^j$. Como se ve, $\beta(a,b)$ se escribe multiplicando término a término y sumando todos los productos. Es como el producto estándar de $\mathbb{R}^m$.*

8.1.2. Efectos de cambios de base

Igual que cuando $\mathbb{E}$ tiene dos bases $\mathcal{U}$ y $\mathcal{V}$, un endomorfismo L tiene dos matrices, también β tiene dos matrices. ¿Cómo se relacionan? Todo es cuestión de calcular $\beta(x,y)$ de dos formas:

$$\beta(x,y) = \operatorname{mat}^{\mathcal{V}}(x)^\top \operatorname{mat}_{\mathcal{V}\mathcal{V}}(\beta) \operatorname{mat}^{\mathcal{V}}(y),$$

$$\begin{aligned}
\beta(x,y) &= \operatorname{mat}^{\mathcal{U}}(x)^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}^{\mathcal{U}}(y) \\
&= \left(\operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \operatorname{mat}^{\mathcal{V}}(x)\right)^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \left(\operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \operatorname{mat}^{\mathcal{V}}(y)\right) \\
&= \operatorname{mat}^{\mathcal{V}}(x)^\top \left(\operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right) \operatorname{mat}^{\mathcal{V}}(y).
\end{aligned}$$

Comparando las expresiones resulta el siguiente teorema fundamental

Teorema 153 *Dadas dos bases $\mathcal{U}$ y $\mathcal{V}$ de $\mathbb{E}$, la relación entre las matrices de β en estas bases es*

$$\operatorname{mat}_{\mathcal{V}\mathcal{V}}(\beta) = \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}). \tag{8.3}$$

Hay dos reglas mnemotécnicas como las que vimos para endomorfismos

$$\beta(x,y) = \operatorname{mat}^{\mathcal{U}}(x)^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}^{\mathcal{U}}(y) \text{ es similar a } 1 = U\frac{1}{UU}U,$$
$$\operatorname{mat}_{\mathcal{V}\mathcal{V}}(\beta) = \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta) \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \text{ es similar a } \frac{1}{VV} = \frac{U}{V}\frac{1}{UU}\frac{U}{V},$$

pero sin olvidar que hay que poner la trasposición $\bullet^\top$ en el lugar correcto (¡esencial!).

En el ejemplo de (8.1) conocemos $\operatorname{mat}_{\mathcal{E}\mathcal{E}}(\beta)$. Tomemos otra base $\mathcal{U}$ de $\mathbb{E} = \mathbb{R}_2[X]$, digamos que $\mathcal{U} = \left(1, 1 + X, 1 + X + X^2\right)$. La matriz $\operatorname{mat}_{\mathcal{U}\mathcal{U}}(\beta)$ se calcula como

$$\operatorname{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\mathbb{E}}) \operatorname{mat}_{\mathcal{E}\mathcal{E}}(\beta) \operatorname{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\mathbb{E}}) = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^\top \begin{pmatrix} 1 & 3 & 9 \\ 2 & 6 & 18 \\ 4 & 12 & 36 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 13 \\ 3 & 12 & 39 \\ 7 & 28 & 91 \end{pmatrix}.$$

Como se ve, pasar de unas matrices a otras basta conocer las matrices de cambio y multiplicarlas. Pesado pero rutinario.

En ciertas circunstancias, para demostraciones teóricas, interesa la forma de (8.3) con sumatorios. Aunque se puede traducir la igualdad matricial a sumatorios (todo es cosa de igualar los n^2 coeficientes) es posiblemente más fácil un cálculo comenzando desde cero. Eso sí, para no perdernos en la jungla de índices usaremos la convención de Einstein y las bases y coordenadas no serán $\mathcal{U}$ y $\mathcal{V}$ sino $\mathcal{U}=(u_1,\ldots,u_n)$ y $(u_{1'},\ldots,u_{n'})$ teniendo coordenadas x^i y $x^{i'}$. Con todo esto,

$$\begin{aligned} b_{i'j'}=\beta\left(u_{i'},u_{j'}\right)&=\beta\left(\sum_{p=1}^{n}c_{i'}^{p}u_p,\sum_{q=1}^{n}c_{j'}^{q}u_q\right)=\sum_{p=1}^{n}c_{i'}^{p}\beta\left(u_p,\sum_{q=1}^{n}c_{j'}^{q}u_q\right)\\ &=\sum_{p=1}^{n}c_{i'}^{p}\left(\sum_{q=1}^{n}c_{j'}^{q}\beta\left(u_p,u_q\right)\right)=\sum_{p,q=1}^{n}c_{i'}^{p}c_{j'}^{q}\beta\left(u_p,u_q\right)=\sum_{p,q=1}^{n}c_{i'}^{p}c_{j'}^{q}b_{pq}. \end{aligned}$$

El lector habrá imaginado con poco esfuerzo que $(c_{i'}^{p})$ es matriz de cambio de base (¿cuál de los dos cambios?) y que $(b_{i'j'})$ y (b_{pq}) son las matrices de β en las bases $\mathcal{U}$ y $\mathcal{U}'$. También debe tener en cuenta que la ecuación matricial (8.3) se traduce en las n^2 ecuaciones escalares de más arriba. Resumimos

$$b_{i'j'}=\sum_{p,q=1}^{n}c_{i'}^{p}c_{j'}^{q}b_{pq}\ ,\ 1\leq i',j'\leq n\quad\text{supuesto que } u_{i'}=\sum_{p=1}^{n}c_{i'}^{p}u_{ppq}\ ,\ 1\leq i'\leq n.$$

Advertimos que si nos dan una fórmula para β con la que calculamos $\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)$ y aparece luego otra base $\mathcal{U}'$ habrá que estudiar si es más conveniente usar el teorema 153 o calcular directamente. Un ejemplo de esto. En el problema 289 se estudia en $\mathbb{E}=\mathbb{R}_n[X]$ la forma bilineal $\beta\left(P(X),Q(X)\right)=P(1)\,Q(1)$ y se pide su matriz para la base estándar $\mathcal{E}$ y para $\mathcal{V}=\left(1,1+X,1+X^2,\ldots,1+X+X^2+\ldots+X^n\right)$. El lector habrá calculado por aplicación directa de la definición que la matriz $\mathrm{mat}_{\mathcal{E}\mathcal{E}}(\beta)=b$ tiene todo unos y que la de β en $\mathcal{V}$ es

$$\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)=\begin{pmatrix} 1\cdot1 & 1\cdot2 & \cdots & 1\cdot(n-1) & 1\cdot n\\ 2\cdot1 & 2\cdot2 & \cdots & 2\cdot(n-1) & 2\cdot n\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ (n-1)\cdot1 & (n-1)\cdot2 & \cdots & (n-1)\cdot(n-1) & (n-1)\cdot n\\ n\cdot1 & n\cdot2 & \cdots & n\cdot(n-1) & n\cdot n \end{pmatrix}=\bar{b}.$$

Más brevemente, $\bar{b}_{ij}=ij$. La matriz de cambio que expresa $\mathcal{V}$ en función de $\mathcal{E}$ es

$$c=\mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id})=\begin{pmatrix} 1 & 1 & \cdots & 1 & 1\\ 0 & 1 & \cdots & 1 & 1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & 1\\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

La teoría (no el cálculo directo) dice que $\bar{b}=c^{\top}bc$, pero opinamos que es más fácil calcular directamente $\bar{b}$ que hacer el producto $\bar{b}=c^{\top}bc$.

Una parte muy importante del capítulo será investigar si se puede sustituir una base $\mathcal{U}$ en la que $\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)$ es "complicada" por otra $\mathcal{V}$ en la que $\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)$ es "sencilla". Nos ahorrará muchos cálculos con índices el recordar algunas cuestiones sobre operaciones elementales limitándonos a matrices cuadradas. Hay tres operaciones elementales con filas y tres con columnas. Las de tipo **(1)** permutan dos filas/columnas; las de tipo **(2)** suman a una fila/columna un múltiplo de otra, y las de tipo **(3)** multiplican una fila/columna por un escalar no nulo. En todos los casos las restantes filas/columnas no se alteran. Si F (respectivamente K) es una operación de filas (columnas) tenemos matrices elementales $\phi=F(I)$ o $\kappa=(I)\,K$ y para cualquier matriz cuadrada a, $F(a)=\phi a$ y $(a)\,K=a\kappa$. Se podían hacer operaciones elementales también con sucesiones de vectores, pero nosotros nos limitamos a bases $(u_1,\ldots,u_n)=\mathcal{U}$. Las operaciones son **(1)** Permutar u_i con u_j **(2)** Sustituir u_i por $u_i+\lambda u_j$ y **(3)** Sustituir u_i por μu_i, $\mu\neq0$. En todos los casos los restantes vectores de $\mathcal{U}$ no se alteran. Nos preguntamos si se pasa $\mathcal{U}$ a $\mathcal{V}$ por una operación elemental E ¿cómo es la matriz de cambio $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})=c$? Si, por ejemplo, E es del

tipo **(2)** en c_k han de estar las coordenadas de v_k en la base $\mathcal{U}$. Si solo se ha alterado $v_i = u_i + \lambda u_j$ tendremos que, excepto para c_i, todas las columnas serán $c_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ con solo 1 en el lugar k. Por otra parte, $c_j = (*, \ldots, *, \lambda, *, \ldots, *)$ con λ en el lugar j y todas las estrellas 0 excepto un 1 en el lugar i. Lo que hemos descrito como c es la matriz $c_k = e_k$ si $k \neq i$ y $c_i = e_i + \lambda e_j$; es decir, $(I)\, K = \kappa$ siendo K la operación columna "sumar a la columna i la columna j multiplicada por λ". Lo que hemos probado, para operaciones tipo **(2)**, pero se hace igualmente para las otras.

Teorema 154 *Si se pasa de $\mathcal{U}$ a $\mathcal{V}$ por la operación elemental E, que corresponde a la operación por columnas K, se tiene que $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}} = (I)\, K = \kappa$, siendo κ la matriz elemental que corresponde a E.*

Un ejemplo. Sea $\dim \mathbb{E} = 5$, $v_4 = u_4 - 9u_2$ y $v_k = u_k$ si $k \neq 4$, luego $i = 4$ y $j = 2$. Entonces

$$c = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -9 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad (v_1, \ldots, v_5) = (u_1, \ldots, u_5) \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -9 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Teorema 155 *Pasemos de $\mathcal{U}$ a $\mathcal{V}$ por una operación elemental E con operaciones homologas F y K para filas y columnas y matrices elementales $\phi = F(I)$ y $\kappa = (I)\, K$. Entonces*

$$\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta) == \kappa^{\top} \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa = \phi\, \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa.$$

Demostración. Por la fórmula (8.3),

$$\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})^{\top} \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = \kappa^{\top} \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa \overset{*}{=} \phi\, \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa$$

con una simple aplicación del teorema 154. Hay que justificar $\overset{*}{=}$; o sea, que $\phi = \kappa^{\top}$, que es inmediato, aunque mentalmente trabajoso, con solo aplicar las definiciones de κ y trasposición. Supongamos por ejemplo que κ se obtenga sumando a la *columna* i de I la *columna* j de I multiplicada por λ y sin alterar las otras *columnas*. Como trasponer es cambiar *filas* por *columnas*, en $\kappa^{\top}$ deducimos que no se han alterado las *filas* distintas de la *fila* i y se ha sumado a la *fila* i la *fila* j multiplicada por λ. ¿Cómo es entonces $\kappa^{\top}$? Es la matriz que resulta al hacer en I las mismas operaciones hechas para transformar I en κ pero sustituyendo *columna(s)* por *fila(s)*.[3] ♣

Si tuviesemos como dato las operaciones $E_1, \ldots, E_h$ que se han realizado para pasar de $\mathcal{U}$ a $\mathcal{V}$, correspondientes a matrices elementales $\kappa_1, \ldots, \kappa_h$ o $\phi_1, \ldots, \phi_h$, se deduciría del teorema 155 que

$$\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta) = \kappa_h^{\top} \cdots \kappa_1^{\top} \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa_1 \cdots \kappa_h = \phi_h \cdots \phi_1\, \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\, \kappa_1 \cdots \kappa_h$$

Ademas, con el teorema 154 se conoce $\mathcal{V}$ en función de $\mathcal{U}$ porque

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}) = (I)\, K_1 \circ \cdots \circ K_h;$$

o sea, aplicando a I las operaciones columna homólogas a las hechas al transformar la base.

No obstante, vamos a realizar todo esto en orden inverso. ¿Qué queremos decir? Nuestro objetivo es mostrar que si β es lo que definiremos como una **forma bilineal simétrica**, que es aquella que cumple $\beta(x, y) = \beta(y, x)$ para todo x, y, podemos encontrar a partir de cualquier base $\mathcal{U}$ otra base $\mathcal{V}$ en donde $\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)$ será sumamente sencilla e informativa. El procedimiento práctico tomará la matriz $\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta) = s$ en una base cualquiera y, trabajando sobre s con operaciones fila y columna, alcanzar la matriz "buena" $\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)$. Este proceso se explicará con más detalle en este capítulo pero vamos a anticipar un ejemplo, Nos da $\beta : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ que en $\mathcal{E}$ tiene matriz

$$b = \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 2 \end{pmatrix}$$

[3] En realidad podríamos tener desde el capítulo primero un teorema que dijese que $\kappa^{\top}$ es la matriz que resulta al hacer en I las operaciones *fila* homólogas a las operaciones *columna* que se han hecho para transformar I en κ pero sustituyendo en la regla *columna(s)* por *fila(s)*. Hay otro aserto similar para $\phi^{\top}$.

Hacemos operaciones fila/columna

$$\begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 4-4\cdot 1 \\ 0 & 1 & 0-4\cdot 0 \\ 4 & 0 & 2-4\cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & -14 \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4-4\cdot 1 & 0-4\cdot 0 & -14-4\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -14 \end{pmatrix} = d$$

Hemos aplicado en $\to$ la operación "restar a la *columna* 3 la *columna* 1 multiplicada por 3" y en $\Rightarrow$ "restar a la *fila* 3 la *fila* 1 multiplicada por 4". Al final nos queda d que es una matriz diagonal. Quiere decirse que si en vez de la base estándar $\mathcal{E}$ usamos la base $\mathcal{V}$ con matriz de cambio

$$\mathrm{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id}) = \begin{pmatrix} 1 & 0 & 0-4\cdot 1 \\ 0 & 1 & 0-4\cdot 0 \\ 0 & 0 & 1-4\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \text{ o sea, } \mathcal{V} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} \right),$$

en $\mathcal{V}$ tiene β matriz diagonal $d = \mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)$. Como comprobación parcial, $\beta(v_1, v_3)$ y $\beta(v_1, v_3)$ son

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} = 0, \qquad \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} = -14.$$

Obtendremos en el futuro como teorema general que para cualquier β *simétrica* podremos construir una base en donde su matriz sea diagonal. Es un teorema de similar importancia al teorema de diagonalización de endomorfismos (pero aquí trabajamos con funciones bilineales, aunque simétricas).

8.1.3. Distinciones entre endomorfismos y funciones bilineales

Hacemos unos comentarios generales para no confundir cuestiones propias de formas bilineales con cuestiones propias de endomorfismos, motivadas las confusiones porque ambos objetos se representan, una vez elegida una base $\mathcal{U}$, por una matriz $n \times n$, siendo $n = \dim(\mathbb{E})$. Como advertencia general: las matrices son "coordenadas de objetos", pero el que diversos objetos tengan el mismo tipo de coordenadas no implica en absoluto que tengan igual naturaleza.

1. Si $\mathbb{E} = \Bbbk^n$, tanto los endomorfismos como las formas bilineales son identificables con matrices $a, b \in \Bbbk^{n\times n}$. En el caso de los endomorfismos, a través de $L(x) = ax$ (se ve x como vector columna), y en el caso de las formas a través de $\beta(x, y) = x^\top by$ (se ven tanto x como y como vectores columna, por eso, para poder multiplicar las matrices, se necesita poner $x^\top$).

2. Si $\mathbb{E}$ es arbitrario y tenemos una base $\mathcal{U}$, podemos asignar matrices a, b de $\Bbbk^{n\times n}$ tanto al endomorfismo L como a la forma β. Las reglas son:

$$\text{si } L(u_j) = \sum_{i=1}^n a_j^i u_j \quad (1 \le j \le n), \text{ entonces } a = \left(a_j^i\right) \in \Bbbk^{n\times n} \text{ es la matriz de } L,$$

$$\text{si } \beta(u_i, u_j) = b_{ij}, \text{ entonces } b = (b_{ij}) \in \Bbbk^{n\times n} \text{ es la matriz de } \beta.$$

   Esto facilita los cálculos porque, si $x = \sum_{i=1}^n x^i u_i$ e $y = \sum_{j=1}^n y^j u_j$, sea con matrices o con sumatorios,

$$\begin{pmatrix} L(x)^1 \\ \vdots \\ L(x)^n \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix}, \quad (L(x))^i = \sum_{j=1}^n a_j^i x^j,$$

$$\beta(x, y) = \left(x^1, \ldots, x^n\right) \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix}, \quad \beta(x, y) = \sum_{i,j=1}^n b_{ij} x^i y^j.$$

3. Con la notación mat, pesada de manejar pero muy fiable, las fórmulas anteriores son

$$\mathrm{mat}^{\mathcal{U}}(L(x)) = \mathrm{mat}^{\mathcal{U}}_{\mathcal{U}}(L)\,\mathrm{mat}^{\mathcal{U}}(x), \quad \beta(x,y) = \left[\mathrm{mat}^{\mathcal{U}}(x)\right]^{\top}\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\,\mathrm{mat}^{\mathcal{U}}(x).$$

La posición de $\mathcal{U}$ se elige para recordarlas por analogía con una cancelación de fracciones.

4. Esto ayuda también al relacionar fórmulas para dos bases distintas. Si tenemos dos bases $\mathcal{U}$ y $\mathcal{V}$,

$$\mathrm{mat}^{\mathcal{V}}_{\mathcal{V}}(L) = \mathrm{mat}^{\mathcal{V}}_{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}^{\mathcal{U}}_{\mathcal{U}}(L)\,\mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) = \left[\mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})\right]^{-1}\mathrm{mat}^{\mathcal{U}}_{\mathcal{U}}(L)\,\mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}),$$

$$\mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta) = \left[\mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})\right]^{\top}\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)\,\mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}).$$

No se relacionan las matrices del mismo modo, aunque siempre interviene $c = \mathrm{mat}^{\mathcal{U}}_{\mathcal{V}}(\mathrm{id}_{\mathbb{E}})$. Con una notación menos florida, las matrices a y a' de L se relacionan por $a' = c^{-1}ac$ mientras que las matrices b y b' de β se relacionan por $b' = c^{\top}bc$.

Ya dijimos en *Funciones lineales* que hay una relación llamada de **semejanza**, **similaridad** o **conjugación** (las matrices son semejantes, similares o conjugadas) dada por $a' \overset{1}{\sim} a$ si existe c *invertible* tal que $a' = c^{-1}ac$. Otra importante relación es la **congruencia** (las matrices son congruentes) dada por $b' \overset{2}{\sim} b$ si existe c *invertible* tal que $b' = c^{\top}ac$. Podemos pues decir que las matrices de un mismo endomorfismo respecto de bases diferentes son semejantes y (aún no se ha probado) que las matrices de una misma forma bilineal *simétrica* respecto de bases diferentes son congruentes.

Cada vez que el lector descubra en un libro clásico de teoría de matrices un teorema que le diga que a es semejante a otra matriz a' "interesante", sepa que subyace un teorema que le dice que un endomorfismo admite una base donde L se expresa de modo "interesante". De modo análogo podrá encontrar teoremas[4] que le dicen que β *simétrica* (supone que $\beta(x,y) = \beta(y,x)$) tiene matriz s, se puede conseguir que s sea congruente con una matriz d diagonal. La traducción es que se podrá llegar a una base donde la matriz d de β sea diagonal; es decir $\beta(v_i, v_j) = 0$ si $i \neq j$. Hay por tanto que tener cierto cuidado. Un valor propio λ de la matriz a tiene interés mientras se vea a como expresión un endomorfismo de $\mathbb{k}^n$ pero no si se ve a como expresión de una función bilineal. Pasa lo mismo al hablar de diagonalizar una matriz a. Hemos dicho que se puede hacer de dos maneras y la base "buena" tiene interés según lo que represente a.

Problema 292 *En un libro clásico de teoría de matrices leemos un teorema que nos dice que una matriz cuadrada $a \in \mathbb{k}^{n\times n}$ con las propiedades*

1. *Su polinomio característico $C(X)$ se puede escribir como $(\lambda_1 - X)^{m_1}\cdots(\lambda_p - X)^{m_p}$, siendo $m_1 + \ldots + m_p = n$*
2. *Para todo $i = 1,\ldots,p$ se cumple que $\mathrm{rg}(a - \lambda_i) = n - m_i$*

es semejante a una matriz diagonal con $\lambda_1,\ldots,\lambda_p$ en ella, apareciendo m_i veces cada λ_i
¿Qué teorema es ese? Explicarlo.

8.2. Formas bilineales simétricas

Diremos que β es **simétrica** (respectivamente **antisimétrica**) si para todo $x, y \in \mathbb{E}$ se tiene $\beta(x,y) = \beta(y,x)$ (respectivamente $\beta(x,y) = -\beta(y,x)$). Obsérvese que en característica 2 la simetría y antisimetría son equivalentes y por ello *supondremos siempre que $\mathbb{k}$ no tiene característica* 2.

Teorema 156 *Es necesario y suficiente para que β sea simétrica (antisimétrica) que haya al menos una base $\mathcal{U}$ donde la matriz de β sea simétrica (antisimétrica). Si esto sucede, en cualquier otra base $\mathcal{V}$ su matriz también será simétrica (respectivamente antisimétrica).*

[4] Con un lenguaje geométrico, lo veremos en este capítulo.

Demostración. La primera parte es muy fácil y queda para el lector. Probamos la segunda. Supongamos que en $\mathcal{U}$ la matriz b de β es simétrica. En otra base $\mathcal{V}$ se tendrá $b' = c^\top bc$. Entonces

$$(b')^\top = \left(c^\top bc\right)^\top = c^\top b^\top \left(c^\top\right)^\top \overset{*}{=} c^\top bc = b'.$$

Se ha usado en $\overset{*}{=}$ que $b^\top = b$ y que trasponer dos veces deja invariante la matriz. ♣

El caso antisimétrico es importante, pero *nos vamos a centrar en una forma bilineal simétrica σ de $\mathbb{E}$*. Reservaremos la letra σ para denotar este tipo de formas y s para su simétrica matriz.

Problema 293 *Sea σ una forma simétrica. Se tienen las* ***identidades polares***

$$\sigma(x,y) = \frac{1}{2}\left[\sigma(x+y,x+y) - \sigma(x,x) - \sigma(y,y)\right] = \frac{1}{4}\left[\sigma(x+y,x+y) - \sigma(x-y,x-y)\right]$$

Se deduce de la primera que si $\sigma \neq 0$ debe existir un x tal que $\sigma(x,x) \neq 0$.

Dijimos en su momento que $\varepsilon(x,y) = \sum_{i=1}^n x^i y^i$ en $\mathbb{E} = \Bbbk^n$ es la forma bilineal más sencilla y sirve como referencia. Sin embargo una forma simétrica σ puede tener un comportamiento muy diferente a ε. Puede suceder $\sigma(x,x) = 0$ pero $x \neq 0$ o que exista $x \neq 0$ tal que para todo y sea $\sigma(x,y) = 0$. Es muy posible que el lector haya tenido contacto con ε para $\Bbbk = \mathbb{R}$, pues es el producto escalar ordinario. Si se piensa que σ generaliza a ε debe quedar claro que en parte sí, pero que hay importantísimas diferencias si buscamos prever el comportamiento de σ en términos de longitud y perpendicularidad. Son una guía intuitiva pero llena de peligros para una σ general. Como ejemplos raros basta considerar en $\mathbb{R}^2$ la forma σ con matriz cero excepto 1 en σ_{11}.

Se define el **espacio nulo** de σ como (no hay dos definiciones pues σ es simétrica)

$$\mathbb{K} = \{x \in \mathbb{E} \mid \forall y \in \mathbb{E},\ \sigma(x,y) = 0\} = \{y \in \mathbb{E} \mid \forall x \in \mathbb{E},\ \sigma(x,y) = 0\}.$$

Es muy fácil ver que $\mathbb{K}$ es un subespacio vectorial. Por ejemplo, si $x_1, x_2 \in \mathbb{K}$ se tiene

$$\sigma(x_1 + x_2, y) = \sigma(x_1, y) + \sigma(x_2, y) = 0 + 0 = 0,$$

de modo que $x_1 + x_2 \in \mathbb{K}$. Vamos a determinar las ecuaciones de $\mathbb{K}$ dada una base $\mathcal{U}$. Partimos de que la condición $\sigma(x,y) = 0$ para todo x equivale a $\sigma(u_i, x) = 0$ para $i = 1, \ldots, n$ (lo verificará el lector). Con las σ_{ij} queda un sistema lineal homogéneo o, equivalentemente, una ecuación matricial

$$\left\{\begin{array}{l} 0 = s(u_1,y) = \sum_{j=1}^n s_{1j}y^j = s_{11}y^1 + \cdots + s_{1n}y \\ 0 = s(u_2,y) = \sum_{j=1}^n s_{2j}y^j = s_{21}x^1 + \cdots + s_{2n}x^m \\ \vdots \\ 0 = s(u_n,y) = \sum_{j=1}^n s_{nj}y^j = s_{n1}x^1 + \cdots + s_{nn}x^n \end{array}\right., \qquad \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix}\begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

En forma más compacta, si $s = (s_{ij})$, el sistema es $sy = 0$. Los $y \in \mathbb{E}$ cuya matriz de coordenadas $\mathrm{mat}^{\mathcal{U}}(y) = \left(y^1, \ldots, y^n\right)$ sea solución de $sy = 0$, constituyen el espacio nulo $\mathbb{K}$. Se define el **nulidad** de σ como la dimensión de $\mathbb{K}$. Es obvio que si tenemos la matriz s de σ en una base $\mathcal{U}$ y esta matriz tiene rango r, la dimensión de $\mathbb{K}$, que es la del espacio de soluciones de $sy = 0$ será $n - r$. El **rango de** σ es la diferencia de dimensiones $n - \dim(\mathbb{K})$, luego, si se dispone de la matriz s, el rango de σ es el de s, *sin que importe la base utilizada para calcularla.* Obsérvese que si a una matriz simétrica a le asignamos un endomorfismo L de $\Bbbk^n$ y una forma simétrica σ de $\Bbbk^n$ por $L(x) = ax$ y $\sigma(x,y) = x^\top ay$ los vectores solución de $ax = 0$ tienen una doble interpretación: en términos de L son los transportados a 0 (los de $\ker(L)$) y en términos de σ los "perpendiculares" a todo vector de $\Bbbk^n$. Ambos tipos de vectores forman un subespacio y su dimensión es el rango de a.

Las formas σ con $\mathbb{K} = 0$ se llaman **no degeneradas** y las que tienen $\mathbb{K} \neq 0$ se llaman **degeneradas**. Viendo intuitivamente $\sigma(x,y) = 0$ como una "perpendicularidad", las formas degeneradas son las que poseen vectores no nulos ortogonales a todos los demás. Para que σ sea no degenerada es condición equivalente que su rango sea $n = \dim(\mathbb{E})$ y esto a su vez equivale a que en cualquier base sea $\mathrm{mat}_{\mathcal{UU}}(\sigma)$ de rango n (equivale a invertible o con determinante no nulo).

Por ejemplo, si nos dan formas σ que en cierta base $\mathcal{U}$ de $\mathbb{E}$ tienen matrices

$$s = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \qquad t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{8.4}$$

vemos que la primera forma tiene rango 3 y nulidad 0 pues tiene determinante 2. La segunda matriz tiene rango 2 y nulidad 1. Si resolvemos

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} y \\ x+z \\ y \end{pmatrix} = 0, \qquad \begin{cases} y = 0 \\ x+z = 0 \end{cases}$$

llegamos a que el segundo espacio nulo es el generado por $(1,0,-1)^\top$. Si nos preguntan si s y t pueden ser matrices de una misma σ para bases diferentes, la respuesta es no. En efecto, si lo fueran, el espacio nulo $\mathbb{K}$ de σ tendría, según se calculase con s o t dimensión 0 o 1.

Problema 294 *Generalizar a dimensión arbitraria; o sea, a σ que en una base tiene matriz s con todo 1 en las dos diagonales junto a la principal y los demás coeficientes cero.* ♦

Solución. El sistema en los casos $n = 4, 5$ resulta igualando a cero los vectores

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} y \\ x+z \\ t+y \\ z \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \\ w \end{pmatrix} = \begin{pmatrix} y \\ x+z \\ t+y \\ w+z \\ t \end{pmatrix}.$$

El sistema del primer caso solo existe la solución 0 luego el rango es cuatro y $\mathbb{K} = 0$, y en el segundo caso son solución los vectores proporcionales a $(-1,0,1,0,-1)^\top$, luego en rango es $5-1=4$ y estos vectores (mejor dicho, los vectores que en la base con la que se calculó s tienen estas coordenadas) forman $\mathbb{K}$. Queda para el lector el caso general. ♦

Problema 295 *Tenemos dos formas simétricas σ_1 y σ_2 con matrices*

$$s_1 = \begin{pmatrix} 1 & & & 0 & & & 1 \\ & \ddots & & \vdots & & \diagup & \\ & & 1 & 0 & 1 & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & 1 & 0 & 1 & & \\ & \diagup & & \vdots & & \ddots & \\ 1 & & & 0 & & & 1 \end{pmatrix} \in \mathbb{k}^{(2k+1)\times(2k+1)}, \qquad s_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 0 \\ 3 & 0 & h \end{pmatrix},$$

s_1 en equis" y s_2 con un parámetro h. Determinar los rangos y los h que dan rango < 3.

Problema 296 *Consideremos $\sigma = f \otimes f$ para $f \in \mathbb{E}^*$. Supuesta $f \neq 0$ probar que* $\dim \mathbb{K} = m-1$ *y que* $\mathbb{K} = \ker(f)$.

Si $\mathbb{K}$ es grande, la matriz de σ se puede conseguir relativamente sencilla porque tendrá muchos ceros.

Teorema 157 *Sea $\mathbb{F}$ un suplementario de $\mathbb{K}$ en $\mathbb{E}$; es decir, $\mathbb{E} = \mathbb{F} \oplus \mathbb{K}$. Si $\{u_1, \ldots, u_r\}$ es base de $\mathbb{F}$ y $\{u_{r+1}, \ldots, u_{r+k}\}$ es base de $\mathbb{K}$ se tiene que la matriz de σ en esa base tiene la forma*

$$\begin{pmatrix} (s_{ij}) & 0_{r\times k} \\ 0_{k\times r} & 0_{k\times k} \end{pmatrix},$$

una matriz $n \times n$ con $n = r + k$ siendo (s_{ij}) de dimensión $r \times r$ y los otros 0 representando matrices $r \times k$, $k \times k$ y $k \times r$. Además la matriz (s_{ij}) de la esquina superior izquierda es invertible.

La demostración es inmediata con la definición de espacio nulo y queda como problema para el lector. Para el siguiente teorema fundamental necesitamos otro.

Teorema 158 *Sea $f : \mathbb{E} \to \mathbb{k}$ una forma lineal y v un vector tal que $f(v) \neq 0$. Se tiene la descomposición en suma directa $\mathbb{E} = \lg(v) \oplus \ker(f)$. En ella cada $x \in \mathbb{E}$ se descompone en la forma*

$$x = \frac{f(x)}{f(v)} v + \left[x - \frac{f(x)}{f(v)} v \right].$$

Demostración. El segundo sumando está en $\ker(f)$, cosa inmediata porque

$$f\left(x-\frac{f(x)}{f(v)}v\right)=f(x)-\frac{f(x)}{f(v)}f(v)=0.$$

Si $x\in \lg(v)\cap\ker(f)$ escribimos $x=\lambda v$ y entonces $0=f(x)=\lambda f(v)$ da $\lambda=0$ y $x=0$. ♣

Teorema 159 (de diagonalización) *Hay una base $\mathcal{W}$ para σ tal que $\sigma(w_i,w_j)=0$ cuando $i\neq j$, luego $\mathrm{mat}_{\mathcal{W}\mathcal{W}}(\sigma)$ es diagonal.*

Demostración. Se hace la demostración por inducción sobre $n=\dim(\mathbb{E})$. Los casos con $n=0,1$ son obvios así como el caso $\sigma=0$ en cualquier dimensión. Supongamos por tanto $\sigma\neq 0$ y cierto el teorema para dimensiones $<n$. Hay algún para (x,y) tal que $\sigma(x,y)\neq 0$ y ya dijimos en el problema 293 que si $\sigma\neq 0$ puede elegirse ese par con $x=y$. Tomamos uno de estos $x=w_1$ y consideramos $f:\mathbb{E}\to\Bbbk$, $f(y)=\sigma(w_1,y)$. El teorema 158 nos da $\mathbb{E}=\lg\{w_1\}\oplus\ker(f)$. Sea $\tau:\ker(f)\times\ker(f)\to\Bbbk$ la restricción de σ. Aplicando la hipótesis inductiva tenemos una base $(w_2,\ldots,w_n)$ de $\ker(f)$ tal que $\tau(w_i,w_j)=0$ para $2\leq i,j$ e $i\neq j$. Pero aunque sea $i=1$ se cumple $\sigma(w_1,w_j)=f(w_j)=0$ para $j\geq 2$. La base $(w_1,w_2,\ldots,w_n)$ cumple el teorema ♣

¿Cómo se calcula en la práctica la base $\mathcal{W}$? Se suele dar una base $\mathcal{U}$ y $s=\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\sigma)$. La idea es hacer operaciones elementales sobre s tal como se han descrito en la sección precedente. Estas operaciones se hacen también sobre $\mathcal{U}$ y, cuando se ha transformado s en d diagonal, $\mathcal{U}$ se ha transformado en $\mathcal{W}$, la base que diagonaliza. Detallamos una estrategia sistemática.

1. Supongamos primeramente que $s_{11}\neq 0$. Restamos a las columnas $2,3,\ldots,,n$ la columna 1 multiplicada por $s_{12}/s_{11},\ldots,s_{1n}/s_{11}$, y queda

$$s\to\begin{pmatrix} s_{11} & s_{12}-\frac{s_{12}}{s_{11}}\cdot s_{11} & \cdots & s_{1n}-\frac{s_{12}}{s_{11}}\cdot s_{11}\\ s_{21} & s_{22}-\frac{s_{12}}{s_{11}}\cdot s_{21} & \cdots & s_{2n}-\frac{s_{12}}{s_{11}}\cdot s_{21}\\ \vdots & \vdots & \ddots & \vdots\\ s_{n1} & s_{n2}-\frac{s_{12}}{s_{11}}\cdot s_{n1} & \cdots & s_{nn}-\frac{s_{12}}{s_{11}}\cdot s_{n1}\end{pmatrix}=\begin{pmatrix} s_{11} & 0 & \cdots & 0\\ s_{21} & t_{22} & \cdots & t_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ s_{n1} & t_{n2} & \cdots & t_{nn}\end{pmatrix}$$

   A continuación se hacen las operaciones homólogas con filas y esta última matriz se convierte en[5]

$$\Rightarrow\begin{pmatrix} s_{11} & 0 & \cdots & 0\\ s_{21}-\frac{s_{12}}{s_{11}}\cdot s_{11} & t_{22} & \cdots & t_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ s_{n1}-\frac{s_{1n}}{s_{11}}\cdot s_{11} & t_{n2} & \cdots & t_{nn}\end{pmatrix}=\begin{pmatrix} s_{11} & 0 & \cdots & 0\\ 0 & t_{22} & \cdots & t_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & t_{n2} & \cdots & t_{nn}\end{pmatrix}=t$$

   Los ceros finales de la primera columna utilizan la simetría de s ya que

$$s_{j1}-\frac{s_{1j}}{s_{11}}\cdot s_{11}=s_{j1}-\frac{s_{j1}}{s_{11}}\cdot s_{11}=0.$$

   Obsérvese que las t no se alteran porque la primera fila tiene todo ceros salvo en el primer lugar. Esto hace que la operación $\Rightarrow$ sea rápida, bastando poner ceros en los lugares $(2,1),(3,1),\ldots,(3,n)$. Las $(n-1)$ operaciones son elementales de tipo **2** y si se corresponden con matrices q_j se tiene $t=q_{(n-1)}^{\top}\cdots q_1^{\top}sq_1\cdots q_{(n-1)}$. La matriz t es simétrica. Hemos pasado de la base inicial $\mathcal{U}$ a una base intermedia $\mathcal{V}$ tal que $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})=q_1\cdots q_{(n-1)}$. Si queremos llevar la cuenta de cómo es $\mathcal{V}$, haremos las operaciones columnas de las q_j a la matriz unidad I.

2. Supongamos, aunque no siempre sucederá, que sea $t_{22}\neq 0$. Haríamos $n-2$ operaciones restando a las columnas $3,4,\ldots,n$ la columna 2 multiplicada por $t_{23}/t_{22},\ldots,s_{2n}/t_{22}$, y luego se hace la operación homóloga por filas. Queda

$$\begin{pmatrix} s_{11} & 0 & 0 & \cdots & 0\\ 0 & t_{22} & 0 & \cdots & 0\\ 0 & 0 & z_{33} & \cdots & z_{3n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & z_{n3} & \cdots & z_{nn}\end{pmatrix}$$

[5] Para ayudar a seguir los cálculos, al hacer la operación homóloga por filas ponemos $\Rightarrow$ en vez de $\to$

de modo que en las dos primeras filas y columnas solo s_{11} y t_{22} son no nulos. Si $z_{33} \neq 0$ se puede repetir el proceso y, en general, contando con que el último término en posición (i,i) sea no nulo, llegar hasta una matriz diagonal con los términos nulos, si los hay, al final.

Tenemos pendiente saber qué se hace si aparecen términos nulos en la diagonal y no al final, lo que pronto trataremos, pero vamos a ver primero un ejemplo.

Sea $s = \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\beta)$ la matriz

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

Las operaciones son

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2-\frac{2}{1}\cdot 1 & 3-\frac{3}{1}\cdot 1 \\ 2 & 1-\frac{2}{1}\cdot 2 & 2-\frac{3}{1}\cdot 2 \\ 3 & 2-\frac{2}{1}\cdot 3 & 1-\frac{3}{1}\cdot 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & -4 \\ 3 & -4 & -8 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & -4 \\ 0 & -4 & -8 \end{pmatrix}.$$

La doble flecha $\Rightarrow$ indica que se hacen las operaciones con filas homólogas de las que en $\to$ se han hecho para columnas. Como la parte teórica nos dice que la forma final ha de ser simétrica y que los coeficientes fuera de la primera columna no se verán alterados, $\Rightarrow$ es muy rápido pues basta poner 0 en la primera columna salvo en el lugar $(1,1)$. Este es el paso de s a t descrito en general, Como $t_{22} = -3 \neq 0$, seguimos y

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & -4 \\ 0 & -4 & -8 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & -4-\left(\frac{-4}{-3}\right)\cdot(-3) \\ 0 & -4 & -8-\left(\frac{-4}{-3}\right)\cdot(-4) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & -4 & -\frac{8}{3} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -\frac{8}{3} \end{pmatrix}$$

Esta última matriz es la de β en una base $\mathcal{V}$ que diagonaliza. Para conocer la base que diagonaliza las operaciones son

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0-\frac{2}{1}\cdot 1 & 0-\frac{3}{1}\cdot 1 \\ 0 & 1-\frac{2}{1}\cdot 0 & 0-\frac{3}{1}\cdot 0 \\ 0 & 0-\frac{2}{1}\cdot 0 & 1-\frac{3}{1}\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\to \begin{pmatrix} 1 & -2 & -3-\left(\frac{-4}{-3}\right)\cdot(-2) \\ 0 & 1 & 0-\left(\frac{-4}{-3}\right)\cdot 1 \\ 0 & 0 & 1-\left(\frac{-4}{-3}\right)\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 & -2 & -\frac{1}{3} \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & 1 \end{pmatrix}.$$

Las tres columnas de la matriz dan una base que diagonaliza. Se puede comprobar que (¡como debe ser!)

$$\begin{pmatrix} 1 & -2 & -\frac{1}{3} \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & 1 \end{pmatrix}^{\top} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -\frac{1}{3} \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -\frac{8}{3} \end{pmatrix}.$$

3. Supongamos que en **1** tuviésemos $s_{11} = 0$. Si hay algún $i > 1$ tal que $s_{ii} \neq 0$ se permuta la columna i con la 1 y luego, en operación homóloga, la fila i con la 1. Ahora tendremos s_{ii} en el lugar $(1,1)$ y podremos proceder como en 1. Si llevamos cuentas de la base, la nueva base es la antigua trasponiendo el vector 1 con el i.

4. Como caso extremo pero posible queda el caso con todo $s_{ii} = 0$. Permutando columnas y filas, siempre en operaciones dobles, podemos llegar a un caso con $s_{12} \neq 0$. La fila 1 de s se sustituye por la suma de la fila 1 y 2 y la fila 2 por la diferencia de la fila 1 y 2, haciéndose a continuación las correspondientes operaciones columna.

$$\begin{pmatrix} 0 & s_{12} & \cdots \\ s_{12} & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \to \begin{pmatrix} 0+s_{12} & 0-s_{12} & \cdots \\ s_{12}+0 & s_{12}-0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \Rightarrow \begin{pmatrix} 2s_{12} & 0 & \cdots \\ 0 & -2s_{12} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Queda $2s_{12} \neq 0$ en posición $(1,1)$ en la nueva matriz y se procede como en **1.**

Consideremos el ejemplo

$$\begin{pmatrix} 0 & 3 & 0 \\ 3 & 0 & 3 \\ 0 & 3 & 0 \end{pmatrix} \to \begin{pmatrix} 0+3 & 3+0 & 0+3 \\ 0-3 & 3-0 & 0-3 \\ 0 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 3 \\ -3 & 3 & -3 \\ 0 & 3 & 0 \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} 3+3 & 3-3 & 3 \\ -3+3 & -3-3 & -3 \\ 0+3 & 0-3 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 3 \\ 0 & -6 & -3 \\ 3 & -3 & 0 \end{pmatrix}.$$

La base $\mathcal{U} = (u_1, u_2, u_3)$ se transforma en $(u_1 + u_2, u_1 - u_2, u_3) = \mathcal{V}$ que cumple $\sigma(v_1, v_1) = 6 \neq 0$.

Problema 297 *Diagonalizar las funciones bilineales σ con $\mathbb{E}, \mathbb{R}^3$ con matrices*

$$\begin{pmatrix} 1 & 1 \\ 1 & h \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 2 & 3 & 4 \end{pmatrix}$$

y dar una base que las diagonalice. ♦

Solución. Para la primera matriz ponemos

$$\left(\begin{array}{cc} 1 & 1 \\ 1 & h \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \to \left(\begin{array}{cc} 1 & 1-1 \\ 1 & h-1 \\ \hline 1 & 0-1 \\ 0 & 1-0 \end{array}\right) = \left(\begin{array}{cc} 1 & 0 \\ 1 & h-1 \\ \hline 1 & -1 \\ 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{cc} 1 & 0 \\ 0 & h-1 \\ \hline 1 & -1 \\ 0 & 1 \end{array}\right).$$

La matriz final arriba es la forma diagonal y las columnas de la matriz final abajo la base que diagonaliza.

En el segundo caso, los cálculos son

$$\left(\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 0 & 0+0 & 0-0 \\ 0 & 0+1 & 0-1 \\ 0 & 1+0 & 1-0 \\ \hline 1 & 0+0 & 0-0 \\ 0 & 1+0 & 1-0 \\ 0 & 0+1 & 0-1 \end{array}\right) = \left(\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 1+1 & -1+1 \\ 0 & 1-1 & -1-1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{array}\right) = \left(\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{array}\right).$$

La matriz final arriba es la forma diagonal y las columnas de la matriz final abajo la base que diagonaliza.

En el tercer caso, pasamos primero las filas/columnas 2 y 3 al final

$$\left(\begin{array}{ccc} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 2 & 3 & 4 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 1 & 2 & 0 \\ -1 & 3 & 1 \\ 3 & 4 & 2 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} -1 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 0 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right).$$

Luego hacemos operaciones para que en la primera columna y fila sean todo ceros salvo el -1,

$$\left(\begin{array}{ccc} -1 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 0 \\ \hline 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right) \to \left(\begin{array}{ccc} -1 & 3+3\cdot(-1) & 1+(-1) \\ 3 & 4+3\cdot 3 & 2+3 \\ 1 & 2+3\cdot 1 & 0+1 \\ \hline 0 & 0+3\cdot 0 & 1+0 \\ 1 & 0+3\cdot 1 & 0+1 \\ 0 & 1+3\cdot 0 & 0+0 \end{array}\right) = \left(\begin{array}{ccc} -1 & 0 & 0 \\ 3 & 13 & 5 \\ 1 & 5 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 3 & 1 \\ 0 & 1 & 0 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 13 & 5 \\ 0 & 5 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 3 & 1 \\ 0 & 1 & 0 \end{array}\right).$$

Ahora quitamos el 5,

$$\left(\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 13 & 5 \\ 0 & 5 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 3 & 1 \\ 0 & 1 & 0 \end{array}\right) \to \left(\begin{array}{ccc} -1 & 0 & 0-\frac{5}{13}\cdot 0 \\ 0 & 13 & 5-\frac{5}{13}\cdot 13 \\ 0 & 5 & 1-\frac{5}{13}\cdot 5 \\ \hline 0 & 0 & 1-\frac{5}{13}\cdot 0 \\ 1 & 3 & 1-\frac{5}{13}\cdot 3 \\ 0 & 1 & 0-\frac{5}{13}\cdot 1 \end{array}\right) = \left(\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 5 & -\frac{12}{13} \\ \hline 0 & 0 & 1 \\ 1 & 3 & -\frac{2}{13} \\ 0 & 1 & -\frac{5}{13} \end{array}\right) \Rightarrow \left(\begin{array}{ccc} -1 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 0 & -\frac{12}{13} \\ \hline 0 & 0 & 1 \\ 1 & 3 & -\frac{2}{13} \\ 0 & 1 & -\frac{5}{13} \end{array}\right).$$

Las bases que se piden están dadas por las columnas de las últimas matrices abajo en el cálculo,

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 3 & -\frac{2}{13} \\ 0 & 1 & -\frac{5}{13} \end{pmatrix}. \blacklozenge$$

Lo que hacemos en general es, a partir de $s = \mathrm{mat}_{\mathcal{UU}}(\beta)$ hacer unas operaciones equivalentes a ir multiplicando por matrices κ_j,

$$\left(\frac{s}{I}\right) \overset{1}{\to} \left(\frac{s\kappa_1}{I\kappa_1}\right) \overset{1}{\Rightarrow} \left(\frac{\kappa_1^\top s\kappa_1}{I\kappa_1}\right) \overset{2}{\to} \left(\frac{\kappa_1^\top s\kappa_1\kappa_2}{I\kappa_1\kappa_2}\right) \overset{2}{\Rightarrow} \left(\frac{\kappa_2^\top \kappa_1^\top s\kappa_1\kappa_2}{I\kappa_1\kappa_2}\right) \overset{3}{\to} \dots$$

hasta llegar a

$$\dots \overset{h}{\to} \left(\frac{\kappa_{h-1}^\top \cdots \kappa_2^\top \kappa_1^\top s\kappa_1\kappa_2\cdots\kappa_{h-1}\kappa_h}{I\kappa_1\kappa_2\cdots\kappa_{h-1}\kappa_h}\right) \overset{h}{\Rightarrow} \left(\frac{\kappa_h^\top \kappa_{h-1}^\top \cdots \kappa_2^\top \kappa_1^\top s\kappa_1\kappa_2\cdots\kappa_{h-1}\kappa_h}{I\kappa_1\kappa_2\cdots\kappa_{h-1}\kappa_h}\right) = \left(\frac{p^\top sp}{p}\right) = \left(\frac{d}{p}\right).$$

Si $s = \mathrm{mat}_{\mathcal{UU}}(\sigma)$ y $p = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$ se tiene que

$$d = p^\top sp = \left[\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right]^\top \mathrm{mat}_{\mathcal{UU}}(\sigma)\, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) = \mathrm{mat}_{\mathcal{VV}}(\sigma);$$

es decir d es la matriz de σ en la base $\mathcal{V}$ relacionada con $\mathcal{U}$ por la matriz de cambio $p = \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$. Independientemente de que d sea o no diagonal, lo que importa es que si hacemos operaciones fila/columna correspondientes a matrices $\kappa_1, \dots, \kappa_h$ y $p = \kappa_1 \cdots \kappa_h$, al llegar al punto

$$\left(\frac{s}{I}\right) \overset{1}{\to} \overset{1}{\Rightarrow} \overset{2}{\to} \overset{2}{\Rightarrow} \dots \overset{h}{\to} \overset{h}{\Rightarrow} \left(\frac{p^\top sp}{p}\right) = \left(\frac{d}{p}\right) \text{ se tendrá } p^\top sp = d.$$

El ordenador comprueba esto en el tercer caso del problema precedente

$$p^\top sp = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 3 & -\frac{2}{13} \\ 0 & 1 & -\frac{5}{13} \end{pmatrix}^T \begin{pmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 3 & -\frac{2}{13} \\ 0 & 1 & -\frac{5}{13} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 0 & -\frac{12}{13} \end{pmatrix} = d.$$

Problema 298 *Dar la forma diagonal de las matrices, supuesto* $\Bbbk = \mathbb{C}$, *y en el tercer caso,* $1 - p^2 \neq 0$ *y* $pq = 1$,

$$a = \begin{pmatrix} i & i & 1 \\ i & 1 & i \\ 1 & i & q \end{pmatrix} \qquad b = \begin{pmatrix} i & 0 & q \\ 0 & p & 0 \\ q & 0 & q \end{pmatrix} \qquad c = \begin{pmatrix} 1 & p & 0 & q \\ p & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix}$$

y, en los dos primeros casos (el tercero es muy largo) una base que diagonalice.

Problema 299 *Al diagonalizar* σ *obtenemos* $(\lambda_1, \dots, \lambda_n)$ *en la diagonal de su matriz. Supuesto* $n \geq 2$ *preguntamos:*

1. *¿Existe si* $\lambda_m = 0$ *algún* $x \neq 0$ *tal que* $\sigma(x, x) = 0$*? ¿Y existe algún* $x \neq 0$ *tal que* $\sigma(x, y) = 0$ *para todo* y*?*
2. *Si todos los* λ_j *son no nulos, ¿cómo se responden las preguntas anteriores?*
3. *¿Cuando es el espacio nulo distinto de* 0*?*
4. *¿Están los* λ_i*, salvo orden de numeración, unívocamente determinados por* σ*?*

En $\Bbbk = \mathbb{C}$ o $\mathbb{R}$ se puede mejorar todavía el teorema de diagonalización.

Teorema 160 *Sea* $\Bbbk = \mathbb{C}$. *Si* σ *es simétrica de rango* r, *hay bases* $\mathcal{V}$ *tales que*

$$\sigma(v_i, v_j) = 0 \quad si\ i \neq j, \qquad \sigma(v_i, v_i) = 1 \quad si\ i \leq r, \qquad \sigma(v_i, v_i) = 1 \quad si\ i > r.$$

Toda base $\mathcal{U}$ *en donde* $\mathrm{mat}_{\mathcal{UU}}(\sigma)$ *sea diagonal con unos y ceros en ella tiene exactamente* r *unos.*

Demostración. Tomamos una base $\mathcal{W}$ como en el teorema 159. Sea h_i una de las raíces cuadradas de $\sigma(w_i, w_i)$ cuando sea $\sigma(w_i, w_i) \neq 0$ de modo que $(h_i)^2 = \sigma(w_i, w_i)$. (Recordamos que todo número complejo tiene raíz cuadrada.) Construimos una nueva base $\mathcal{V}$ dejando $v_j = w_j$ si $\sigma(w_j, w_j) = 0$ y cambiando $v_j = (1/h_j)\, w_j$ en caso contrario. La base $\mathcal{V}$ cumple el teorema una vez reordenada para que vayan primero todos los v_i con $\sigma(v_i, v_i) = 1$.

El número r no depende de la base construida, que no es única, sino solo de σ, pues es su rango. ♣

Por ejemplo, las matrices

$$\begin{pmatrix} i & i & i \\ i & i & i \\ i & i & 1+i \end{pmatrix} \text{ y } \begin{pmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 2 & 3 & 4 \end{pmatrix} \text{ se pueden diagonalizar a } \begin{pmatrix} i & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ y } \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -12 \end{pmatrix}$$

y el teorema previo nos dice que se puede mejorar el resultado obteniendo matrices diagonales

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Por supuesto, en la segunda matriz, aunque todos sus coeficientes estén en $\mathbb{Z}$, como $\mathbb{Z} \subset \mathbb{C}$, se la considera como matriz compleja. Si las bases (w_1, w_2, w_3) dan la forma diagonal sin ser 1 todos los términos, se ha de dar un último paso definiendo

$$(v_1, v_2, v_3) = \left(\left(\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right) w_1, w_2, w_3\right) \text{ o } (v_1, v_2, v_3) = \left(\frac{1}{i} w_1, w_2, \frac{1}{i\sqrt{12}} w_3\right)$$

porque

$$\sqrt{i} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i, \quad \frac{1}{\sqrt{i}} = \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i, \quad \sqrt{-1} = i, \quad \sqrt{-12} = i\sqrt{12}.$$

Si no nos piden la base, basta observar que las matrices de los ejemplos tiene rangos 2 y 3 y por ello sabemos cuántos unos y ceros van en la diagonal.

Nos planteamos un teorema análogo a este en el caso $\Bbbk = \mathbb{R}$.

Teorema 161 (de Sylvester) *Sea $\Bbbk = \mathbb{R}$ y σ forma simétrica de rango r. Existen p, q con $p + q = r$ y bases $\mathcal{V}$ tales que*

$$\sigma(v_i, v_j) = 0 \text{ si } i \neq j, \quad \sigma(v_i, v_i) = 0 \text{ si } i > r, \quad \sigma(v_i, v_i) = 1 \text{ si } i \leq p, \quad \sigma(v_j, v_j) = -1 \text{ si } p < j \leq p + q.$$

*Los números p, q, r dependen solo de σ, no de la base elegida (**ley de la inercia**).*[6]

Demostración. Tomamos una base $\mathcal{W}$ como en el teorema 159. Podemos suponerla ordenada de forma que los $\sigma(w_i, w_i) > 0$ sean para $1 \leq i \leq p$ y los $\sigma(w_i, w_i) < 0$ sean para $p + 1 \leq i \leq p + q$. Tiene que ser $p + q = r$, el rango de σ ya que, al ser la matriz diagonal, r es el rango de la matriz y de la forma. Solo queda retocar la base para que sea $\sigma(w_i, w_i) = \pm 1$ cuando $\sigma(w_i, w_i) \neq 0$. Definimos

$$v_i = \frac{1}{\sqrt{|\sigma(w_i, w_i)|}} w_i, \quad 1 \leq i \leq p + q = r, \qquad v_j = w_j, \quad j > r.$$

Esta base cumple las condiciones porque para $i \leq r = p + q$.

$$\sigma(v_i, v_i) = \frac{\sigma(w_i, w_i)}{|\sigma(w_i, w_i)|} = \operatorname{sg}(\sigma(w_i, w_i)) = \pm 1,$$

siendo $\operatorname{sg}(r)$ el signo de $r \in \mathbb{R}$, que se define como $-1, 0, +1$ según sea $r < 0$, $r = 0$ o $r > 0$.

Probamos la segunda parte. Supongamos que haya bases $\mathcal{V}_1$ y $\mathcal{V}_2$ donde $p_1 \neq p_2$ y $q_1 \neq q_2$. Como r es el rango, que no depende de la base, sí que tiene que ser $p_1 + q_1 = p_2 + q_2 = r$. Digamos por

[6] Leo en el texto *Linear Algebra* de Kaye y Wilson, capítulo 7, que el nombre no es por su relación con la inercia sino porque Sylvester dijo (en inglés): *If Isaac Newton can have a law of inertia, so can I.* Nada de esto dice la Wikipedia ni el MacTutor, pero en este última fuente hay una larga biografía donde se lee que golpeó, con un bastón de estoque (sword stick) y casi dejo por muerto a un alumno que leía el periódico en su clase (entonces no había móviles). El lector sacará sus consecuencias de cómo eran las clases antaño y el autor las sacará del hecho que la Facultad le ayudó muy poco y Sylvester se tuvo que marchar.

ejemplo que sea $p_1 > p_2$. Llamamos $\mathbb{F}$ al subespacio generado por $\left\{v_{(1)1}, \ldots, v_{(1)p_1}\right\}$ y $\mathbb{G}$ al generado por $\left\{v_{(2)p_2+1}, \ldots, v_{(2)n}\right\}$. Es muy fácil ver para $x \neq 0$ que si $x \in \mathbb{F}$ es $\sigma(x,x) > 0$ y que si $x \in \mathbb{G}$ es $\sigma(x,x) \leq 0$. Aplicamos la fórmula de Grassmann a $\mathbb{F}$ y $\mathbb{G}$ y, como $\dim(\mathbb{F}+\mathbb{G}) \leq n$,

$$\begin{aligned}\dim(\mathbb{F}\cap\mathbb{G}) &= \dim\mathbb{F}+\dim\mathbb{G}-\dim(\mathbb{F}+\mathbb{G}) \geq \dim\mathbb{F}+\dim\mathbb{G}-n \\ &\geq p_1+(n-p_2)-n = p_1-p_2 > 0.\end{aligned}$$

Hay pues un $x \neq 0$ en $\mathbb{F}\cap\mathbb{G}$ que cumple $\sigma(x,x) > 0$ y $\sigma(x,x) \leq 0$. Contradicción. ♣

En el caso $\Bbbk = \mathbb{R}$ vemos que los parámetros p, q y r determinan unívocamente la forma "óptima" de la matriz de σ. Hay dependencia entre p, q y r porque $p+q=r$. Todos los autores llaman r al rango de σ pero hay muchas elecciones sobre cuál es el segundo parámetro que se usará y qué nombre se le dará. Elegiremos llamar **signatura** de σ al par de números (p,q). Claramente, conocida la signatura se conoce el rango $r = p+q$. A diferencia del caso complejo donde el conocimiento de la forma diagonal con unos se tiene en cuanto se sabe el rango, aquí cuesta más.

Problema 300 *Hallar el rango y signatura de las matrices reales (los coeficientes no escritos son 0)*

$$a = \begin{pmatrix} 1 & 1 & & & & \\ 1 & 1 & & & & \\ & & 1 & -1 & & \\ & & -1 & 1 & & \\ & & & & -1 & -1 \\ & & & & -1 & -1 \end{pmatrix}, \qquad b = \begin{pmatrix} 0 & 1 & & & & \\ 1 & 0 & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & 0 & 1 \\ & & & & 1 & 0 \end{pmatrix}.$$

8.3. Formas cuadráticas

Dada σ forma simétrica en $\mathbb{E}$ podemos definir[7] la función $Q : \mathbb{E} \to \Bbbk$ por $Q(x) = \sigma(x,x)$. Entonces Q será la **forma cuadrática asociada a** σ y σ la **forma polar de** Q. El **rango y signatura** (esto último si $\Bbbk = \mathbb{R}$) de Q son los de σ por definición. Se puede obtener σ si se conoce Q porque

$$\sigma(x,y) = \frac{1}{2}\left[Q(x+y) - Q(x) - Q(y)\right] = \frac{1}{4}\left[Q(x+y) - Q(x-y)\right]$$

tal como se vio en el problema 293. Si se conoce $s = \mathrm{mat}_{\mathcal{UU}}(\sigma)$, sin estas fórmulas, se tiene

$$Q(x) = \sigma(x,x) = \sum_{i,j=1}^{n} s_{ij}x^i x^j \overset{*}{=} \sum_{h=1}^{n} s_{hh}\left(x^h\right)^2 + 2\sum_{i<j} s_{ij}x^i x^j, \tag{8.5}$$

habiéndose usado en $\overset{*}{=}$ que $s_{ij} = s_{ji}$. Esta segunda forma de expresar Q, aunque menos simétrica, nos será útil para cálculos. Los términos $s_{hh}\left(x^h\right)^2$ se llaman **términos cuadráticos.** Damos un ejemplo para $\mathbb{E} = \mathbb{R}^3$, la base estándar y

$$s = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 5 & 0 \\ -2 & 0 & -1 \end{pmatrix}.$$

Entonces, usando paréntesis en $\left(x^i\right)^2$ para evitar confusiones,

$$Q\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \left(x^1, x^2, x^3\right)\begin{pmatrix} 1 & 0 & -2 \\ 0 & 5 & 0 \\ -2 & 0 & -1 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = \left(x^1\right)^2 + 5\left(x^2\right)^2 - \left(x^3\right)^2 - 4x^1x^3.$$

Igual que con $\mathrm{mat}_{\mathcal{UU}}(\sigma) = s$ se conoce Q, también conocida Q como a la derecha en (8.5), se tiene s. Los coeficientes de la diagonal son los que multiplican a los términos cuadráticos, los s_{ij} con $i < j$ son *la mitad* de los coeficientes de $x^i x^j$, y se completa s con $s_{ij} = s_{ji}$. Un ejemplo:

$$\text{si } Q(x) = \left(x^1\right)^2 - 2\left(x^3\right)^2 + 2x^1x^2 + 6x^1x^3 - x^2x^3, \text{ entonces } \mathrm{mat}_{\mathcal{UU}}(\sigma) = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 0 & -\frac{1}{2} \\ 3 & -\frac{1}{2} & -2 \end{pmatrix}.$$

[7] Acordamos al tratar formas simétricas y ahora al tratar $Q(x) = \sigma(x,x)$ que $\Bbbk$ no tiene característica 2.

Nuestro objetivo es expresar Q en una base adecuada como suma de términos cuadráticos. De hecho esto ya se tiene por el teorema 159, pues en las coordenadas que da $\mathcal{W}$ se consigue. En efecto,

$$\text{si } \operatorname{mat}_{\mathcal{W}\mathcal{W}}(\sigma) = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}, \text{ entonces } Q(x) = d_1\left(x^1\right)^2 + \ldots + d_n\left(x^n\right)^2 = \sum_{i=1}^{n} d_i\left(x^i\right)^2.$$

Podemos incluso añadir que el número de $d_i \neq 0$ es el rango de σ. Además, si $\Bbbk = \mathbb{C}$, podremos escribir en coordenadas adecuadas $Q(x) = \left(x^1\right)^2 + \ldots + \left(x^r\right)^2$, siendo r el rango de σ; y si $\Bbbk = \mathbb{R}$, $Q(x) = \left(x^1\right)^2 + \ldots + \left(x^p\right)^2 - \left(x^{p+1}\right)^2 - \ldots - \left(x^{p+q}\right)^2$ con p y q unívocamente determinados por σ. Todos los teoremas sobre σ tienen una traducción sencilla para su Q asociada. Hay sin embargo una cuestión paralela pero con ciertas diferencias: la **reducción de Lagrange**. Es un método para pasar de una expresión polinomial de Q como (8.5) a otra puramente cuadrática $Q(x) = d_1\left(x^1\right)^2 + \ldots + d_n\left(x^n\right)^2$. En lugar de centrarse en cambiar la base "mala" $\mathcal{U}$, en donde Q tiene una expresión complicada, a otra base "buena" en el que Q la tenga "buena" (por la simplicidad de $Q(x) = d_1\left(x^1\right)^2 + \ldots + d_n\left(x^n\right)^2$), lo que busca la reducción de Lagrange es cambiar las coordenadas $\left(x^i\right)$ asociadas a $\mathcal{U}$ por otras $\left(y^j\right)$ en las que sea $Q(x) = d_1\left(y^1\right)^2 + \ldots + d_n\left(y^n\right)^2$. El foco se centra en cambiar *coordenadas* y no cambiar bases. En realidad, si asociamos "coordenadas" con "base dual" vemos que lo que se hace es cambiar de la base dual $\mathcal{U}^*$ a $\mathcal{W}^*$ en lugar de cambiar de $\mathcal{U}$ a $\mathcal{W}$.

De momento no entramos en problemas computacionales.

Problema 301 *Sea $Q : \mathbb{R}^2 \to \mathbb{R}$ una forma cuadrática. Se supone que existen vectores $p, q \in \mathbb{R}^2$ tales que $Q(p) > 0$ y $Q(q) < 0$. Probar con el teorema 161 que hay una base $\mathcal{W}$ en donde Q se escribe como $Q = \left(w^1\right)^2 - \left(w^2\right)^2$.*

Problema 302 *Sea $Q : \mathbb{E} \to \mathbb{C}$ una forma cuadrática en un espacio complejo con $\dim \mathbb{E} \geq 2$. Probar que siempre existe un $w \neq 0$ tal que $Q(w) = 0$. Indicación: estudiar la ecuación de segundo grado $Q(u + \lambda v) = 0$. ¿Por qué pedimos que sea $\dim \mathbb{E} \geq 2$?*

Lo que sigue es materia optativa. Tiene el interés de poner de relieve la importancia de la base dual, relacionada con las coordenadas, y dar la reducción de Lagrange (tema muy clásico) que puede servir como método alternativo a los vistos en la sección anterior para clasificar Q y σ. El esfuerzo a realizar para cálculos manuales es muy similar entre uno y otro método pero el de la sección *Formas bilineales simétricas* es mejor si se quiere conocer también la base que diagonaliza. El lector decidirá si quiere estudiar esta materia optativa.

8.3.1. La reducción de Lagrange

Cada vez que en $\mathbb{E}$ tenemos una base $\mathcal{U} = (u_1, \ldots, u_n)$, podemos asignar coordenadas x^i a x dadas por $x = \sum_{i=1}^{n} x^i u_i$. Dijimos al tratar el espacio dual que la función u^i que asigna a x su coordenada i (es decir, $x^i = u^i(x)$) es lineal. Las funciones $u^1, \ldots, u^n$ de $\mathbb{E}$ en $\Bbbk$ están en el espacio vectorial $\mathbb{E}^*$ y de hecho forman base, que al haber sido originada por $\mathcal{U}$ se denota por $\mathcal{U}^*$, y es la base dual de $\mathcal{U}$. Aunque resulte raro, la fórmula $\mathrm{id}_{\mathbb{E}} = \sum_{i=1}^{n} u^i u_i$ es perfectamente correcta porque $x = \mathrm{id}_{\mathbb{E}}(x) = \sum_{i=1}^{n} u^i(x)\, u_i$.

Si escribimos $Q(x) = \sum_{i,j=1}^{n} s_{ij} x^i x^j$ como en (8.5), tenemos como expresión alternativa "quitando la variable", $Q = \sum_{i,j=1}^{n} s_{ij} u^i u^j$, siendo s la matriz de σ en la base $\mathcal{U}$. Suele tenderse a escribir fórmulas "con variable", como por ejemplo "$2\cos(x)\sin(x) = \sin(2x)$", porque la alternativa es "$2\cos\cdot\sin = \sin\circ T$ donde $T : \mathbb{R} \to \mathbb{R}$ es $T(x) = 2x$", cosa mucho más dificil de manejar. Nosotros seguiremos utilizando $Q(x) = \sum_{i,j=1}^{n} s_{ij} x^i x^j$, pero para seguir algunas explicaciones tendrá el lector que sustituir mentalmente x^i por u^i o $u^i(x)$. Esta ampliación de notación (un "estiramiento") es muy frecuente en Matemáticas. Tiene riesgos pero se supone que el contexto evita confusiones y que la agilidad de manejo compensa esos riesgos.

Un ejemplo. Partimos de

$$Q = Q\left(x^1, x^2\right) = \left(x^1\right)^2 + h\left(x^2\right)^2 + 2x^1x^2$$

y hacemos el cambio ventajoso de coordenadas $x^1 = y^1 - y^2$, $x^2 = y^2$ (no decimos de donde viene la inspiración) que da

$$\left(x^1\right)^2 + h\left(x^2\right)^2 + 2x^1x^2 = \left(y^1 - y^2\right)^2 + h\left(y^2\right)^2 + 2\left(y^1 - y^2\right)y^2 = \left(y^1\right)^2 + (h-1)\left(y^2\right)^2.$$

Estamos cambiando la base $\left(u^1,u^2\right)$ asociada a $\left(x^1,x^2\right)$ por la base $\left(v^1,v^2\right)$ asociada a $\left(y^1,y^2\right)$ y por esto $x^1=y^1-y^2$, $x^2=y^2$ equivale a $u^1=v^1-v^2$, $u^2=v^2$. Si Q es en $\mathcal{V}^*$ suma de términos cuadráticos es porque σ se diagonaliza con $\mathcal{V}$. Por supuesto, si se conoce la relación entre $\mathcal{U}^*$ y $\mathcal{V}^*$, que es como conocer la relación entre coordenadas, se conoce asimismo la relación entre $\mathcal{U}$ y $\mathcal{V}$. Efectivamente, para $c\in\Bbbk^{n\times n}$ equivalen $c=\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}\left(\mathrm{id}_{\mathbb{E}}\right)$ y $\mathrm{mat}^{\mathcal{U}}\left(x\right)=c\,\mathrm{mat}^{\mathcal{V}}\left(x\right)$, luego si sabemos la expresión da las coordenadas antiguas en función de las nuevas, conocemos, *con la misma matriz*, la expresión de la base nueva en función de la base antigua. En el ejemplo de más arriba,

$$\begin{pmatrix} x^1\\ x^2\end{pmatrix}=\begin{pmatrix} 1 & -1\\ 0 & 1\end{pmatrix}\begin{pmatrix} y^1\\ y^2\end{pmatrix}\text{ luego }\left(v_1,v_2\right)=\left(u_1,u_2\right)\begin{pmatrix} 1 & -1\\ 0 & 1\end{pmatrix}\text{ que es }\begin{cases} v_1=u_1\\ v_2=-u_1+u_2\end{cases}.$$

La reducción de Lagrange es un algoritmo y hay dos herramientas que en muchos casos no se utilizan, pero hay que conocer porque si no se bloquea el procedimiento.

1. La primera es permutar las coordenadas, que equivale a permutar las bases. Si el lector quiere llevar también las cuentas del cambio de base[8] le decimos que si P es una permutación de $\{1,\ldots,n\}$, el que sea $x^j=y^{P(j)}$, alternativamente $u^j=v^{P(j)}$, equivale a que $v_{P(j)}=u_j$, $j=1,\ldots,n$. (¡Ojo! fijarse donde va la P). Es un razonamiento corto pero enrevesado. Si suponemos cierto $v_{P(j)}=u_j$ para todo j calculamos

$$\left(u^i-v^{P(i)}\right)\left(u_j\right)=u^i\left(u_j\right)-v^{P(i)}\left(u_j\right)=\delta^i_j-v^{P(i)}\left(v_{P(j)}\right)=\delta^i_j-\delta^{P(i)}_{P(j)}=\delta^i_j-\delta^i_j=0$$

   y, al anularse $u^i-v^{P(i)}$ sobre una base de $\mathbb{E}$, debe ser $u^i-v^{P(i)}=0$. Recíprocamente, si $u^j=v^{P(j)}$ para todo j calculamos

$$u^j\left(v_{P(i)}-u_i\right)=v^{P(j)}\left(v_{P(i)}\right)-u^j\left(u_i\right)=\delta^{P(j)}_{P(i)}-\delta^j_i=0,$$

   luego $v_{P(i)}-u_i$ tiene todas las coordenadas nulas y ha de ser $v_{P(i)}-u_i$.

2. La segunda herramienta se usa si en la expresión (8.5) todo $s_{ii}=0$ porque esto impide completar el cuadrado que es lo principal de la reducción de Lagrange. Supongamos pues que

$$Q\left(x\right)=2\sum_{1\le i<j\le n}s_{ij}x^ix^j=2s_{12}x^1x^2+2\sum_{j=3}^{n}s_{1j}x^1x^j+2\sum_{j=3}^{n}s_{2j}x^2x^j+2\sum_{3\le i<j\le n}s_{ij}x^ix^j.$$

   El caso $Q=0$ es trivial y sin interés, así que supondremos $Q\neq 0$. Permutando si se necesita las coordenadas con **1**, podemos suponer que $s_{12}\neq 0$. Hacemos $x^1=y^1+y^2$ y $x^2=y^1-y^2$ dejando $x^j=y^j$ para $j\ge 3$. Obsérvese que $x^1x^2=\left(y^1\right)^2-\left(y^2\right)^2$ y

$$Q\left(y\right)=2s_{12}\left(\left(y^1\right)^2-\left(y^2\right)^2\right)+\sum_{j=3}^{n}s_{1j}\left(y^1+y^2\right)y^j+\sum_{j=3}^{n}s_{2j}\left(y^1-y^2\right)y^j+2\sum_{3\le i<j\le n}s_{ij}y^iy^j.$$

   En la nueva $Q\left(y\right)$ los coeficientes de $\left(y^1\right)^2$ y $\left(y^2\right)^2$ son $2s_{12}$ y $-2s_{12}$, ambos no nulos.

Como ejemplo damos, con letras diferentes en vez de superíndices,[9]

$$Q\left(x,y,x\right)=xy+xz+yz=\left(u+v\right)\left(u-v\right)+\left(u+v\right)z+\left(u-v\right)z=u^2+2uz-v^2=Q\left(u,v,z\right).$$

Entramos ahora propiamente en la **reducción de Lagrange** cuyo paso esencial es la llamada **completación del cuadrado.** Elija el lector si prefiere mirar primero un ejemplo y luego la teoría general o al revés. Tomamos $Q\left(x,y,x\right)=xy+xz+yz$. Para tener un término cuadrático hacemos el cambio de más arriba y $Q\left(u,v,z\right)=u^2+2uz-v^2$. Entonces con $\pm z^2$ completamos un cuadrado y

$$\begin{aligned}Q\left(x,y,x\right)=Q\left(u,v,z\right)&=u^2+2zu-v^2=u^2+2zu+z^2-z^2-v^2\\&=\left(u+z\right)^2-z^2-v^2=w^2-z^2-v^2=Q\left(w,v,z\right).\end{aligned}$$

[8] Ya le hemos advertido que, en nuestra opinión, es mejor el procedimiento con matrices.

[9] Esto es conveniente para cálculos con $n\le 4$ así como buscar solo un nuevo nombre para la coordenada que se modifica sin cambiar el de las demás.

siendo $w = u + z$. Si queremos llevar la cuenta de las coordenadas, que es muy pesado casi siempre, conviene hacerlo desde las primeras a las últimas. El cálculo es

$$\left\{\begin{array}{l} x = u + v = (w - z) + v \\ y = u - v = (w - z) - v \\ z = z \end{array}\right. , \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} w \\ v \\ z \end{pmatrix}$$

Si llamamos M a la matriz, ella expresa las coordenadas antiguas respecto de las nuevas, luego expresa también la base nueva respecto de la antigua. Por consiguiente, las columnas de M forman una base que diagonaliza Q como prueba

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix}^{\top} \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

El razonamiento general es este. Tras hacer modificaciones de las coordenadas si fuese necesario, podemos suponer que tenemos unas coordenadas x en donde $s_{11} \neq 0$. Se divide la expresión de Q en tres partes

$$Q = s_{11} \left(x^1\right)^2 + 2 \sum_{i=2}^{n} s_{1j} x^1 x^j + \sum_{j,k=2}^{n} s_{jk} x^j x^k.$$

y se opera buscando que aparezca el cuadrado de un binomio

$$\begin{aligned} Q &= s_{11} \left[\left(x^1\right)^2 + 2x^1 \sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j \right] + \sum_{j,k=2}^{n} s_{jk} x^j x^k \\ &= s_{11} \left[\left(x^1\right)^2 + 2x^1 \sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j + \left(\sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j \right)^2 \right] - \left(\sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j \right)^2 + \sum_{j,k=2}^{n} s_{jk} x^j x^k \\ &= s_{11} \left[x^1 + \sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j \right]^2 - \left(\sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} x^j \right)^2 + \sum_{j,k=2}^{n} s_{jk} x^j x^k. \end{aligned}$$

Las nuevas coordenadas y la nueva expresión de Q son[10]

$$y^1 = x^1 + \sum_{i=2}^{n} \frac{s_{1i}}{s_{11}} x^i, \quad y^j = x^j \text{ para } j \geq 2,$$

$$Q = s_{11} \left(y^1\right)^2 - \left(\sum_{i=2}^{n} \frac{s_{1j}}{s_{11}} y^j \right)^2 + \sum_{j,k=2}^{n} s_{jk} y^j y^k = s_{11} \left(y^1\right)^2 + R\left(y^2, \ldots, y^n\right) = s_{11} \left(y^1\right)^2 + R\left(x^2, \ldots, x^n\right)$$

Aquí ya se tiene un procedimiento inductivo que permite aplicar a $R\left(y^2, \ldots, y^n\right) = R\left(x^2, \ldots, x^n\right)$ operaciones similares hasta llegar a tener solo términos cuadráticos.

Se puede decir que si las coordenadas originales son $\left(u^1, \ldots, u^n\right)$ y las finales son $\left(v^1, \ldots, v^n\right)$ y la relación entre ellas en forma matricial es

$$\begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix} = \begin{pmatrix} c_1^1 & \cdots & c_n^1 \\ \vdots & \ddots & \vdots \\ c_1^n & \cdots & c_n^n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix},$$

entonces $\operatorname{mat}^{\mathcal{U}}(x) = c \operatorname{mat}^{\mathcal{V}}(x)$, luego $c = \operatorname{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$. Con otras palabras, las columnas de c dan en coordenadas la base $\mathcal{V}$ que diagonaliza σ en función de la base original $\mathcal{U}$.

En el ejemplo de más arriba vemos primero que a $Q(x, y, x)$ no se le puede aplicar la completación del cuadrado, pero tras pasar de (x, y, z) a (u, v, z) sí que se puede hacer con $Q(u, v, z) = u^2 + 2uz - v^2$. Completando el cuadrado llegamos a de $p^2 + qp$, que requiere que aparezcan denominadores, llegamos a $w^2 - z^2 - v^2 = Q(w, v, z)$. La reducción de Lagrange puede ser quizás mejor si se busca principalmente la signatura en el caso $\Bbbk = \mathbb{R}$, que reduce a contar signos. En todo caso, son problemas pesados.

[10] Reiteramos que en la práctica no se suele cambiar en el cálculo $x^2, \ldots, x^n$ por $y^2, \ldots, y^n$.

Problema 303 *Diagonalizar en función del parámetro $h \in \mathbb{R}$ la forma Q*

$$Q(x,y,z) = 2x^2 + 4xy + 6xz + 2hyz,$$

y calcular su signatura. ♦

Solución. Calculamos por una parte

$$\begin{aligned} 2x^2 + 4xy + 6xz &= 2\left(x^2 + 2xy + 3xz\right) \\ &= 2\left(x^2 + 2x\left(y + \frac{3}{2}z\right)\right) \\ &= 2\left(x^2 + 2x\left(y + \frac{3}{2}z\right) + \left(y + \frac{3}{2}z\right)^2\right) - 2\left(y + \frac{3}{2}z\right)^2 \\ &= 2\left(x + y + \frac{3}{2}z\right)^2 - 2y^2 - 6yz - \frac{9}{2}z^2 \end{aligned}$$

y, con el cambio $u = x + y + \frac{3}{2}z$, queda

$$Q(x,y,z) = 2u^2 - 2y^2 + (2h-6)yz - \frac{9}{2}z^2 = 2u^2 - 2\left(y^2 + (3-h)yz\right) - \frac{9}{2}z^2 = Q(u,y,z).$$

Seguimos calculando

$$y^2 + (3-h)yz = y^2 + 2\frac{3-h}{2}yz + \left(\frac{3-h}{2}z\right)^2 - \left(\frac{3-h}{2}z\right)^2 = \left[y + \frac{3-h}{2}z\right]^2 - \left(\frac{3-h}{2}z\right)^2$$

Así pues, si $v = y + \frac{3-h}{2}z$ obtenemos

$$Q(x,y,z) = Q(u,y,z) = 2u^2 - 2v^2 + 2\left(\frac{3-h}{2}z\right)^2 - \frac{9}{2}z^2 = 2u^2 - 2v^2 + \frac{h(h-6)}{2}z^2 = Q(u,v,z).$$

Vamos con la signatura. El grafo de $h^2 - 6h$ es una parábola que corta a OX en $h = 0, 6$. En el intervalo $(0,6)$ tendrá $h^2 - 6h$ signo opuesto al que tenga fuera de $[0,6]$. Tomando $h = 1$ vemos que $h^2 - 6h$ es < 0 en $(1,6)$ y > 0 en $(1,6)$. Como consecuencia **(a)** la signatura es $(1,1)$ si $h = 0, 6$; **(b)** la signatura es $(2,1)$ si $h \notin (0,6)$; **(c)** la signatura es $(1,2)$ si $h \in [0,6]$.

Aunque el problema no lo pide, vamos a determinar la base que diagonaliza, siquiera para ver lo laborioso que es. Tenemos

$$\begin{cases} x = u - y - \frac{3}{2}z = u - v - \frac{h}{2}z \\ y = v - \frac{3-h}{2}z \\ z = z \end{cases} \quad \text{equivalente a} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -1 & -\frac{h}{2} \\ 0 & 1 & -\frac{3-h}{2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ z \end{pmatrix}.$$

Tal como hemos dicho en general, si $\left(u^1, u^2, u^3\right) = (x,y,z)$, las coordenadas originales, y $\left(v^1, v^2, v^3\right) = (u,v,z)$ las coordenadas nuevas que diagonalizan, la matriz c de la ecuación anterior es $\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})$. Por consiguiente, las tres columnas de c son los tres vectores de $\mathcal{V}$ expresados en términos de $\mathcal{U}$. Efectivamente,

$$\begin{pmatrix} 1 & -1 & -\frac{h}{2} \\ 0 & 1 & -\frac{3-h}{2} \\ 0 & 0 & 1 \end{pmatrix}^T \begin{pmatrix} 2 & 2 & 3 \\ 2 & 0 & h \\ 3 & h & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & -\frac{h}{2} \\ 0 & 1 & -\frac{3-h}{2} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & \frac{h(h-6)}{2} \end{pmatrix},$$

que es $\left[\mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id})\right]^\top \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\sigma)\, \mathrm{mat}_{\mathcal{V}}^{\mathcal{U}}(\mathrm{id}) = \mathrm{mat}_{\mathcal{V}\mathcal{V}}(\beta)$. ♦

Problema 304 *Aplicar la reducción de Lagrange a*

$$Q_1(x,y,z) = xy + 2xz \quad y \quad Q_2(x,y,z) = x^2 + y^2 + z^2 + 2xy + 2yz.$$

dando su rango y signatura.

El siguiente problema pregunta sobre σ pero quizás sea mejor trabajar con Q asociada a σ.

Problema 305 *Nos dan una forma σ cuya matriz a en una base $\mathcal{U}$ tiene todo ceros excepto la primera fila y columna todo con unos. Hallar su forma diagonal. El problema vale para cualquier $\Bbbk$ pero supongamos $\Bbbk = \mathbb{R}$. Dar la signatura. Si es $\Bbbk = \mathbb{C}$. dar coordenadas con las que Q asociada a σ sea suma de cuadrados. Nota. Si resulta difícil hacer el caso con $m = 4$.*

En otras cuestiones hemos dicho cómo se pueden poner cuantos problemas se quieran de tipo numérico conociendo de antemano la solución. Aquí también puede hacerse. Sea $d \in \mathbb{R}^{n\times n}$ una matriz diagonal con p veces 1, q veces -1 y el resto 0. Sin duda c tiene signatura (p,q). Si c es una matriz invertible, $b = c^\top dc$ tiene también signatura (p,q).

Problema 306 *Justificar la afirmación precedente.*

Otro procedimiento similar cuando $\Bbbk = \mathbb{R}$ consiste en elegir formas lineales $f^1, \ldots, f^r$ *independientes* con $r \leq n$ y $p, q \in \mathbb{N}$ tales que $p+q = r$. Se define $\sigma = f^1 \otimes f^1 + \ldots + f^p \otimes f^p - f^{p+1} \otimes f^{p+1} - \ldots - f^r \otimes f^r$; o sea,

$$\sigma(x,y) = f^1(x) f^1(y) + \ldots + f^p(x) f^p(y) - f^{p+1}(x) f^{p+1}(x) - \ldots - f^r(x) f^r(x).$$

Problema 307 *Definimos $\varepsilon_1 = 1, \ldots, \varepsilon_p = 1$, $\varepsilon_{p+1} = -1, , \ldots, \varepsilon_r = -1$. Probar que existe una base $\mathcal{V} = (v_1, \ldots, v_n)$ tal que $f^i(v_j) = \varepsilon_i$ si $i = j \leq r$ y $f^i(v_j) = 0$ en los demás casos. Probar también que $\mathcal{V}$ diagonaliza σ cuya signatura será (p,q).*

8.4. Una aplicación al Análisis

Es muy probable que el lector conozca cómo se estudian los **máximos**, **mínimos** y **puntos estacionarios** de una función $f : I \to \mathbb{R}$, siendo I un intervalo abierto. Se dice que f tiene un máximo (mínimo) en $z \in I$ si existe $\varepsilon > 0$ tal que para todo x de I tal que $|x - z| < \varepsilon$ se tenga $f(x) \leq f(z)$ ($f(x) \geq f(z)$). Si se tiene la hipótesis de que se pueden calcular las dos primeras derivadas de f, se prueba que es necesario, aunque no suficiente, para que z sea máximo o mínimo que se tenga $f'(z) = 0$. Geométricamente es como decir que la tangente al grafo en $(z, f(z))$ es una recta horizontal. Se dice que z es estacionario si $f'(z) = 0$, luego los máximos y mínimos son estacionarios. La razón por la que un punto estacionario puede no ser ni máximo ni mínimo es fácil de visualizar. Pasa que, aunque la tangente sea horizontal, puede suceder, por ejemplo, que al moverse a la derecha de z suba el grafo y baje si nos movemos hacia la izquierda. El lector sabrá que si $f''(z) < 0$ o $f''(z) > 0$, el punto estacionario es respectivamente un máximo o un mínimo. Aunque no pretendemos desarrollar la teoría formalmente sí que conviene tener una idea de cómo se hace esto. La fórmula de Taylor nos permite aproximar f por la función polinómica de segundo grado

$$P(x) = f(z) + f'(z)(x-z) + \tfrac{1}{2} f''(z)(x-z)^2 \quad \text{que es} \quad P(x) = f(z) + \tfrac{1}{2} f''(z)(x-z)^2 \text{ si } f'(z) = 0.$$

Aunque puede ser $f(x) \neq P(x)$, la aproximación es suficientemente buena para que $f(x) - f(z)$ y $P(x) - P(z)$ *tengan el mismo signo* (¡es la clave!) ¿Qué significa que z sea un mínimo? pues que $f(x) - f(z)$ es ≥ 0 para x cerca de z. Ahora tenemos que, como $f(z) = P(z)$, se cumple que el signo de $f(x) - f(z)$ es el de $P(x) - P(z)$, que es el de $f''(z)(x-z)^2$ (porque $P(z) = f(z)$ y $f'(z) = 0$) y, al haber un cuadrado, es el signo de $f''(z)$. Esta es la base de la demostración de porqué $f'(z) = 0$ y $f'(z) > 0$ permite saber que z es un mínimo.

La razón de este contacto con el Análisis es porque si f depende de varias variables hay entonces una teoría similar donde aparecen las formas lineales y las formas cuadráticas, y el teorema 161 *(de Sylvester) juega un papel esencial.*

Supondremos para simplificar $f : \mathbb{R}^n \to \mathbb{R}$. Si $z \in \mathbb{R}^n$ se puede definir en muchos casos para f lo análogo a $f'(z)$, la derivada primera en z. La cuestión es que $f'(z)$ no va a ser un número, sino una función lineal $f'(z) : \mathbb{R}^n \to \mathbb{R}$. Es aún más complicado generalizar $f''(z)$, que tampoco es un número sino una función $f''(z) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ que es *bilineal simétrica*. No entramos ahora en cómo se calculan $f'(z)$ y $f''(z)$ pero sí diremos que los puntos estacionarios se definen por la condición $f'(z) = 0$ y si visualizamos el grafo de f (cosa posible si $\mathbb{R}^n = \mathbb{R}^2$), el ser estacionario significa que el plano tangente

al grafo en $(z, f(z)) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^3$ es un plano horizontal. No pretendemos que el lector sepa calcular $f'(z)$ y $f''(z)$ sino que basta que admita que se pueden escribir en coordenadas de modo respectivo

$$f'(z)(x-z) = (f_1(z), \ldots, f_n(z)) \begin{pmatrix} x^1 - z^1 \\ \vdots \\ x^n - z^n \end{pmatrix},$$

$$f''(z)(x-z, x-z) = \left(x^1 - z^1, \ldots, x^{n1} - z^n\right) \begin{pmatrix} f_{11}(z) & \cdots & f_{11}(z) \\ \vdots & \ddots & \vdots \\ f_{11}(z) & \cdots & f_{11}(z) \end{pmatrix} \begin{pmatrix} x^1 - z^1 \\ \vdots \\ x^n - z^n \end{pmatrix}.$$

No hay que asustarse con tanto símbolo pues solo queremos decir que $f'(z)$, como toda función lineal de $\mathbb{R}^n \to \mathbb{R}$, tendrá una matriz $1 \times n$ cuyos coeficientes son los $f_j(z)$ dependientes de z. Análogamente, $f''(z)$ es bilineal y, como $(x-z)$ va repetido, $f''(z)(x-z, x-z)$ se conoce con la forma cuadrática $Q_z(v) = f''(z)(v, v)$ cuya matriz tiene coeficientes $f_{ij}(z)$. Si z es estacionario, $f'(z) = 0$ y lo análogo al polinomio $P(x)$ con una variable es otro de n variables $x^1, \ldots, x^n$,

$$P(x) = f(z) + f'(z)(x-z) + f''(z)(x-z, x-z) = f(z) + f''(z)(x-z, x-z) = f(z) + Q_z(x-z).$$

Sucede, aunque cuesta mucho más probarlo, que se mantiene la situación en una variable: $f(x) - f(z)$ y $P(x) - P(z)$, aun siendo distintas, son muy aproximadas si x está cerca de z y la aproximación es suficientemente buena para que $f(x) - f(z)$ y $P(x) - P(z)$ tengan el mismo signo. El tener, por ejemplo, $f(x) - f(z) \geq 0$ para x cerca de z (condición de mínimo) equivale a $P(x) - P(z) \geq 0$ para x cerca de z. *Y aquí está lo notable:* $P(x) - P(z) = Q_z(x-z)$ *ya que* $P(z) = f(z)$ *y nuestro problema se reduce a informarnos del signo de* $Q_z(x-z)$. ¡Pero con el teorema de Sylvester esto es posible! Si la signatura es $(n, 0)$, en coordenadas adecuadas $Q_z(x-z)$ es una suma de cuadrados, luego es ≥ 0 y en z hay un mínimo. Si la signatura fuera $(0, n)$ se tendría que $Q_z(x-z)$ es suma de cuadrados con $-$ delante y habría un máximo. ¿Qué podemos decir de las situaciones intermedias con signatura (s, t) y $s, t \geq 1$. Pues que moviéndonos en ciertas direcciones subimos y en otras bajamos y z es entonces un **punto de silla**.[11] Dicho con símbolos, consideramos un x concreto cerca de z y la curva $C(t) = f(z + t(x-z))$ que cumple $C(0) = z$. Comparamos

$$C(t) + C(0) = f(z + t(x-z)) - f(z) \quad \text{y} \quad P(z + t(x-z)) - P(z) = Q_z(z + t(x-z))$$

Según sea $z + t(x-z)$ se tendrá que $Q_z(z + t(x-z)) > 0$ o $Q_z(z + t(x-z)) < 0$, que es lo característico de un punto de silla.

Para calcular $f'(z)$ y $f''(z)$ hay que saber derivar. Vamos a dar, sin ninguna justificación cómo se hace cuando $f : \mathbb{R}^2 \to \mathbb{R}$ es un *polinomio* en las variables x, y, digamos que $f(x, y) = x^2y^2 - x^2 - y^2$ (los superíndices son exponentes). Las derivadas parciales respecto a x e y se obtienen fijando x (o bien y), dejando correr la otra variable, y derivando en el punto correspondiente como si la variable fijada fuese una constante. En nuestro ejemplo,

$$\frac{\partial f(x,y)}{\partial x} = \frac{d}{dx}\left(x^2y^2 - x^2 - y^2\right) = 2xy^2 - 2x, \quad \frac{\partial f(x,y)}{\partial y} = \frac{d}{dy}\left(x^2y^2 - x^2 - y^2\right) = 2yx^2 - 2y.$$

Para $z = (x, y)$ se tiene

$$f'(z) = f'(x, y) = \left(2xy^2 - 2x, 2yx^2 - 2y\right) = \left(2x\left(y^2 - 1\right), 2y\left(x^2 - 1\right)\right)$$

Los puntos estacionarios se obtienen resolviendo

$$\begin{cases} 2x\left(y^2 - 1\right) = 0 \\ 2y\left(x^2 - 1\right) = 0 \end{cases} \quad \text{que dan } (0,0), (1,1), (1,-1), (-1,1), (-1,-1)$$

[11] En francés *point de selle* y en inglés *saddle point*. En ambos casos se refiere a la silla de montar a caballo. que tiene una concavidad/convexidad pronunciada. Como ahora se monta menos a caballo el autor ofrece otra descripción. Imaginamos un collado con una montaña al Norte y otra al Sur, un valle a la derecha Este y otro a la izquierda Oeste de la línea de cumbres. Sea z el centro del collado. Si se va desde la cima Sur a la Norte a través de z, primero se baja y luego se sube; si se va desde el valle Este al valle Oeste a través de z, primero se sube y luego se baja. Al ir entre cimas, en z se alcanza la altura mínima, y al ir entre valles la altura máxima.

Para conocer $f''(z)$, que tendrá una matriz 2×2 hay que calcular las segundas derivadas parciales, que son

$$\frac{\partial^2 f(x,y)}{\partial x\partial x}=\frac{d}{dx}\left(\frac{\partial f(x,y)}{\partial x}\right)=2y^2-2,\quad \frac{\partial^2 f(x,y)}{\partial x\partial y}=\frac{d}{dx}\left(\frac{\partial f(x,y)}{\partial y}\right)=4xy$$

$$\frac{\partial^2 f(x,y)}{\partial y\partial x}=\frac{d}{dy}\left(\frac{\partial f(x,y)}{\partial x}\right)=4xy,\quad \frac{\partial^2 f(x,y)}{\partial y\partial y}=2x^2-2.$$

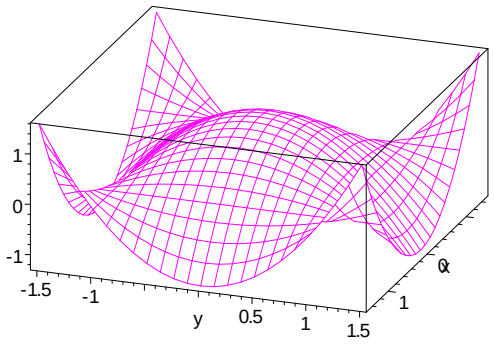
En cualquier z la derivada $f''(z)$ tiene matriz

$$f''(z)=\begin{pmatrix}2y^2-2 & 4xy\\ 4xy & 2x^2-2\end{pmatrix}$$

En los cinco puntos estacionarios las matrices son, haciendo $x,y=0,\pm 1$ según corresponda,

$$f''(0,0)=\begin{pmatrix}-2 & 0\\ 0 & -2\end{pmatrix},\ f''(1,1)=f''(-1,-1)=\begin{pmatrix}0 & 4\\ 4 & 0\end{pmatrix},\ f''(1,-1)=f''(-1,1)=\begin{pmatrix}0 & -4\\ -4 & 0\end{pmatrix}.$$

En el origen $(0,0)$ la matriz es definida negativa y ahí tenemos un máximo y en los otros casos La signatura es $(1,1)$ y hay puntos de silla. Damos un dibujo como ayuda



Como casos de máxima sencillez están $f(x,y)=xy$ y $f(x,y)=x^2+y^2$ donde el lector comprobará enseguida que $z=(0,0)$ es el único punto estacionario, siendo punto de silla para $f(x,y)=xy$ y mínimo para $f(x,y)=x^2+y^2$. Obviamente, si se saben calcular las derivadas para funciones más complejas o de más variables, se pueden hacer problemas más difíciles.

8.5. Cuádricas afines

$\mathbb{E}$ será un espacio con $\Bbbk=\mathbb{R}$ de dimensión n, $\sigma:\mathbb{E}\times\mathbb{E}\to\mathbb{R}$ una forma bilineal simétrica *no nula* con $Q(x)=\sigma(x,x)$ su cuádrica asociada, $f:\mathbb{E}\to\mathbb{R}$ una forma lineal, y $\alpha\in\mathbb{R}$ una constante. Con esto consideramos una función

$$\phi:\mathbb{E}\to\mathbb{R},\quad \phi(x)=\sigma(x,x)+2f(x)+\alpha.$$

El caso más importante en la práctica se obtiene tomando $s\in\mathbb{R}^{n\times n}$ simétrica no nula y $F\in\mathbb{R}^{1\times n}$ una matriz fila. Entonces

$$\phi:\mathbb{R}^n\to\mathbb{R},\quad \phi(x)=x^\top sx+2Fx+\alpha=\sum_{i,j=1}^{n}s_{ij}x^ix^j+2\sum_{k=1}^{n}F_kx^k+\alpha$$

y ϕ es una **función polinomial de segundo grado**. Se define una **cuádrica** $\mathcal{C}$ como un conjunto $\mathcal{C}=\{x\in\mathbb{E}\,|\phi(x)=0\}$, siendo ϕ como acabamos de describir. Hacemos varias advertencias. La primera es que puede ser $\mathcal{C}=\emptyset$ o un subespacio afín estricto de $\mathbb{E}$. En muchos casos vamos a descartar estos casos de la definición de cuádrica. La segunda es que una cosa es el subconjunto $\mathcal{C}$ y otra la ecuación $\phi(x)=0$ que lo define. Es obvio que si $\lambda\neq 0$, las ecuaciones $\phi(x)=0$ y $\lambda\phi(x)=0$ son distintas pero dan la misma cuádrica $\mathcal{C}$. Cuesta mucho más probar y nosotros lo admitiremos, que si $\mathcal{C}$ no es el vacío o un subespacio afín y tiene dos ecuaciones $\phi_1(x)=0$ y $\phi_2(x)=0$, existe un $\lambda\neq 0$ tal que $s_1=\lambda s_2$, $f_1=\lambda f_2$ y $\alpha_1=\lambda\alpha_2$. Por consiguiente, salvo en casos extremos, descartables para los teoremas interesantes, $\mathcal{C}$

determina la terna (σ, f, α) salvo factor multiplicativo. Esto será importante porque podremos decir, por ejemplo, que σ es no degenerada si lo es una de las posibles σ y entonces lo serán todas.

Si $\mathcal{R} = (z, u_1, \ldots, u_n)$ es una referencia afín recordamos que las coordenadas de x en $\mathcal{R}$ son $\left(x^1, \ldots, x^n\right)$ definidas por $x - z = \sum_{i=1}^{n} x^i u_i$. La **ecuación de $\mathcal{C}$ en la referencia** $\mathcal{R}$ es $\phi_{\mathcal{R}}\left(x^1, \ldots, x^n\right) = 0$ siendo $\phi_{\mathcal{R}}\left(x^1, \ldots, x^n\right) = \phi(x)$. Veremos que existen referencias $\mathcal{R}$ donde la ecuación es sencilla, informativa y permite clasificar las cuádricas.

8.5.1. Centros de una cuádrica

El concepto de **centro de la cuádrica** será esencial. Dado $z \in \mathbb{E}$ tenemos la función afín $\Sigma_z(x) = 2z - x$ que es la simetría respecto de z. *No es una función lineal.* Sin embargo tiene una sencilla interpretación visual. Pensemos en dos puntos x y x' simétricos con respecto a z. Esto es como decir que $x - z = -(x' - z)$ y por tanto $x' = 2z - x$. El caso $z = 0$ da $\Sigma_0(x) = -x$ y todo es muy intuitivo. Diremos que z es un **centro de** $\mathcal{C}$, supuesta $\mathcal{C}$ no vacía, si para cada $x \in \mathcal{C}$ se tiene que también $\Sigma_z(x) \in \mathcal{C}$. Si tenemos como caso especial $\phi(x) = \sigma(x, x) + \alpha$, luego $f = 0$, vemos que $z = 0$ es un centro de $\mathcal{C}$ ya que $\phi(-x) = \sigma(-x, -x) + \alpha = \sigma(x, x) + \alpha = \phi(x)$, que es 0 cuando $x \in \mathcal{C}$. Una cuádrica puede tener muchos centros y también no tener ninguno o simplemente uno. Advertimos que es posible que $z \notin \mathcal{C}$.

Teorema 162 *La simetría central $\Sigma_z(x) = 2z - x$ cumple $\phi(\Sigma_z(x)) = \phi(x) + 4(\sigma(z, z - x) + f(z - x))$. Si la cuádrica no está contenida en un hiperplano afín se tiene que z es un centro de $\mathcal{C}$ si y solo si la forma lineal $g_z(y) = \sigma(z, y) + f(y)$ es la forma nula.*

Demostración. Lo primero es un puro cálculo. Sean como sean x y z,

$$\begin{aligned}
\phi(\Sigma_z(x)) &= \sigma(2z - x, 2z - x) + 2f(2z - x) + \alpha \\
&= 4\sigma(z, z) - 2\sigma(z, x) - 2\sigma(x, z) + \sigma(x, x) + 4f(z) - 2f(x) + \alpha \\
&\overset{1}{=} 4\sigma(z, z) - 4\sigma(z, x) + \sigma(x, x) + 4f(z) - 2f(x) + \alpha \\
&\overset{2}{=} [\sigma(x, x) + 2f(x) + \alpha] + 4\sigma(z, z) - 4\sigma(z, x) + 4f(z) - 4f(x) \\
&= \phi(x) + 4[\sigma(z, z - x) + f(z - x)] = \phi(x) + 4g_z(z - x).
\end{aligned}$$

Si $g_z = 0$ y $x \in \mathcal{C}$ se tiene $\phi(\Sigma_z(x)) = \phi(x) + 4g_z(z - x) = 0$, luego z es un centro. Recíprocamente, sea z un centro. Si fuese $g_z \neq 0$, entonces $g_z(x) = g_z(z)$ sería la ecuación de un hiperplano afín. Como $\phi(\Sigma_z(x)) = \phi(x) + 4g_z(z - x)$ lleva a $g_z(x) = g_z(z)$ para $x \in \mathcal{C}$, vemos que $g_z \neq 0$ es imposible. ♣

Podemos calcular centros de $\mathcal{C}$ tomando una base $\mathcal{U}$ de $\mathbb{E}$ arbitraria y preguntando si existen $z \in \mathbb{E}$ tales que $g_z(y) = \sigma(z, y) + f(y)$ sea la forma cero sobre $\mathbb{E}$. En tal caso, el teorema 162 da z como centro. Evidentemente, $g_z = 0$ si y solo si $g_z(u_j) = 0$ para $u_j \in \mathcal{U}$, que es un sistema lineal,

$$\left\{\begin{array}{c} s_{11}z^1 + \ldots + s_{1n}z^n + f_1 = 0 \\ \vdots\vdots \\ s_{n1}z^1 \ldots + s_{nn}z^n + f_n = 0 \end{array}\right. , \quad \text{o} \quad \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix} + \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{8.6}$$

El sistema no es homogéneo si $f \neq 0$, pero cada posible solución z es un centro. En particular, si s es invertible, que es como decir que σ es no degenerada, $\mathcal{C}$ tiene exactamente un centro. Igualmente vemos que, si la ecuación $\phi(x) = 0$ cumple que $f = 0$, obtenemos, porque el sistema es homogéneo, que sea como sea σ hay al menos un centro. En general, si añadiendo a s la columna $(f_1, \ldots, f_n)^\top$, no aumenta el rango de la matriz ampliada, es porque tenemos un centro. Como ejemplo sencillo en $\mathbb{R}^2$,

$$\mathcal{C} : \phi\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (2, 2)\begin{pmatrix} x \\ y \end{pmatrix} - 8 = 0.$$

No es vacía porque $(-2, -2) \in \mathcal{C}$, y vemos que $z = (-1, -1)^\top$ es centro suyo pero $z \notin \mathcal{C}$.

8.5.2. Ecuación normalizada si hay centros

Vamos a determinar la ecuación normalizada de $\mathcal{C} : \phi(x) = \sigma(x, x) + 2f(x) + \alpha = 0$ en una referencia $\mathcal{R}$. Será esencial una buena elección del origen z de la referencia, que será un centro de $\mathcal{C}$ si $\mathcal{C}$ los tiene.

El caso en que $\mathcal{C}$ to tenga centros queda para la sección siguiente. Llamaremos **base de Sylvester de** σ a una base $\mathcal{V}=(v_1,\ldots,v_n)$ que diagonaliza σ como en el teorema de Sylvester (teorema 161); es decir, con signatura (p,q) y rango $r=p+q$,

$$\left\{\begin{array}{l}\sigma(v_i,v_i)=1 \text{ si } i\leq p\\ \sigma(v_j,v_j)=-1 \text{ si } p<j\leq p+q\end{array}\right. \qquad \left\{\begin{array}{l}\sigma(v_i,v_j)=0 \text{ si } i\neq j\\ \sigma(v_k,v_k)=0 \text{ si } k>r\end{array}\right. \tag{8.7}$$

Teorema 163 *Existe una referencia* $\mathcal{R}=(z,z+v_1,\ldots,z+v_n)$ *donde* z *es un centro de* $\mathcal{C}$ *y* $\mathcal{V}$ *una base de Sylvester tal que* $\mathcal{C}$ *tiene ecuación*

$$\phi_{\mathcal{R}}\left(x^1,\ldots,x^n\right)=\sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+\phi(z)=0.$$

Posiblemente $z\notin\mathcal{C}$*, pero si* $z\in\mathcal{C}$ *la ecuación se simplifica hasta* $\sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2=0$.

Demostración. Vamos a necesitar el desarrollo de $\phi(x)$ para $x=z+y$. Sean como sean y,z,

$$\begin{aligned}\phi(z+y)&=\sigma(z+y,z+y)+2f(z+y)+\alpha\\&=\sigma(z,z)+\sigma(z,y)+\sigma(y,z)+\sigma(y,y)+2f(z)+2f(y)+\alpha\\&=2\sigma(z,y)+2f(y)+\sigma(y,y)+[\sigma(z,z)+2f(z)+\alpha]\\&=2[\sigma(z,y)+f(y)]+\sigma(y,y)+\phi(z)\end{aligned}$$

Sea z un centro, por lo que $g_z(t)=\sigma(z,t)+f(t)$ es la forma cero por el teorema 162. Si $y=z-x$,

$$\phi(x)=\phi(z+y)=\sigma(y,y)+\phi(z)=\sigma(x-z,x-z)+\phi(z).$$

Para cualquier base de Sylvester de σ, al ser $x-z=y=\sum_{i=1}^{n}x^iv_i$, resulta

$$\begin{aligned}0=\phi_{\mathcal{R}}\left(x^1,\ldots,x^n\right)=\phi(x)=\phi\left(z+\sum_{i=1}^{n}x^iu_i\right)=\sigma\left(\sum_{i=1}^{n}x^iu_i,\sum_{i=1}^{n}x^iu_i\right)+\phi(z)\\=\sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+\phi(z)\end{aligned}$$

por las propiedades de $\mathcal{V}$. La última afirmación del enunciado es evidente. ♣

Sea $\mathcal{C}:x^2+y^2+2xy+2x+2y=0$ en $\mathbb{R}^2$ que reescribimos como

$$\mathcal{C}:\phi\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x\\y\end{pmatrix}^{\top}\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+2(1,1)\begin{pmatrix}x\\y\end{pmatrix}=0.$$

Con operaciones obvias

$$\left(\begin{array}{cc}1&1\\\hline 1&0\\0&1\end{array}\right)\to\left(\begin{array}{cc}1&1-1\\\hline 1&0-1\\0&1-0\end{array}\right)=\left(\begin{array}{cc}1&0\\\hline 1&-1\\0&1\end{array}\right)\Rightarrow\left(\begin{array}{cc}1&0\\\hline 1&-1\\0&1\end{array}\right)$$

y la base $\mathcal{V}=\left((1,0)^{\top},(-1,1)^{\top}\right)$ cumple (8.7). En la referencia cuyo origen es el centro $(-1/2,-1/2)^{\top}$,

$$\mathcal{R}=\left(\begin{pmatrix}-1/2\\-1/2\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}-1\\1\end{pmatrix}\right)$$

la ecuación de $\mathcal{C}$ en las coordenadas asociadas (p,q) será $p^2=1$ ya que $\phi\left((-1/2,-1/2)^{\top}\right)=-1$. La cónica está formada por dos rectas paralelas $p=\pm1$ en las coordenadas del teorema.

8.5.3. Ecuación normalizada si no hay centros

Recordamos que $\mathbb{K}=\{x\in\mathbb{E}\mid \sigma(x,y)=0\ \forall y\in\mathbb{E}\}$ es el espacio nulo de σ y en una base de Sylvester $\mathcal{V}$ es $\mathbb{K}=\lg(v_{r+1},\ldots,v_n)$. *Será crucial el que f no puede anularse sobre $\mathbb{K}$ porque no hay centros.* En efecto, demostramos que si f es nula sobre $\mathbb{K}$, tiene que haber centros. Todo es cosa de calcular con una base de Sylvester teniendo en cuenta que $f(v_{r+1})=\ldots=f(v_n)=0$. Sabemos por el teorema 162 que z será un centro si $g_z(x)=\sigma(z,x)+f(x)$ es nula para todo x. Pero en coordenadas esto es

$$\begin{aligned}0&=z^1x^1+\ldots+z^px^p-z^{p+1}x^{p+1}-\ldots-z^rx^r+f_1x^1+\ldots+f_px^p+f_{p+1}x^{p+1}+\ldots+f_rx^r\\&=\left(z^1+f_1\right)x^1+\ldots+\left(z^p+f_p\right)x^p+\left(f_{p+1}-z^{p+1}\right)x^{p+1}+\ldots+\left(f_r-z^r\right)x^r\end{aligned}$$

y valdría como centro cualquier z verificando $z^1+f_1=\ldots=f_r-z^r=0$.

Sabido que f no es nula sobre $\mathbb{K}$, tomamos un $w\in\mathbb{K}-\ker(f)$ cumpliendo $f(w)=1$. Puede elegirse λ tal que $v=\lambda w$ cumpla $\phi(v)=0$. Esto es fácil porque, al estar v en $\mathbb{K}$,

$$\phi(v)=\sigma(v,v)+2f(v)+\alpha=2\lambda f(w)+\alpha,\ \text{y se toma}\ \ \lambda=-\alpha/2.$$

Todos estos cálculos se hacen porque pronto veremos que v (que, por cierto, está en $\mathcal{C}$ al ser $\phi(v)=0$) es un buen origen para la "buena" referencia. En la demostración del teorema 163 se ha probado que sean como sean z,y se tiene

$$\phi(z+y)=2\left[\sigma(z,y)+f(y)\right]+\sigma(y,y)+\phi(z).$$

Pues bien, si hacemos $z=v$ esto se reduce a $\phi(v+y)=\sigma(y,y)+2f(y)$.

Teorema 164 *Sea $\mathcal{C}:\sigma(x,x)+2f(x)+\alpha=0$ sin centros. Existe v en $\mathbb{K}\cap\mathcal{C}$ tal que $2f(z)+\alpha=0$, en cuyo caso para $x=z+y$ se tiene como ecuación de $\mathcal{C}$,*

$$\mathcal{C}:\phi(x)=\sigma(x-v,x-v)+2f(x-v).$$

Se puede elegir una base de Sylvester $\mathcal{V}$ de modo que sea $f(v_{r+1})=\ldots=f(v_{n-1})=0$ y $f(v_n)=1$ y en la referencia $\mathcal{R}=(v,\mathcal{V})$ se tiene la ecuación

$$\mathcal{C}:\phi_{\mathcal{R}}\left(x^1,\ldots,x^n\right)=\sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+2x^n=0.$$

Demostración. Si hacemos $x=v+y$ tenemos $\phi(x)=\sigma(x-v,x-v)+2f(x-v)$ por el cálculo anterior al enunciado. Tomamos una base $\mathcal{U}=(u_1,\ldots,u_m)$ con $u_n=w$ y $f(u_1)=\ldots=f(u_{n-1})=0$. Si calculamos la matriz s de σ en esta base, al estar $w=u_n$ en $\mathbb{K}$ se cumple que la última fila y la última columna de s son nulas. Hacemos operaciones fila y columna en la matriz para reducir s a forma diagonal con diagonal $(1,\ldots,1,-1,\ldots,-1,0,\ldots,0)$ con p unos y q menos uno (o sea, (p,q) es la signatura de s y σ). Tenemos una base de Sylvester $\mathcal{V}=(v_1,\ldots,v_n)$ con $v_n=w$ porque no ha habido que tocar la última fila y columna de s. Además, como los v_i para $i<n$ son combinación lineal de $(u_1,\ldots,u_{n-1})$ debe ser $f(v_1)=\ldots=f(v_{n-1})=0$. Tomemos como referencia $\mathcal{R}=(v,v_1,\ldots,v_n)$. Sabemos que x tiene coordenadas $\left(x^1,\ldots,x^n\right)$ si $x=z+\sum_{h=1}^{n}x^hv_h$. Como $\mathcal{V}$ es base de Sylvester, $f(v_1)=\ldots=f(v_{n-1})=0$ y $f(v_n)=f(v)=1$, obtenemos enseguida

$$\phi\left(x^1,\ldots,x^n\right)=\sigma\left(\sum_{h=1}^{n}x^hv_h,\sum_{h=1}^{n}x^hv_h\right)+2f\left(\sum_{h=1}^{n}x^hv_h\right)=\sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+2x^n$$

y el teorema está probado. ♣

Uniendo los teoremas 163 y 164 tenemos las ecuaciones normalizadas.

Teorema 165 *Existe una referencia afín donde $\mathcal{C}$ tiene ecuación del tipo*

$$\textbf{\textit{(I)}}\ \sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+\beta=0\quad o\quad \textbf{\textit{(II)}}\ \sum_{i=1}^{p}\left(x^i\right)^2-\sum_{j=p+1}^{p+q}\left(x^j\right)^2+2x^{r+1}=0,$$

siendo (p,q) la signatura de σ.

Mejor que la demostración formal describiremos los pasos para llegar a las ecuaciones.

1. Determinar si $\mathcal{C}$ tiene o no tiene centros. Estos centros, según vimos en (8.6), son soluciones de ese sistema lineal. Si solo queremos saber si hay o no centros, basta poner la condición de compatibilidad $\operatorname{rg}(s) = \operatorname{rg}(s \mid f)$ (el teorema de Frobenius) y, si se quiere, se pueden calcular los centros. Si hay centros, la ecuación será del tipo **I**; si no, del tipo **II**.

2. Calcular la signatura (p, q) de σ. Con lo que hasta ahora sabemos, no hay más remedio que diagonalizarla, pero luego explicaremos un atajo.

3. Tras conocer (p, q) hay varias posibilidades:

   a) Hay centros y uno de ellos z está en $\mathcal{C}$. Entonces la ecuación es del tipo **I** en el teorema 165 con $\beta = 0$. La ecuación cambiada de signo da la misma cuádrica.

   b) Hay centros pero ninguno en $\mathcal{C}$.Entonces la ecuación es del tipo **I** en el teorema 165 con $\beta \neq 0$. Lo que se suele hacer es considerar como ecuación estándar la que resulta al dividir por $-\beta$ para que queden ecuaciones típicas en aplicaciones con $x^2 + y^2 - z^2 = 1$.

   c) No hay centros. Para ello es necesario que σ sea degenerada. Entonces $p+q < n$ y la ecuación es del tipo **II** en el teorema 165. A veces se prefiere cambiar la coordenada x^n por $-\frac{1}{2}x^n$ para obtener una ecuación del tipo $x^n = \sum_{i=1}^{p} \left(x^i\right)^2 - \sum_{j=p+1}^{p+q} \left(x^j\right)^2$ y ver $\mathcal{C}$ como el grafo de una función de $\mathbb{R}^{n-1}$ en $\mathbb{R}$.

Cuando nos pidan dar la ecuación normalizada de $\mathcal{C}$ se entiende que es una ecuación como las del teorema 165 salvo factor constante no nulo. Como en otros problemas similares, es mucho más fácil conocer el tipo de ecuación (con ambigüedades de signo) que calcular la referencia.

Problema 308 *Encontrar una ecuación tipo **I** o **II** para la cuádrica (se supone $\mathbb{E} = \mathbb{R}^3$)*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}^{\top} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 2 \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}^{\top} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 4 = 0. \blacklozenge$$

Solución. Tenemos $C(X) = -(X-3)X^2$ que tiene una raíz positiva y ninguna negativa, luego $(p, q) = (1, 0)$. Además

$$\operatorname{rg}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 1 \quad \text{y} \quad \operatorname{rg}\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \end{pmatrix} = 2$$

luego $\mathcal{C}$ no tiene centros y tendrá una ecuación tipo **II** de forma concreta $u^2 + 2v = 0$. $\blacklozenge$

Problema 309 *Ídem para*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}^{\top} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 2(0, 0, h) \begin{pmatrix} x \\ y \\ z \end{pmatrix} + h = 0$$

siendo $h \in \mathbb{R} - \{0\}$. (Se entiende que la respuesta dependerá del valor de h.)

Problema 310 *Sea z un centro de $\mathcal{C}$. Probar que el conjunto de todos los centros es el subespacio afín $z + \mathbb{K}$, siendo $\mathbb{K}$ el subespacio nulo de σ.*

8.5.4. Tipos de cónicas y cuádricas en dimensión 2 y 3

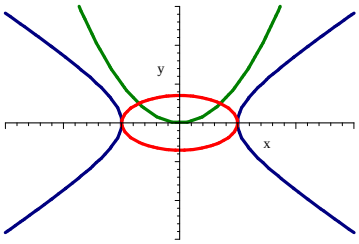
Enumeramos los tipos de cónicas ($n = 2$) y cuádricas ($n = 3$) que no son $\emptyset$ o un subespacio afín. Lo hacemos dando su ecuación distinguida y con (x, y) o (x, y, z) en vez de $\left(x^1, x^2\right)$ o $\left(x^1, x^2, x^3\right)$. Esta es

la tabla para $n = 2$.

¿centros?	¿alguno en $\mathcal{C}$?	signatura	ejemplo tipo	nombre
sí	sí	$(1,1)$	$x^2 - y^2 = 0$	par de rectas secantes
sí	no	$(1,0)$	$x^2 = 1$	par de rectas no secantes
sí	no	$(2,0)$	$x^2 + y^2 = 1$	elipse
sí	no	$(1,1)$	$x^2 - y^2 = 1$	hipérbola
no		$(1,0)$	$x^2 + 2y = 0$	parábola

Hemos seguido la costumbre de poner $x^2 + y^2 = 1$ en vez de $x^2 + y^2 - 1 = 0$.

La gráficas de una elipse (rojo), hipérbola (azul) y parábola (verde) están en la siguiente figura.



Puede extrañar que digamos que $x^2 + y^2 = 1$ es una elipse y no un círculo. La libertad que tenemos de elegir coordenadas es grande y esto hace que se distorsionen las longitudes (no se ha definido formalmente "longitud") lo que hace que, vista la figura en esas coordenadas, no haya achatamientos. Cuando tratemos la clasificación euclidiana esto no pasará.

En el caso $n = 3$ la tabla es

¿centros?	¿alguno en $\mathcal{C}$?	signatura	ejemplo tipo	**nombre**
si	si	$(1,1)$	$x^2 - y^2 = 0$	par de planos secantes
sí	si	$(2,1)$	$x^2 + y^2 - z^2 = 0$	cono
sí	no	$(1,0)$	$x^2 = 1$	par de planos no secantes
sí	no	$(2,0)$	$x^2 + y^2 = 1$	cilindro (elíptico)
sí	no	$(3,0)$	$x^2 + y^2 + z^2 = 1$	elipsoide
sí	no	$(1,1)$	$x^2 - y^2 = 1$	cilindro hiperbólico
sí	no	$(2,1)$	$x^2 + y^2 - z^2 = 1$	hiperboloide de una hoja
sí	no	$(1,2)$	$x^2 - y^2 - z^2 = 1$	hiperboloide de dos hojas
no		$(1,0)$	$x^2 - y = 0$	cilindro parabólico
no		$(2,0)$	$x^2 + y^2 - z = 0$	paraboloide (elíptico)
no		$(1,1)$	$x^2 - y^2 - z = 0$	paraboloide hiperbólico

Es casi seguro que el lector estará un tanto confundido en cuanto a cómo se ha hecho la tabla (¡lo difícil es que estén todos los que son y sean todos los que están!), la lógica al elegir los nombres, y la idea de la superficie del espacio que representan.

Primero van los casos con centros y alguno de ellos en $\mathcal{C}$, luego los que tienen centros pero fuera de $\mathcal{C}$, y finalmente los que no tienen centros. La elección de terminología es más difícil de explicar así como lo que las superficies representan. Descartemos por obvios los pares de planos, luego quedan nueve cuádricas.

1. La palabra "cono" se refiere a un subconjunto de $\mathbb{R}^3$ formado por una unión de rectas que pasan por el origen y por los puntos de una curva en un plano que no pasa por él. Si imaginamos el círculo $x^2 + y^2 = 1$ en el plano $z = 1$ resulta el **cono de revolución**[12] ordinario, que es el que aquí tiene ecuación $x^2 + y^2 - z^2 = 0$. El lector puede convencerse cortando el subconjunto de ecuación $x^2 + y^2 - z^2 = 0$ por planos horizontales $z = r$ y obtendrá círculos $x^2 + y^2 = r^2$. Si en vez de

[12] La expresión "de revolución" que se va a usar varias veces no es más que una ayuda a la intuición basada en que una superficie de revolución es la que se obtiene al girar una curva del espacio en torno a un eje.

$x^2+y^2=1$ ponemos $2x^2+y^2=1$ sale el **cono** (llamado **cono elíptico** porque el plano $z=r$ corta en elipses) pero nadie usa el adjetivo. Se podría imaginar una hipérbola o parábola en $z=1$ y hacer la misma operación que hemos propuesto para llegar al cono de revolución. Desde luego se tendría un "cono" pero no sería una cuádrica (se necesitaría una ecuación polinómica de grado ≥ 3). Por eso, mientras nos limitemos a cuádricas se puede hablar del cono $x^2+y^2-z^2=0$, llamándole incluso "cono elíptico" pero no va a haber "cono hiperbólico" o "cono parabólico".

2. Hay otra figura similar al cono, que es el cilindro, y ahí sí que aparecen **cilindros elípticos**, **hiperbólicos** y **parabólicos**. Si se tiene una curva en el plano $z=0$ y se trazan todas las rectas verticales que se apoyan en ella, se tiene un **cilindro**. El caso más sencillo es el de la curva $x^2+y^2=1$, que da el **cilindro de revolución** (que aquí llamamos **cilindro elíptico**). Si como curva en $z=0$ se toman hipérbolas o parábolas se obtienen los cilindros hiperbólicos y parabólicos. Sus ecuaciones son como las de la elipse, hipérbola y parábola pero "sin coordenada z".

3. El **paraboloide** ordinario es $x^2+y^2=z$, que hemos llamado **paraboloide elíptico**. Cortando por planos $z=\kappa$ se obtienen elipses o círculos contando con $z>0$. Es una copa, con simetría de revolución si $\alpha=\beta$, con el fondo en el origen. El **paraboloide hiperbólico** es, por ejemplo, $x^2-y^2=z$ y es el grafo de $z=x^2-y^2$, que representa una silla de montar o un collado entre dos montañas. Esto se ve bien cortando por planos $z=\kappa$ que no dan hipérbolas como sección.

4. El **elipsoide** es la generalización natural de la elipse $x^2+y^2=1$ a tres dimensiones con $x^2+y^2+z^2=1$. Tiene forma de lenteja o balón de rugby y si se corta por planos con x, y o z constante, resultan, si no hay intersección vacía, elipses o un punto en casos extremos. Los **hiperboloides de una y dos hojas** son, si se toma la ecuación más sencilla, $x^2+y^2-z^2=1$ (una hoja) y $x^2-y^2-z^2=1$ (dos hojas). El de una hoja tiene la forma de las chimeneas de las centrales térmicas. Cortando por planos $z=\kappa$ (constante) resultan elipses (círculos) $x^2+y^2=1+\kappa^2$ cualquiera que sea κ luego la chimenea tiene eje vertical. El de dos hojas está formado por dos copas simétricas respecto a $x=0$. En efecto, si cortamos por planos con normal horizontal $x=\kappa$ (constante), la ecuación de la intersección es $y^2+z^2=\kappa^2-1$, que es vacía si $|\kappa|<1$ pero que da elipses (círculos) si $|\kappa|>1$ y puntos, los fondos de las copas, si $|\kappa|=1$. En cualquiera de los dos hiperboloides, al cortar por planos $y=\kappa$ o $z=\kappa$ obtenemos hipérbolas; por ejemplo $x^2-y^2=1+\kappa^2$.

5. Para ayudar a la memoria del lector en esta jungla de nombres y apellidos le diremos en tono informal que "cono" no hay más que uno, el elíptico, que se visualiza como un cono de revolución; cilindros hay tres y el calificativo corresponde a la curva que lo genera al deslizarse en paralelo al eje OZ; y paraboloides hay dos, el elíptico y el hiperbólico, dependiendo el calificativo de las curvas que aparecen al cortar por planos $z=\kappa$ (horizontales). Los hiperboloides tienen un tipo de calificativo diferente (de una o dos hojas). Si se dibujan, el primero es una superficie "continua" que al cortarse por planos $z=\kappa$ siempre hay puntos. El segundo es "discontinuo" pues al cortar por $z=0$ queda $x^2+y^2=-1$ que no tiene solución. Y un último aviso al lector: no hay paraboloides ni cilindros de dos hojas, conos hiperbólicos o paraboloides parabólicos, ni otros más que su calenturienta imaginación puede inventar.

Problema 311 *Determinar para que valores de $h\in\mathbb{R}$ la cónica*

$$\phi\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}x\\y\end{pmatrix}^{\top}\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+2\begin{pmatrix}h\\1\end{pmatrix}^{\top}\begin{pmatrix}x\\y\end{pmatrix}+1=0$$

es una parábola. ⧫

Solución. La signatura de s es $(1,0)$ ya que $C_s(X)=X(X-2)$. Para tener una parábola se necesita que no tenga centros; es decir, que sea

$$\operatorname{rg}\begin{pmatrix}1&1&h\\1&1&1\end{pmatrix}>1,$$

lo que es posible para $h\neq 1$ exclusivamente. ⧫

Problema 312 *Para la cuádrica $\mathcal{C}$ de $\mathbb{R}^3$ de ecuación*

$$\phi\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}x\\y\\z\end{pmatrix}^{\top}\begin{pmatrix}h&1&0\\1&h&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}+2\begin{pmatrix}0\\0\\1\end{pmatrix}^{\top}\begin{pmatrix}x\\y\\z\end{pmatrix}+1=0$$

dependiente de $h \in \mathbb{R}$, determinar los valores de h, si existen, para los que $\mathcal{C}$ sea un hiperboloide, distinguiendo que sea de una o dos hojas.

Problema 313 *Lo mismo que en el problema anterior con h al final en vez de 1; o sea*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}^\top \begin{pmatrix} h & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 2\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} + h = 0.$$

Problema 314 *¿Cómo tiene que ser $(p,q,r) \neq 0$ para que $\mathcal{C}$ dada por*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}^\top \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 2\begin{pmatrix} p \\ q \\ r \end{pmatrix}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$$

sea un cilindro? ¿Hay alguno de los tres tipos de cilindro que nunca pueda darse?¿Cuáles son sus centros?

8.5.5. Cálculos alternativos para clasificar cuádricas

En una referencia $\mathcal{R}$, la cuádrica $\mathcal{C}$ tendrá ecuación

$$\phi_{\mathcal{R}}\left(x^1,\ldots,x^n\right) = \sum_{i,j=1}^{n} x^i s_{ij} x^j + 2\sum_{h=1}^{n} F_h x^h + \alpha,$$

siendo $s \in \mathbb{R}^{n\times n}$ simétrica, $F \in \mathbb{R}^{1\times n}$ (matriz fila) y $\alpha \in \mathbb{R}$. Si definimos,

$$M\left(\phi_{\mathcal{R}}\right) = M\left(s,F,\alpha\right) = \begin{pmatrix} \alpha & F_1 & \cdots & F_n \\ F_1 & s_{11} & \cdots & s_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ F_n & s_{n1} & \cdots & s_{nn} \end{pmatrix} = \begin{pmatrix} \alpha & F \\ F^\top & s \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

tenemos una fórmula matricial para $\phi_{\mathcal{R}}$ como alternativa a la polinómica, que es

$$\phi_{\mathcal{R}}\left(x\right) = \left(1,x^1,\ldots,x^n\right)\begin{pmatrix} \alpha & F_1 & \cdots & F_n \\ F_1 & s_{11} & \cdots & s_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ F_n & s_{n1} & \cdots & s_{nn} \end{pmatrix}\begin{pmatrix} 1 \\ x^1 \\ \vdots \\ x^n \end{pmatrix} = (1,x)\begin{pmatrix} \alpha & F \\ F^\top & s \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}$$

Estas fórmulas ayudan a relacionar $\phi_{\mathcal{R}}$ y $\phi_{\mathcal{S}}$ cuando tenemos dos referencias afines $\mathcal{R}$ y $\mathcal{S}$. Si $\left(x^1,\ldots,x^n\right)$ e $\left(y^1,\ldots,y^n\right)$ son las coordenadas de p en $\mathcal{R}$ y $\mathcal{S}$ sabemos que hay una relación entre coordenadas $x = cy + h$ con $c \in \mathbb{R}^{n\times n}$ invertible y $h \in \mathbb{R}^n$. Coviene poner esto también con matrices $(n+1)\times(n+1)$ en la forma

$$\begin{pmatrix} 1 \\ x^1 \\ \vdots \\ x^n \end{pmatrix} = \begin{pmatrix} 1 & F_1 & \cdots & F_n \\ h^1 & c_1^1 & \cdots & c_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ h^n & c_1^n & \cdots & c_n^n \end{pmatrix}\begin{pmatrix} 1 \\ y^1 \\ \vdots \\ y^n \end{pmatrix} \text{ o bien } \begin{pmatrix} 1 \\ x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix}.$$

Tenemos entonces

$$\begin{aligned}\phi\left(p\right) = \phi_{\mathcal{R}}\left(x\right) &= \begin{pmatrix} 1 \\ x \end{pmatrix}^\top \begin{pmatrix} \alpha & F \\ F^\top & s \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix} = \left[\begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix}\right]^\top \begin{pmatrix} \alpha & F \\ F^\top & s \end{pmatrix}\begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ y \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}^\top \begin{pmatrix} \alpha & F \\ F^\top & s \end{pmatrix}\begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix},\end{aligned}$$

que comparada con

$$\phi\left(p\right) = \phi_{\mathcal{S}}\left(y\right) = \begin{pmatrix} 1 \\ y \end{pmatrix}^\top \begin{pmatrix} \beta & G \\ G^\top & t \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix},$$

nos lleva a que

$$M\left(\phi_{\mathcal{S}}\right) = \begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}^\top M\left(\phi_{\mathcal{R}}\right)\begin{pmatrix} 1 & 0 \\ h & c \end{pmatrix}.$$

Teorema 166 *Las signaturas de $M(\phi_{\mathcal{S}})$ y $M(\phi_{\mathcal{S}})$ y de las matrices s y t son las mismas.*

Demostración. Los primero se debe al cálculo, que muestra que $M(\phi_{\mathcal{S}})$ y $M(\phi_{\mathcal{S}})$ congruentes. Si se comparan las cajas en posición $(2,2)$ de $M(\phi_{\mathcal{S}})$ y $M(\phi_{\mathcal{S}})$ se comprueba que $t = c^{\top} sc$ y esta nueva congruencia da la igualdad de signaturas. ♣

Supongamos que $\mathcal{R}$ sea la referencia "mala" en donde inicialmente nos dan la ecuación de $\mathcal{C}$, y que $\mathcal{S}$ sea la referencia "buena" en donde $\phi_{\mathcal{S}}(y) = 0$ toma la forma sencilla del teorema 165. El rango y signatura de $M(\phi_{\mathcal{R}})$ serán los mismos que los de $M(\phi_{\mathcal{S}})$ y lo mismo con s y t, pero la simplicidad estructural de $M(\phi_{\mathcal{S}})$ permite reconstruirla con solo saber su signatura y la de t. *¡Es el punto esencial!* Enunciamos las posibilidades de $M(\phi_{\mathcal{S}})$ cuando representa una ecuación en la forma normalizada del teorema teorema 165. En ellas

1. $I_{p.q}$ es una matriz $n \times n$ diagonal con diagonal $(1,\dots,1,-1,\dots,-1,0,\dots,0)$ habiendo p unos y q menos unos.
2. Representamos con 0 tanto el número 0 como el vector 0 de $\mathbb{R}^{n\times 1}$ o $\mathbb{R}^{1\times n}$, y $\beta \in \mathbb{R} - \{0\}$.
3. $u = (0,\dots,0,1)$ con solo un 1 al final. Si aparece u entendemos que $p+q<n$.

Las posibilidades son

$$(\mathbf{1})\ M(\phi_{\mathcal{S}}) = \begin{pmatrix} 0 & 0 \\ 0 & I_{p,q} \end{pmatrix} \text{ con signaturas } (p,q) \text{ para } M(\phi_{\mathcal{S}}) \text{ e } I_{p,q}$$

$$(\mathbf{2})\ M(\phi_{\mathcal{S}}) = \begin{pmatrix} \beta & 0 \\ 0 & I_{p,q} \end{pmatrix} \text{ con signaturas } (p,q) \text{ para } I_{p,q} \text{ y } (p+1,q) \text{ o } (p,q+1) \text{ para } M(\phi_{\mathcal{S}}),$$

$$(\mathbf{3})\ M(\phi_{\mathcal{S}}) = \begin{pmatrix} 0 & u \\ u^{\top} & I_{p,q} \end{pmatrix} \text{ con signaturas } (p,q) \text{ para } I_{p,q} \text{ y } (p+1,q+1) \text{ para } M(\phi_{\mathcal{S}}).$$

Problema 315 *Comprobar que en **(3)** la signatura de $M(\phi_{\mathcal{S}})$ es $(p+1,q+1)$.* ♦

Solución. El problema es sobre todo elegir una buena notación. Vamos a descomponer $M(\phi_{\mathcal{S}})$, que es $(n+1)\times(n+1)$ en cajas de diferente tamaño que vamos a describir. En la parte central irá una caja $J \in \mathbb{R}^{(n-1)\times(n-1)}$ que es la matriz que se obtiene al quitar a $I_{p,q}$ su última fila y columna. ¡Atención! como $p+q<n$, todo lo que se quita a $I_{p,q}$ para pasar a J son ceros, que irán en la última fila y columna de $M(\phi_{\mathcal{S}})$. Representaremos por 0_{n-1} matrices fila o columna con $n-1$ huecos y todo ceros, y los 0 o 1 de las esquinas son números. Con todo esto la forma final de $M(\phi_{\mathcal{S}})$ es

$$M(\phi_{\mathcal{S}}) = \begin{pmatrix} 0 & 0_{n-1} & 1 \\ 0_{n-1} & J & 0_{n-1} \\ 1 & 0_{n-1} & 0 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

Un ejemplo con $n=1$ y $p=q=1$ sería

$$M(\phi_{\mathcal{S}}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad 0_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \text{ o } (0,0,0).$$

Así puestas las cosas, permutamos columnas y filas pasando la última fila y columna al lugar 2,

$$\begin{pmatrix} 0 & 0_{n-1} & 1 \\ 0_{n-1} & J & 0_{n-1} \\ 1 & 0_{n-1} & 0 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 0_{n-1} \\ 0_{n-1} & 0_{n-1} & J \\ 1 & 0 & 0_{n-1} \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 1 & 0_{n-1} \\ 1 & 0 & 0_{n-1} \\ 0_{n-1} & 0_{n-1} & J \end{pmatrix}.$$

Con operaciones fila y columna que afectan solo a las dos primeras, se pasa enseguida a

$$\begin{pmatrix} 0 & 1 & 0_{n-1} \\ 1 & 0 & 0_{n-1} \\ 0_{n-1} & 0_{n-1} & J \end{pmatrix} \to\Rightarrow \begin{pmatrix} 2 & 0 & 0_{n-1} \\ 0 & -2 & 0_{n-1} \\ 0_{n-1} & 0_{n-1} & J \end{pmatrix},$$

y como J es diagonal con p unos y q menos unos en ella, la signatura de $M(\phi_{\mathcal{S}})$ es $(p+1,q+1)$. ♦

Problema 316 *Si* $\mathcal{C} : x^2 + 2xy + py^2 + 4x + 2 = 0$ *y* $p > 0$, *determinar su tipo de ecuación.* ♦

Solución. Las matrices a estudiar son

$$s = \begin{pmatrix} 1 & 1 \\ 1 & p \end{pmatrix}, \qquad M = \begin{pmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & p \end{pmatrix}$$

Sean (h, k) y (H, K) sus respectivas signaturas. Para calcular (h, k) hacemos las operaciones

$$\begin{pmatrix} 1 & 1 \\ 1 & p \end{pmatrix} \overset{1}{\to} \begin{pmatrix} 1 & 0 \\ 1 & p-1 \end{pmatrix} \overset{2}{\to} \begin{pmatrix} 1 & 0 \\ 0 & p-1 \end{pmatrix}$$

y lo mismo con M,

$$\begin{pmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & p \end{pmatrix} \overset{1}{\to} \begin{pmatrix} 2 & 0 & 0 \\ 2 & -1 & 1 \\ 0 & 1 & p \end{pmatrix} \overset{2}{\to} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & p \end{pmatrix} \overset{3}{\to} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & p+1 \end{pmatrix} \overset{4}{\to} \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & p+1 \end{pmatrix}$$

Distinguimos los casos $0 < p < 1$, $p = 1$, $p > 1$

En el caso $0 < p < 1$ es $(h, k) = (1, 1)$ y $(H, K) = (2, 1)$. La ecuación es $u^2 - v^2 + 1 = 0$ salvo factor positivo.

En el caso $p = 1$ es $(h, k) = (1, 0)$ y $(H, K) = (2, 1)$. La ecuación es $u^2 + 2v = 0$

En el caso $1 < p$ es $(h, k) = (2, 0)$ y $(H, K) = (2, 1)$. La ecuación es $u^2 + v^2 + 1 = 0$ salvo factor positivo.

Los tres casos corresponden en la clasificación a una hipérbola, una parábola y el conjunto vacío. ♦

Problema 317 *Supongamos que solo se quiere saber si la cuádrica está en uno de los grupos* ***(a)*** *con centros y alguno en* $\mathcal{C}$, ***(b)*** *con centros pero ninguno en* $\mathcal{C}$, *y* ***(c)*** *sin centros. Probar que basta para ello conocer los rangos* r *y* R *de* s *y* M *porque el estar en* ***(a)***,***(b)*** *y* ***(c)*** *equivale respectivamente a* $r = R$, $r = R - 1$ *y* $r = R - 2$.

Capítulo 9

Productos euclidianos

9.1. Fundamentos

En este capítulo $\mathbb{E}$ será, salvo advertencia, un espacio real y $\omega(x,y) = \langle x, y\rangle$ un producto euclidiano.

9.1.1. Principales definiciones y ejemplos

Un **producto escalar** o **euclidiano**[1] en $\mathbb{E}$ es una forma bilineal simétrica ω en $\mathbb{E}$ tal que $\omega(x,x) > 0$ si $x \neq 0$. La condición $\omega(x,x) > 0$ si $x \neq 0$ es precisamente la condición de que ω sea **definida positiva**. No se necesita dimensión finita para definir estos productos. El ejemplo más importante, el arquetipo[2] de producto euclidiano, es el de $\mathbb{E} = \mathbb{R}^n$ dado por

$$\varepsilon(x,y) = x^1y^1 + \ldots + x^ny^n = \sum_{i=1}^{n} x^iy^i = \left(x^1, \ldots, x^n\right)\begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} = x^\top y,$$

que se llamará el **producto (escalar o euclidiano) estándar** y juega un papel similar a la base estándar de $\mathbb{R}^m$. Se llama **espacio euclidiano** al par $(\mathbb{E}, \omega)$ siendo $\mathbb{E}$ un espacio real y ω un producto euclidiano. Al hablar del "par" $(\mathbb{E}, \omega)$ se entiende que si cambiamos ω por otro producto euclidiano η manteniendo $\mathbb{E}$ se tiene un espacio euclidiano distinto $(\mathbb{E}, \eta)$. Llamaremos a $(\mathbb{R}^m, \varepsilon)$ el **espacio euclidiano estándar**. Casi siempre trabajaremos con ω fijo y usaremos la notación alternativa $\omega(x,y) = \langle x, y\rangle$. Hay varios conceptos fundamentales al disponer de un producto euclidiano ω.[3]

1. La **norma**, **módulo**, o **longitud** del vector $x \in \mathbb{E}$ es $\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\omega(x,x)}$. Obsérvese que la raíz existe pues $\omega(x,x) \geq 0$. En $(\mathbb{R}^m, \varepsilon)$ es inmediato que

$$\|x\| = \sqrt{(x^1)^2 + \ldots + (x^m)^2} = \sqrt{\sum_{i=1}^{m} (x^i)^2}$$

   y representa en el sentido intuitivo (al menos para $m = 1, 2, 3$) la longitud de la flecha con que se visualiza x en $\mathbb{R}^m$. Cualquier ω cumple que $\|x\| \geq 0$, $\|\lambda x\| = |\lambda| \|x\|$, y que $\|x\| = 0$ si y solo si $x = 0$, todo ello fácil de verificar.

2. Un **vector unitario** x es el que cumple $\|x\| = 1$. Si $x \neq 0$, se le puede **normalizar**, que es construir con él $x/\|x\|$. Este vector es unitario pues

$$\left\langle \frac{x}{\|x\|}, \frac{x}{\|x\|} \right\rangle = \frac{1}{\|x\|^2}\langle x, x\rangle = \frac{\|x\|^2}{\|x\|^2} = 1.$$

[1] En inglés se usa mucho *inner product* aunque *producto interno* o *interior* es menos frecuente en español. La frecuente notación $\omega(x,y) = x \cdot y$ hace que en inglés se use también *dot product.* Salvo por razones de espacio en cálculos, no usaremos $x \cdot y$ pues es fácil confundirlo con $\circ$ el signo de composición de funciones.

[2] arquetipo: modelo original y primario en un arte u otra cosa (según la RAE).

[3] Se entiende que la norma o longitud, la perpendicularidad, y el ser base ortogonal u ortonormal dependen de la elección de ω. Se podría usar una notación como $\|x\|_\omega$ para indicar la norma de x según ω, pero no es necesario pues casi siempre solo interviene un producto euclidiano. (Eso sí, que puede ser diferente de lo que se considera "estándar" o fácil de intuir.)

3. Dos vectores x, y son **ortogonales** o **perpendiculares** si $\langle x, y\rangle = 0$. Damos la manera intuitiva de justificar la perpendicularidad. Si tenemos un triángulo de lados A, B, C y longitudes respectivas a, b, c, el **teorema de Pitágoras** dice que son equivalentes el ser $a^2 = b^2 + c^2$ y ser perpendiculares los lados B y C.[4] Como

$$\langle x+y, x+y\rangle = \langle x, x\rangle + \langle y, y\rangle + 2\langle x, y\rangle \text{ y } 2\langle x, y\rangle = \|x+y\|^2 - \|x\|^2 - \|y\|^2$$

deducimos para el triángulo cuyos lados son las flechas de los vectores $x+y, x$ e y que la condición $\langle x, y\rangle = 0$ equivale a $\|x+y\|^2 = \|x\|^2 + \|y\|^2$, y por tanto también a que las flechas de los vectores x e y sean perpendiculares en cualquier dibujo o esquema en donde las longitudes de las flechas estén correctamente representadas.

4. Un conjunto de vectores *no nulos* es **ortogonal** si para todo par de elementos x, y en él se tiene $\langle x, y\rangle = 0$. Frecuentemente hablaremos de **bases ortogonales**. La idea intuitiva es la de un conjunto de flechas (los vectores) perpendiculares entre sí. Una sucesión $(x_1, \ldots, x_k)$ de vectores ortogonales es independiente porque si tenemos $0 = \sum_{i=1}^{k} \lambda^i x_i$ multiplicamos por $\langle \bullet, x_j\rangle$ y resulta $0 = \lambda^i \|x_j\|^2$ y al ser $\|x_j\|^2 \neq 0$ debe ser $\lambda^i = 0$ con i arbitrario.

5. Un conjunto de vectores es **ortonormal** si es ortogonal y *además* todos sus elementos son unitarios. Frecuentemente hablaremos de **bases ortonormales.** La idea intuitiva es la de un conjunto de flechas (los vectores) perpendiculares entre sí y *además,* de longitud 1. De un conjunto ortogonal se pasa fácilmente a otro ortonormal normalizando sus elementos. En particular, si $\mathcal{V}$ es base ortogonal, $\mathcal{W}$ dada por

$$w_i = \frac{1}{\sqrt{\omega(v_i, v_i)}} v_i = \frac{1}{\|v_i\|} v_i, \quad 1 \leq i \leq n,$$

es ortonormal. En el espacio euclidiano estándar, la base estándar es ortonormal.

En lo que sigue se usará con frecuencia $\langle x, y\rangle$ en vez de $\omega(x, y)$ y $\|x\|^2$ en vez de $\omega(x, x)$ mientras no intervenga más que un producto ω.

Damos ejemplos fundamentales de productos euclidianos.

1. Tomamos $a \in \mathbb{R}^{n\times n}$ invertible y definimos $s = a^\top a$ que es simétrica, luego $\omega(x, y) = x^\top s y$ es bilineal simétrica. Además

$$\omega(x, x) = x^\top a^\top a x = (ax)^\top (ax) = \sum_{i=1}^{n} \left((ax)^i\right)^2 \geq 0.$$

Como a es invertible, $x \neq 0$ implica $ax \neq 0$ y $\omega(x, x) > 0$.

2. Sea $\mathbb{E}$ un espacio de funciones definidas en un intervalo $[a, b]$, al menos continuas, de modo que se pueden integrar. La definición es

$$\omega(f(t), g(t)) = \int_a^b f(t)\, g(t)\, dt.$$

Si $f(x) \neq 0$ (no es la función 0), la función $f^2(x)$ es ≥ 0 y estrictamente positiva al menos en un pequeño intervalo, luego, por propiedades básicas de la integral, $\omega(f(x), f(x)) > 0$.

3. En $\mathbb{E}$ de dimensión n tenemos n formas lineales *independientes* $f^1 \ldots, f^n$. Recordamos que $L : \mathbb{E} \to \mathbb{R}^n$ definido por $L(x) = \left(f^1(x), \ldots, f^n(x)\right)^\top$ es inyectiva porque la matriz $\left(f_j^i\right)$ de las formas en cualquier base tiene rango n. Con las formas definimos la forma $\omega(x, y) = f^1(x) f^1(y) + \ldots. + f^n(x) f^n(y)$. Claramente es bilineal simétrica y si $x \neq 0$ es $L(x) \neq 0$ y $\omega(x, y) = \left(f^1(x)\right)^2 + \ldots. + \left(f^n(x)\right)^2 > 0$. Claramente ω es un producto euclidiano.

4. Se toma $\mathbb{E} = \mathbb{R}^{n\times n}$ y se define $\omega(a, b) = \operatorname{tr}\left(a^\top b\right)$. Es fácil ver que $\omega(a, b) = \sum_{i,j=1}^{n} a_j^i b_j^i$, luego $\omega(a, a) = 0$ equivale a $\sum \left(a_j^i\right)^2 = 0$ y $a = 0$.

Hay algunos ejemplos derivados de estos que irán apareciendo en problemas. Primero vamos a aprovechar el trabajo del capítulo anterior.

[4] Suele limitarse el teorema de Pitágoras a que si el triángulo es rectángulo se tiene que $a^2 = b^2 + c^2$. Sin embargo es igual de importante el recíproco; esto es, que si $a^2 = b^2 + c^2$, el triángulo es rectángulo. Si en una cuerda de 12 metros se hace un nudo cada metro y se forma un triángulo con ella de lados $3, 4, 5$, los dos lados más cortos señalan un ángulo recto. Dicen que lo usaban los constructores del antiguo Egipto.

9.1.2. Bases ortogonales y ortonormales

Teorema 167 *Todo espacio euclidiano de dimensión finita tiene bases ortonormales.*

Demostración. Como ω es bilineal simétrica por el teorema de diagonalización y el de Sylvester (teoremas 159 y 161) existirá una base $(u_1,\ldots,u_n)$ con $\omega(u_i,u_i)=1$ y $\omega(u_i,u_j)=0$ para $i\neq j$, que se suele sintetizar en $\omega(u_i,u_j)=\delta_{ij}$ con las deltas de Kronecker. Esto es como decir que $(u_1,\ldots,u_n)$ es ortonormal. ♣

Advertimos que habrá otro procedimiento para conseguir estas bases: el de Gram-Schmidt.[5] Las bases ortogonales y ortonormales son muy útiles pues en ellas se hacen con facilidad muchos cálculos.

Teorema 168 *En una base ortogonal $\mathcal{V}$ u ortonormal $\mathcal{W}$ se tienen las expresiones en coordenadas para $x\in\mathbb{E}$,*

$$x=\sum_{i=1}^{n}\frac{\langle x,v_i\rangle}{\|v_i\|^2}v_i=\sum_{i=1}^{n}\langle x,w_i\rangle w_i,\quad \langle x,y\rangle=\sum_{i=1}^{n}\langle x,w_i\rangle\langle y,w_i\rangle \ \textbf{\textit{(identidad de Parseval).}}$$

Si L es un endomorfismo de $\mathbb{E}$, la matriz de L en una base ortonormal $\mathcal{W}$ está dada por las condiciones equivalentes

$$L(w_j)=\sum_{i=1}^{n}\langle w_i,L(w_j)\rangle w_i,\quad L_j^i=\langle w_i,L(w_j)\rangle.$$

Demostración. Sea $x=\sum_{i=1}^{n}x^iv_i$. Multiplicando por v_j,

$$\langle x,v_j\rangle=\left\langle\sum_{i=1}^{n}x^iv_i,v_j\right\rangle=\sum_{i=1}^{n}x^i\langle v_i.v_j\rangle=x^j\langle v_j,v_j\rangle=x^j\|v_j\|^2$$

puesto que $\langle v_i.v_j\rangle=0$ cuando $i\neq j$. De ahí sale $x^j=\langle x,v_j\rangle/\|v_j\|^2$ y la fórmula primera.

Si $L(w_j)=\sum_{k=1}^{n}L_j^kw_k$, al aplicar $\langle w_i,\bullet\rangle$ queda

$$\langle w_i,L(w_j)\rangle=\left\langle w_i,\sum_{k=1}^{n}L_j^kw_k\right\rangle=\sum_{k=1}^{n}L_j^k\langle w_i,w_k\rangle=\sum_{k=1}^{n}L_j^k\delta_{ik}=L_j^i.$$

♣

Problema 318 *En $\mathbb{R}^2$ definimos ω por*

$$\omega(x,y)=\left(x^1,x^2\right)\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}y^1\\y^2\end{pmatrix}.$$

Probar que es un producto euclidiano, calcular el módulo de $u=(1,1)$, dar otro vector $v\neq0$ perpendicular a $(1,1)$. Construir una base ortogonal en donde esté u y dar una base ortonormal. ♦

Solución. Si al reducir la matriz a la forma diagonal aparecen dos coeficientes >0, entonces ω será definida positiva. Con las técnicas del capítulo anterior

$$\left(\begin{array}{cc}1&1\\\hline 1&2\end{array}\right)\to\left(\begin{array}{cc}1&1-1\\\hline 1&2-2\end{array}\right)=\left(\begin{array}{cc}1&0\\\hline 1&0\end{array}\right)\Rightarrow\left(\begin{array}{cc}1&0\\\hline 1&-1\end{array}\right)\to\left(\begin{array}{cc}1&0\\\hline 1&-1\\0&1\end{array}\right).$$

La base $\mathcal{W}=\left((1,0)^\top,(-1,1)^\top\right)$ es ortonormal. Por ejemplo,

$$(1,0)\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}=0,\quad(-1,1)\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}-1\\1\end{pmatrix}=1.$$

[5] Este, en cierto sentido difícil de precisar ahora, es el mismo que el de los teoremas 159 y 161).

Se tiene

$$\left\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\|^2 = (1,1) \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5,$$

luego u no es unitario y tiene $\|u\| = \sqrt{5}$. Para completar u a una base ortogonal basta buscar (x, y) tal que

$$(1,1) \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2x + 3y = 0$$

y sirve, por ejemplo, $v = (3, -2)$. Las normas de u y v son $\sqrt{5}$ y $\sqrt{13}$ y dividiendo por ellas a u y v obtenemos una base ortonormal. ♦

Problema 319 *En $\mathbb{R}^2$ se toma ω definido por la matriz*

$$s = \begin{pmatrix} 1 & p \\ p & q \end{pmatrix}.$$

Probar que es un producto euclidiano si y solo si $\Delta = \det(s) = q - p^2 > 0$. ¿Puede ser en algún caso la base estándar base ortonormal? ¿Puede ser ω un producto euclidiano si $p < 0$? ¿Y si es $q \leq 0$?

Problema 320 *Dos productos euclidianos ω y η en $\mathbb{R}^3$ tienen en la base estándar matrices*

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \quad y \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 4 \end{pmatrix}.$$

Determinar bases ortonormales y, en ellas, las terceras coordenadas de $z = (2,2,2)^\top$. ♦

Solución (del primer caso). Se hacen las operaciones

$$\left(\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 1 & 1-1 & 0 \\ 1 & 2-1 & 1 \\ 0 & 1-0 & 3 \\ \hline 1 & 0-1 & 0 \\ 0 & 1-0 & 0 \\ 0 & 0-0 & 1 \end{array}\right) = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 3 \\ \hline 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 3 \\ \hline 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

$$\to \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 1-1 \\ 0 & 1 & 3-1 \\ \hline 1 & -1 & 0-(-1) \\ 0 & 1 & 0-1 \\ 0 & 0 & 1-0 \end{array}\right) = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ \hline 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ \hline 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{array}\right)$$

Vemos que $\mathcal{W} = \left((1,0,0)^\top, (-1,1,0)^\top, (1,-1,1)^\top\right)$ es base ortogonal pero no ortonormal porque

$$\mathrm{mat}_{\mathcal{W}\mathcal{W}}(\omega) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Sin embargo si se modifica $\mathcal{W}$ a $\left(w_1, w_2, \left(1/\sqrt{2}\right) w_3\right)$ sí que tenemos una base ortonormal. Entonces

$$\omega\left(\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right) = \frac{1}{\sqrt{2}} (2,2,2) \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = 2\sqrt{2}$$

es la tercera coordenada de z en $\mathcal{W}$. ♦

Problema 321 *Consideramos el espacio $\mathbb{E}$ de las funciones de la forma $f(x) = \alpha \cos x + \beta \sin x$, $\alpha, \beta \in \mathbb{R}$ con el producto euclidiano $\omega(f(x), g(x)) = \int_0^\pi f(x) g(x)\, dx$. Decir si $(\cos x, \sin x)$ es base ortogonal u ortonormal y si no, encontrar una. Hacer el problema análogo con $\mathbb{E}$ los polinomios de grado ≤ 1, límites de integración $(0,1)$ y la base estándar.*

Problema 322 *En* $\mathbb{E}$ *de dimensión* 2 *se toman dos formas lineales* no nulas $f, g : \mathbb{E} \to \mathbb{R}$ *y se define*

$$\omega : \mathbb{E} \times \mathbb{E} \to \mathbb{R}, \quad \omega(x, y) = (f \otimes g + g \otimes f)(x, y) = f(x) g(y) + g(x) f(y).$$

¿Es un producto euclidiano? Si lo fuera, calcular una base ortonormal.

En los ejemplos de productos euclidianos hemos dado dos a lor que volvemos: **(a)** Tomamos $a \in \mathbb{R}^{n\times n}$ invertible y el producto euclidiano en $\mathbb{R}^n$, $\omega_1(x, y) = x^\top a^\top a y$; y **(b)** tomamos $f^1, \ldots, f^n \in \mathbb{E}^*$ *independientes* y el producto euclidiano en $\mathbb{E}$ dado por $\omega_2(x, y) = \sum_{h=1}^n f^h(x) f^h(y)$. El problema que sigue ayuda en las fórmulas de puro cálculo.

Problema 323 *Probar que en el caso de* $\omega_1(x, y) = x^\top a^\top a y$ *se tiene como fórmula alternativa* $\omega_1(x, y) = \varepsilon(ax, ay)$ *con* ε *el producto estándar de* $\mathbb{R}^n$ *y que la matriz de* ω_1 *en* $\mathcal{E}$ *es* $a^\top a$

Probar en el caso de ω_2 *que si* $\mathcal{U}$ *es una base y en su base dual* $\mathcal{U}^*$ *tenemos* $f^i = a_1^i u^1 + \ldots + a_n^i u^n$ *(de modo equivalente,* $\left(a_1^i, \ldots, a_n^i\right) = \mathrm{mat}_{\mathcal{U}}\left(f^i\right)$ *) se cumple entonces que*

$$\omega_2(x, y) = \sum_{h=1}^n f^h(x) f^h(y) = \left(x^1, \ldots, x^n\right) a^\top a \left(y^1, \ldots, y^n\right)^\top$$

siendo x^i, y^j *las coordenadas de* $x, y \in \mathbb{E}$ *en* $\mathcal{U}$. *Como consecuencia* $\mathrm{mat}_{\mathcal{U}\mathcal{U}}(\omega_2) = a^\top a$.

Problema 324 *Sea* $\mathbb{E} = \lg(\cos t, \sin t)$, *subespacio*[6] *del espacio de las funciones continuas de* $[0, 1]$ *en* $\mathbb{R}$. *Las funciones* f^1 *y* f^2 *serán* $f^1(x) = x'(\pi)$ *(la derivada de la* función x *en* $t = \pi$*) y* $f^2(x) = \int_0^1 x(t)\, dt$. *Calcular* $\omega(v, v)$ *para* $v(t) = \cos t - \sin t$.

Un producto euclidiano ω tiene ciertas propiedades elementales de uso continuo. En primer lugar, al escribir la matriz $s = \mathrm{mat}_{\mathcal{U}\mathcal{U}}(\omega)$ en cualquier base, no puede darse $s_{ii} \leq 0$ porque sería $\omega(u_i, u_i) = s_{ii} \leq 0$ con $u_i \neq 0$ y ω es definida positiva. Sí que puede ser $s_{ij} \leq 0$ cuando $i \neq j$. Otra cuestión es que si $\mathbb{F}$ es un subespacio de $\mathbb{E}$ y $\omega_{\mathbb{F}} : \mathbb{F} \times \mathbb{F} \to \mathbb{R}$ es la **restricción** de ω a $\mathbb{F} \times \mathbb{F}$, se cumple que $\omega_{\mathbb{F}}$ es también un producto euclidiano. Así pues, $(\mathbb{F}, \omega_{\mathbb{F}})$ es otro espacio euclidiano. Al hablar simplemente del **subespacio euclidiano** $\mathbb{F}$, entenderemos que el producto es $\omega_{\mathbb{F}}$ y es frecuente escribir el subíndice $\mathbb{F}$ tan solo si conviene por énfasis o claridad.

Otra cuestión es que ω es no degenerada. Recordemos que esto significa que si $x \in \mathbb{E}$ verifica que $\omega(x, y) = 0$ para todo y, tiene que ser $x = 0$. De modo equivalente, si $x \neq 0$ debe existir otro $y \in \mathbb{E}$ tal que $\omega(x, y) \neq 0$. Un producto euclidiano es no degenerado porque si $x \neq 0$ y tomamos $y = x$ obtenemos $\omega(x, x) > 0$. Las formas no degeneradas tienen rango $n = \dim(\mathbb{E})$ y su matriz en cualquier base es invertible, así que los productos euclidianos tienen estas propiedades. Una forma no degenerada σ, y en particular un producto euclidiano ω, cumplen que si $\sigma(x, y) = 0$ para todo y, tiene que ser $x = 0$. Esto es útil para probar que $x_1 = x_2$ (suelen ser los dos lados de una ecuación) verificando que para y arbitrario $\sigma(x_1, y) = \sigma(x_2, y)$.

Problema 325 *En* $\mathbb{R}^2$ *se toma* σ *con matriz*

$$s = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

en la base estándar. Probar que σ *es no degenerada pero no es un producto euclidiano. Encontrar vectores* $u \neq 0$ *tales que* $\sigma(u, u) = 0$ *y subespacios* $\mathbb{F}$ *de* $\mathbb{E} = \mathbb{R}^2$ *tales que la restricción* $\bar{\sigma} = \sigma_{\mathbb{F}}$ *de* σ *a* $\mathbb{F}$ *no es no degenerada.*[7]

Una forma no degenerada, como σ en el problema 325 muestra la diferencia que separa a un producto euclidiano de estas formas. El aspecto más relevante es la existencia de vectores $u \neq 0$ tales que $\sigma(u, u) = 0$, que se llaman **vectores isótropos**, y no existen para los espacios euclidianos. Representarían vectores no nulos de longitud cero o vectores perpendiculares a si mismos.

Problema 326 *Preguntamos si* ω *es un producto euclidiano en los siguientes casos*

1. $\mathbb{E} = \mathbb{R}_n[X]$ *y* $\omega(P(X), Q(X)) = P(0) Q(0) + P(1) Q(1) + \ldots + P(n) Q(n)$.
2. $\mathbb{E}$ *es el espacio de las funciones continuas de* $[0, 1]$ *en* $\mathbb{R}$ *y* $\omega(f(t), g(t)) = \int_0^{1/2} f(t) g(t)\, dt$.

[6] Puede usarse que $\mathcal{B} = (\cos t, \sin t)$ es base de $\mathbb{E}$.

[7] Sintácticamente es más correcto decir que $\sigma_{\mathbb{F}}$ es degenerada, pero queremos subrayar que el ser no degenerada es una propiedad que se pierde al restringirse a subespacios.

9.1.3. Matrices y determinantes de Gram

Si $(x_1,\ldots,x_r)$ es una sucesión de vectores de $\mathbb{E}$ se define su **matriz de Gram** por

$$G(x_1,\ldots,x_r)=\begin{pmatrix} \langle x_1,x_1\rangle & \langle x_1,x_2\rangle & \cdots & \langle x_1,x_{r-1}\rangle & \langle x_1,x_r\rangle \\ \langle x_2,x_1\rangle & \langle x_2,x_2\rangle & \cdots & \langle x_2,x_{r-1}\rangle & \langle x_2,x_r\rangle \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \langle x_{r-1},x_1\rangle & \langle x_{r-1},x_2\rangle & \cdots & \langle x_{r-1},x_{r-1}\rangle & \langle x_{r-1},x_r\rangle \\ \langle x_2,x_1\rangle & \langle x_r,x_2\rangle & \cdots & \langle x_r,x_{r-1}\rangle & \langle x_r,x_r\rangle \end{pmatrix}$$

y $\Gamma(x_1,\ldots,x_r)=\det(G(x_1,\ldots,x_r))$ es el **determinante de Gram**. Si tomamos una base $\mathcal{U}=(u_1,\ldots,u_n)$, se tiene que $G(u_1,\ldots,u_n)=\mathrm{mat}_{\mathcal{UU}}(\omega)$ y es por tanto una matriz invertible. Si $(x_1,\ldots,x_r)$ es independiente y $\mathbb{F}$ es el subespacio que genera, la restricción $\omega_{\mathbb{F}}$, como ya dijimos, es un producto euclidiano luego $G(x_1,\ldots,x_r)$ es invertible.

Problema 327 *Si $\mathcal{U}=(u_1,\ldots,u_n)$ es base ortonormal y $x_1,\ldots,x_r$ se expresan como $x_j=\sum_{i=1}^n a_j^i u_i$, la matriz de Gram es $G(x_1,\ldots,x_r)=a^\top a$.*

Teorema 169 *Las matrices y determinantes de Gram tienen las siguientes propiedades:*

1. *$G(x_1,\ldots,x_r)$ es invertible (que equivale a $\Gamma(x_1,\ldots,x_r)\neq 0$) si y solo si $(x_1,\ldots,x_r)$ es independiente. En particular, si $\mathcal{U}=(u_1,\ldots,u_n)$ es una base de $\mathbb{E}$, cualquier submatriz $G\left(u_{i_1},\ldots,u_{i_p}\right)$ de $G(u_1,\ldots,u_n)$ ha de ser invertible.*
2. *$\Gamma(x_1,\ldots,x_r)\geq 0$ (y, por **1**, nulo justamente si los vectores son dependientes).*

Demostración. Más arriba se probó que si $(x_1,\ldots,x_r)$ es independiente debe ser $G(x_1,\ldots,x_r)$ invertible. Para la recíproca basta ver que si $(x_1,\ldots,x_r)$ es dependiente debe ser $G(x_1,\ldots,x_r)$ no invertible. Esto es fácil porque una fila o columna de $G(x_1,\ldots,x_r)$ es combinación de las otras. Por ejemplo, si $x_1=\lambda_1x_2+\ldots+\lambda_rx_r$ se tiene $\langle x_1,x_j\rangle=\lambda_1\langle x_2,x_j\rangle+\ldots+\lambda_r\langle x_r,x_j\rangle$ y la fila 1 es combinación de las restantes.

Si $(x_1,\ldots,x_r)$ es dependiente, $G(x_1,\ldots,x_r)=0$ (ya probado). Si $(x_1,\ldots,x_r)$ es independiente se considera $\mathbb{F}=\lg(x_1,\ldots,x_r)$ y una base ortonormal $(u_1,\ldots,u_r)$ de $\mathbb{F}$. El problema 327 nos da $G(x_1,\ldots,x_r)=a^\top a$ y entonces $\Gamma(x_1,\ldots,x_r)=\det\left(a^\top a\right)=\det(a)^2\geq 0$. Como a es matriz que expresa una base en función de otra, a es invertible, y debe ser $\Gamma(x_1,\ldots,x_r)>0$. ♣

La condición **1** da condiciones *necesarias* para saber a partir de la matriz s simétrica de σ en una base cualquiera si σ es un producto euclidiano. Por ejemplo, consideremos

$$s=\begin{pmatrix} 1 & h & 1 \\ h & 2 & 0 \\ 1 & 0 & h \end{pmatrix},\quad h\in\mathbb{R}.$$

Hay valores de h para los que podemos asegurar que σ *no es* un producto euclidiano. Por ejemplo, la primera submatriz 2×2 tiene determinante $2-h^2$ y si $2-h^2\leq 0$, σ no es un producto euclidiano. Se necesita $|h|<\sqrt{2}$. Podemos tantear si vale $h=1$ pero $\det(s)=2h-h^3-2$ es vale -1, luego $h=1$ tampoco da un producto euclidiano aunque sea $1<\sqrt{2}$. Mas adelante, el criterio de Sylvester nos permitirá determinar *de forma sistemática* todos los valores de h con los que se tiene un producto euclidiano.

Hay una vía más directa[8] para probar que $(x_1,\ldots,x_r)$ dependiente equivale a $\Gamma(x_1,\ldots,x_r)=0$. Si hay dependencia, existen $\lambda^1,\ldots,\lambda^r\in\mathbb{R}$ no todos nulos tales que $0=\lambda^1x_1+\ldots+\lambda^rx_r$. Multiplicando por $x_1,\ldots,x_r$ tenemos el sistema lineal

$$\begin{cases} 0=\lambda^1\langle x_1,x_1\rangle+\ldots+\lambda^r\langle x_1,x_r\rangle \\ \qquad\qquad\vdots\vdots \\ 0=\lambda^1\langle x_r,x_1\rangle+\ldots+\lambda^r\langle x_r,x_r\rangle \end{cases}.$$

[8] En el teorema 167 probamos que si $(x_1,\ldots,x_r)$ era independiente se tenía $\Gamma(x_1,\ldots,x_r)>0$ utilizando que el espacio $\mathbb{F}=\lg(x_1,\ldots,x_r)$ tenía una base ortonormal. Ciertos textos no disponen de ese resultado inicialmente y esperan a desarrollar el procedimiento de Gram-Schmidt. En ese caso, el problema que sigue es mucho más directo.

Problema 328 *Relacionando la invertibilidad de la matriz de un sistema lineal con la existencia de soluciones no nulas, probar que* $(x_1,\ldots,x_r)$ *dependiente equivale a* $\Gamma(x_1,\ldots,x_r)=0$.

Problema 329 *Consideramos* $G(x_1,\ldots,x_r)$. *Supongamos que para un* $q<r$ *se cumple* $\Gamma(x_1,\ldots,x_q)=0$. *Probar que entonces* $\Gamma(x_1,\ldots,x_r)=0$*?* [9] *Si tomamos* $1\leq i_1<i_2<\ldots<i_q\leq r$ *y* $\Gamma(x_{i_1},\ldots,x_{i_r})=0$, *¿se cumple* $\Gamma(x_1,\ldots,x_r)=0$*?*

Las matrices de Gram tienen aplicaciones curiosas. Sea $\mathbb{E}=\mathcal{C}([0,1],\mathbb{R})$ el espacio de funciones continuas de $[0,1]$ en $\mathbb{R}$. Nos dar n funciones $x_1,\ldots,x_n$ de $\mathbb{E}$ y queremos saber si son independientes; o sea, si existen $\lambda^1,\ldots,\lambda^n\in\mathbb{R}$ tales que $\sum_{i=1}^n\lambda^i x_i(t)=0$ para todo $t\in[0,1]$. Si tomamos un producto euclidiano *cualquiera* en $\mathbb{E}$, la dependencia equivale a $\Gamma(x_1,\ldots,x_n)=0$. Puede ser más complejo verificar que $\Gamma(x_1,\ldots,x_n)=0$ que la dependencia, pero no deja de ser una herramienta, que puede compararse con el **Wronskiano**.

Problema 330 *Estudiar la independencia de las funciones* $\left(\cos t,\sin t,\cos^2 t,\sin^2 t\right)$.

Problema 331 *Hacer el problema análogo para* $\left(\cos t-\sin t,\cos^2 t-\sin^2 t\right)$

Solución. Solo hay que calcular $\int_{-\pi}^{\pi}(\cos t-\sin t)\left(\cos^2 t-\sin^2 t\right)dt=0$ y la matriz de Gram, será 2×2 y diagonal. Como el producto euclidiano es definido positivo, los términos de la diagonal (que no hay que calcular) serán no nulos y no nulo el determinante de Gram será no nulo, Por consiguiente hay independencia. ♠

Hay una fórmula sencilla, muy similar a las fórmulas de las matrices y determinantes de Gram, que luego usaremos en el producto vectorial. Puede generalizarse a espacios euclidianos abstractos, pero nos limitamos al caso $(\mathbb{R}^n,\varepsilon)$.

Teorema 170 (identidad de Gram) *Sean* $(x_1,\ldots,x_n)$ *e* $(y_1,\ldots,y_n)$ *sucesiones en* $\mathbb{R}^n$, *y* $x,y\in\mathbb{R}^{n\times n}$ *las matrices obtenidas yuxtaponiendo columnas. Entonces,* $\det(x)\det(y)=\det\left(\langle x_i,y_j\rangle\right)_{1\leq i,j\leq n}$.

Demostración. Sea M la matriz de los $\langle x_i,y_j\rangle$. Como $\langle\bullet,\bullet\rangle=\varepsilon$ el producto euclidiano estándar en $\mathbb{R}^n$, tenemos

$$\langle x_i,y_j\rangle=\sum_{h=1}^n x_i^h y_j^h=\sum_{h=1}^n\left(x^\top\right)_h^i y_j^h=\left(x^\top y\right)_j^i;\ \text{o sea, } M=x^\top y.$$

Entonces, $\det(M)=\det\left(x^\top y\right)=\det\left(x^\top\right)\det(y)=\det(x)\det(y)$. ♣

Si $(x_1,\ldots,x_n)=(y_1,\ldots,y_n)$ tenemos como caso particular que $\det(x)^2=\Gamma(x_1,\ldots,x_n)$.

9.1.4. Matrices ortogonales

Estudiamos un tipo muy importante de matrices: las **matrices ortogonales**. ¡Ojo! no hay matrices *ortonormales* aunque veremos que estas matrices ortogonales tienen una estrecha relación con las bases *ortonormales*. Diremos que $a\in\mathbb{R}^{n\times n}$ es matriz ortogonal si $a^\top a=I=aa^\top$ (hay parte superflua porque $a^\top a=I$ implica $I=aa^\top$). Se suele decir con palabras que a es ortogonal si su inversa es su traspuesta. Hay que recordar porque se usa con mucha frecuencia el teorema

Teorema 171 *Las matrices ortonormales tienen la propiedades siguientes*

1. *Son invertibles y tienen determinante* ±1.
2. *El que* $a\in\mathbb{R}^{n\times m}$ *sea* ortogonal *equivale a que la sucesión* $(a_1,\ldots,a_n)$ *de sus columnas sea una familia* ortonormal *de* $\mathbb{R}^n$ *con el producto estándar* ε. *Otro tanto es cierto sustituyendo columnas por filas.*
3. *La matriz unidad es ortogonal, el producto de dos matrices ortogonales es ortogonal y la inversa de una matriz ortogonal es ortogonal. Por tanto, las matrices ortogonales forman un grupo (no conmutativo) llamado* ***grupo ortogonal****.*

[9] Si para $a\in\Bbbk^{n\times n}$ se tiene que el determinante de la submatriz $b\in\Bbbk^{q\times q}$ formada por las primeras q filas y columnas de a cumple $\det(b)=0$, puede darse que $\det(a)\neq0$.

Problema 332 *Demostrar estas propiedades de las matrices ortogonales.*

Teorema 172 *Sean $\mathcal{U} = (u_1, \ldots, u_n)$ una base ortonormal de $\mathbb{E}$ y $\mathcal{V} = (v_1, \ldots, v_n)$ otra base, relacionadas entre sí por la matriz de cambio $c \in \mathbb{R}^{n\times n}$; es decir, $v_j = \sum_{i=1}^n c_j^i u_i$, $1 \leq j \leq n$. Para que $\mathcal{V}$ sea base ortonormal, es necesario y suficiente que c sea ortogonal.*

Demostración. Si $\mathcal{V}$ es ortonormal, con la identidad de Parseval,

$$\delta_{pq} = \langle v_p, v_q\rangle = \left\langle \sum_{i=1}^n c_p^i u_i, \sum_{j=1}^n c_q^j u_j \right\rangle = \sum_{i=1}^n c_p^i c_q^i = \sum_{i=1}^n \left(c^\top\right)_i^q c_q^i = \left(c^\top c\right)_p^q,$$

que dice en coordenadas que $c^\top c = I$. Recíprocamente, $c^\top c = I$ nos da

$$\delta_{pq} = \left(c_i^\top c\right)_p^q = \sum_{i=1}^n \left(c^\top\right)_i^q c_q^i = \sum_{i=1}^n c_p^i c_q^i = \left\langle \sum_{i=1}^n c_p^i u_i, \sum_{j=1}^n c_q^j u_j \right\rangle = \langle v_p, v_q\rangle$$

(como se ve es lo anterior pero las implicaciones van en otro orden) y $\mathcal{V}$ es base ortonormal. ♣

Problema 333 *Probar que las matrices ortogonales 2×2 son las de la forma*

$$a = \begin{pmatrix} p & -Dq \\ q & Dp \end{pmatrix}, \quad D = \det(a)$$

con $D = \pm 1$ y con p, q tales que $p^2 + q^2 = 1$. Lo usual es encontrar estas matrices escritas con $p = \cos\theta$ y $q = \sin\theta$.

Problema 334 *Probar que si a es a la vez ortogonal y triangular superior es entonces diagonal con solo ± 1 en ella.*

9.2. Un primer contacto con el producto vectorial

En toda esta subsección trabajamos en $\mathbb{E} = \mathbb{R}^3$ con el producto estándar ε. Hay una fórmula rápida para poder asignar al par de vectores (u, v) de $\mathbb{R}^3$ un tercer vector w ortogonal a u y v. Definimos el **producto vectorial de** u y v, que denotaremos por $u \times v$ por

$$u \times v = \begin{vmatrix} u^1 & v^1 & e_1 \\ u^2 & v^2 & e_2 \\ u^3 & v^3 & e_3 \end{vmatrix} = \begin{vmatrix} u^2 & v^2 \\ u^3 & v^3 \end{vmatrix} e_1 - \begin{vmatrix} u^1 & v^1 \\ u^3 & v^3 \end{vmatrix} e_2 + \begin{vmatrix} u^1 & v^1 \\ u^2 & v^2 \end{vmatrix} e_3.$$

siendo (e_1, e_2, e_3) la base estándar de $\mathbb{R}^3$. Hemos dado una definición y una regla mnemotécnica. La definición es la segunda. El "determinante" 3×3 no tiene sentido estrictamente hablando, porque en la tercera columna no hay números sino vectores. Queremos decir que si se desarrolla formalmente el determinante por la tercera columna se obtiene $u \times v$. Es muy sencillo comprobar que

1. El producto vectorial es antisimétrico y lineal en cada variable; o sea $u \times v = -v \times u$ y

   $$(u + u') \times v = u \times v + u' \times v, \quad u \times (v + v') = u \times v + u \times v' \quad \lambda(u \times v) = (\lambda u) \times v = u \times (\lambda v).$$

2. Dados tres vectores u, v, w se cumple que

   $$\langle u \times v, w\rangle = \det(u, v, w) = \begin{vmatrix} u^1 & v^1 & w^1 \\ u^2 & v^2 & w^2 \\ u^3 & v^3 & w^3 \end{vmatrix}. \tag{9.1}$$

3. El producto vectorial es perpendicular a cada factor porque $\langle u, u \times v\rangle = \langle v, u \times v\rangle = 0$.[10]

[10] Cualquier tratamiento, por elemental que sea del producto vectorial, se ve sobre todo como un modo de asignar a cada par de vectores otro perpendicular a ambos.

4. Se verifica que $u \times v \neq 0$ si y solo si (u, v) es independiente. De hecho, si (u, v) es independiente $(u, v, u \times v)$ es base de $\mathbb{R}^3$.

Esto último se verifica con (9.1). Si (u, v) es independiente, se completa (u, v) a una base (u, v, w). Entonces, $\det(u, v, w) \neq 0$, pero $\det(u, v, w) = \langle u \times v, w \rangle$, y esto implica $u \times v \neq 0$. Recíprocamente, si $u \times v \neq 0$, (9.1) implica ahora que

$$\det(u, v, u \times v) = \langle u \times v, u \times v \rangle = \|u \times v\|^2 > 0,$$

luego $(u, v, u \times v)$ es una base de $\mathbb{R}^3$. En particular, (u, v) es independiente. Estas fórmulas y alguna más que luego veremos sobre $\|u \times v\|$ se usan con gran frecuencia.

Queremos entrar ahora en fórmulas más complejas como el **producto escalar de productos vectoriales** o el **doble producto vectorial**. En estas fórmulas complejas es ventajoso recordar que $x = y$ si y solo si para todo $z \in \mathbb{E}$ es $\langle x, z \rangle = \langle y, z \rangle$ (en nuestro caso será $\mathbb{R}^3$ con $\varepsilon = \langle \bullet, \bullet \rangle$). Parece que esto va a complicar todo pero no es así porque si se quiere probar que $u \times v = w$, tendríamos que probar que $\langle u \times v, z \rangle = \langle w, z \rangle$ pero $\langle u \times v, z \rangle = \det(u, v, z)$ lo que facilita los cálculos más de lo que pudiera parecer. Reescribimos la identidad de Gram del teorema 170 para $\left(\mathbb{R}^3, \varepsilon\right)$

$$\det(a, b, c) \det(u, v, w) = \begin{vmatrix} \langle a, u \rangle & \langle a, v \rangle & \langle a, w \rangle \\ \langle b, u \rangle & \langle b, v \rangle & \langle b, w \rangle \\ \langle c, u \rangle & \langle c, v \rangle & \langle c, w \rangle \end{vmatrix}.$$

Teorema 173 (producto escalar de productos vectoriales) *Sean $a, b, u, v \in \mathbb{R}^3$. Tenemos que*

$$\langle a \times b, u \times v \rangle = \begin{vmatrix} \langle a, u \rangle & \langle a, v \rangle \\ \langle b, u \rangle & \langle b, v \rangle \end{vmatrix}.$$

En particular, $\|u \times v\|^2 = \|u\|^2 \left\|v^2\right\| - \langle u, v \rangle^2 = \Gamma(u, v)$.

Demostración. Si $u = \lambda v$ se ve enseguida que la fórmula es cierta, pues sale $0 = 0$. Descartemos ese caso. Entonces (u, v) es independiente y $w = u \times v \neq 0$. Para $c = w = u \times v$ la identidad de Gram es

$$\begin{vmatrix} \langle a, u \rangle & \langle a, v \rangle & \langle a, w \rangle \\ \langle b, u \rangle & \langle b, v \rangle & \langle b, w \rangle \\ 0 & 0 & \|u \times v\|^2 \end{vmatrix} = \det(a, b, u \times v) \det(u, v, u \times v).$$

Los lados izquierdo y derecho son respectivamente

$$\|u \times v\|^2 \begin{vmatrix} \langle a, u \rangle & \langle a, v \rangle \\ \langle b, u \rangle & \langle b, v \rangle \end{vmatrix} \quad \text{y} \quad \langle a \times b, u \times v \rangle \langle u \times v, u \times v \rangle = \langle a \times b, u \times v \rangle \|u \times v\|^2,$$

utilizándose (9.1) para el lado derecho. Cancelamos $\|u \times v\|^2 \neq 0$ y obtenemos el teorema. ♣

Es muy sencillo probar que la base estándar cumple $e_i \times e_j = e_k$ si (i, j, k) es permutación circular de $(1, 2, 3)$. (Las permutaciones circulares de $(1, 2, 3)$ son, de las seis permutaciones de $\{1, 2, 3\}$, las que llevan $(1, 2, 3)$ a $(1, 2, 3)$, $(2, 3, 1)$ y $(3, 1, 2)$.) Si $\mathcal{U} = (u_1, u_2, u_3)$, ¿seguirá esto siendo cierto? Si hacemos $u_1 \times u_2 = v$ sabemos que v es ortogonal a u_1 y u_2 y que, por el teorema 173, se tiene que $\|v\| = 1$. Esto lleva a $u_1 \times u_2 = v = \lambda u_3$ con $\lambda = \pm 1$. Sin embargo, *no puede deducirse* que sea $\lambda = 1$ y $u_1 \times u_2 = u_3$. En efecto, si se toma $\mathcal{U} = (-e_1, e_2, e_3)$ se sigue teniendo una base ortonormal pero $u_1 \times u_2 = -(e_1 \times e_2) = -e_3 \neq u_3$. Esto es un aviso de que *no todas las bases ortonormales son "buenas" relativas al producto vectorial*. Solo podemos decir que $u_1 \times u_2 = u_3$ equivale a que sea

$$\det(u_1, u_2, u_3) = \begin{vmatrix} u_1^1 & u_2^1 & u_3^1 \\ u_1^2 & u_2^2 & u_3^2 \\ u_1^3 & u_2^3 & u_3^3 \end{vmatrix} = 1.$$

Esto lleva a distinguir entre las bases $\mathcal{U}$ de $\mathbb{R}^n$ ortonormales con $\det(u_1, \ldots, u_n) = +1$ y $\det(u_1, \ldots, u_n) = -1$. La cuestión es de mucho calado y se tratará en la próxima sección.

Teorema 174 (doble producto vectorial) *Para $a, b, c \in \mathbb{R}^3$ se tiene que*[11]

$$a \times (b \times c) = \langle a, c\rangle\, b - \langle a, b\rangle\, c \quad y \quad (a \times b) \times c = \langle a, c\rangle\, b - \langle b, c\rangle\, a$$

En particular, el producto vectorial no es *asociativo.*

Demostración. Aplicaremos que $u = v \iff \forall z,\ \langle u, z\rangle = \langle v, z\rangle$. Con esto y el teorema 173 en $\overset{*}{=}$,

$$\begin{aligned}\langle a \times (b \times c), z\rangle &= \det(a, (b \times c), z) = \det(z, a, b \times c) = \langle z \times a, b \times c\rangle \\ &= \langle a \times z, c \times b\rangle \overset{*}{=} \begin{vmatrix} \langle a, c\rangle & \langle a, b\rangle \\ \langle z, c\rangle & \langle z, b\rangle \end{vmatrix}.\end{aligned}$$

Por otra parte, con el teorema 173,

$$\langle \langle a, c\rangle\, b - \langle a, b\rangle\, c, z\rangle = \langle a, c\rangle \langle b, z\rangle - \langle a, b\rangle \langle c, z\rangle = \begin{vmatrix} \langle a, c\rangle & \langle a, b\rangle \\ \langle c, z\rangle & \langle b, z\rangle \end{vmatrix}$$

La primera fórmula se sigue porque los determinantes son iguales. La otra queda para el lector. ♣

Problema 335 *Probar la segunda fórmula del teorema precedente.*

Problema 336 *Sean $a, b \in \mathbb{R}^3$ con $\|a\| = 1$. Planteamos la ecuación $a \times z = b$. Probar que tiene solución si y solo si $\langle a, b\rangle = 0$. En tal caso ¿cuáles son las soluciones? Indicación: elegir una buena base.*

9.3. El producto vectorial. Orientaciones

Se puede considerar esta sección como materia optativa.

El producto vectorial es definible en un espacio euclidiano tridimensional arbitrario, pero tiene que estar *orientado*. Antes de definir "orientación" diremos que lo que se ha hecho en la sección anterior considera $(\mathbb{E}, \omega) = \left(\mathbb{R}^3, \varepsilon\right)$, y que la orientación "no se nota" porque hay una base "preferente" que es la base estándar, permitiendo tratar el concepto de modo más sencillo.

Dijimos al tratar los determinantes que en $\mathbb{E}$, de dimensión n pero con $\Bbbk$ arbitrario, nos gustaría tener una función semejante al determinante pero que en vez de actuar sobre endomorfismos actuara sobre sucesiones de vectores $(x_1, \ldots, x_n)$ en $\mathbb{E}$. El camino natural parece ser tomar una base $\mathcal{U}$ e identificar $(x_1, \ldots, x_n)$ con la matriz $x^{\mathcal{U}} \in \Bbbk^{n\times n}$ formada yuxtaponiendo los vectores columna $\mathrm{mat}^{\mathcal{U}}(x_j) \in \Bbbk^n$ y, a continuación, definir

$$\Delta_{\mathcal{U}} : \mathbb{E} \times \ldots \times \mathbb{E} = \mathbb{E}^n \longrightarrow \mathbb{R}, \quad \Delta_{\mathcal{U}}(x_1, \ldots, x_n) = \det\left(x^{\mathcal{U}}\right) = \det\left(\mathrm{mat}^{\mathcal{U}}(x_1), \ldots, \mathrm{mat}^{\mathcal{U}}(x_n)\right).$$

Sospechamos que $\Delta_{\mathcal{U}}$ depende de cómo se elija $\mathcal{U}$ y así es. En efecto, si $\mathcal{V}$ es otra base tenemos que $\mathrm{mat}^{\mathcal{V}}(y) = \mathrm{mat}^{\mathcal{V}}_{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\, \mathrm{mat}^{\mathcal{U}}(x)$. Abreviamos $c = \mathrm{mat}^{\mathcal{V}}_{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})$ y pedimos al lector que pruebe que $x^{\mathcal{V}} = cx^{\mathcal{U}}$, lo que lleva inmediatamente a

$$\Delta_{\mathcal{V}}(x_1, \ldots, x_n) = \det\left(x^{\mathcal{V}}\right) = \det(c) \det\left(x^{\mathcal{U}}\right) = \det\left(\mathrm{mat}^{\mathcal{V}}_{\mathcal{U}}(\mathrm{id}_{\mathbb{E}})\right) \Delta_{\mathcal{U}}(x_1, \ldots, x_n).$$

Esto dice que basta tomar bases cuya matriz de cambio c tenga $\det(c) \neq 1$ para obtener $\Delta_{\mathcal{V}} \neq \Delta_{\mathcal{U}}$.

Suponemos $\Bbbk = \mathbb{R}$ y que $\mathbb{E}$ tiene un producto euclidiano.

Recordamos que las matrices de cambio entre bases ortonormales tienen determinante ± 1 (**1** en el teorema 171) y nos planteamos un segundo intento para definir "adecuadamente" $\Delta_{\mathcal{U}}$. La idea es limitarnos a que $\mathcal{U}$ sea una base ortonormal. Esto casi tiene éxito porque si $\mathcal{V}$ es otra base ortonormal, $\Delta_{\mathcal{V}} = \pm\Delta_{\mathcal{U}}$. A partir de aquí hay dos salidas.

1. Conformarse con el valor absoluto $|\Delta_{\mathcal{U}}|$. Esta función de $\mathbb{E}^n \to \mathbb{R}$ no depende de la base ortonormal elegida. Sin entrar en materia podemos decir que tiene una interpretación interesante ya que $|\Delta_{\mathcal{U}}(x_1, \ldots, x_n)|$ es el volumen del paralelepípedo (prisma oblícuo) con los vectores $x_1, \ldots, x_n$ como aristas. (Es un poco vago, pero no podemos ahora entretenernos.)

[11] Regla mnemotécnica: el lejano menos el próximo (del vector fuera del paréntesis $(\bullet \times \bullet)$).

2. Resignarse a elegir una base como preferente, digamos $\mathcal{U}$, y saber que si por alguna razón necesitamos $\Delta_{\mathcal{V}}$, debemos quizás meter un signo menos. Es el camino que vamos a seguir, pero con un matiz importante: no vamos a tomar *una* base preferente sino una *clase* de bases preferentes.

El formalismo adecuado es este. Sea $\mathrm{B}(\mathbb{E})$ conjunto de todas las bases del espacio $\mathbb{E}$. De momento no pedimos que haya un producto euclidiano, pero sí que sea $\Bbbk = \mathbb{R}$. Establecemos en $\mathrm{B}(\mathbb{E})$ una relación de equivalencia que describimos. Dadas $\mathcal{U}$ y $\mathcal{V}$ hay matrices de cambio entre ellas c y d con $c = d^{-1}$. Si solo nos preocupa el *signo* de su determinante, es igual mirar una que otra. La relación es $\mathcal{U} \sim \mathcal{V}$ si el signo del determinante de una de las (y de hecho de ambas) matrices de cambio es > 0. Es muy sencillo ver que $\sim$ es relación de equivalencia. *El conjunto cociente, formado por sus clases, tiene solo dos elementos.* Consideremos en efecto $\mathcal{U} = (u_1, \ldots, u_n)$ y $\mathcal{V} = (-u_1, u_2, \ldots, u_n)$ (solo un signo $-$). La matriz de cambio es diagonal con $(-1, 1, \ldots, 1)$ en la diagonal (solo un signo $-$) luego $\mathcal{U} \nsim \mathcal{V}$ y hay al menos dos clases. Dada una tercera base $\mathcal{W}$ se tendrá

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(\mathrm{id}_{\mathbb{E}}) = \mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(\mathrm{id}_{\mathbb{E}})\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{V}}(\mathrm{id}_{\mathbb{E}}) \text{ y } \det\left(\mathrm{mat}_{\mathcal{U}}^{\mathcal{W}}(\mathrm{id}_{\mathbb{E}})\right) = -\det\left(\mathrm{mat}_{\mathcal{V}}^{\mathcal{W}}(\mathrm{id}_{\mathbb{E}})\right).$$

Esto muestra que si $\mathcal{W} \nsim \mathcal{U}$ sí que se tiene $\mathcal{W} \sim \mathcal{V}$, luego las clases de $\mathcal{U}$ y $\mathcal{V}$ son las únicas que hay. Hemos dividido pues a $\mathrm{B}(\mathbb{E})$ en dos partes y cada parte o clase, por definición es una **orientación de** $\mathbb{E}$. Por consiguiente, independientemente de que pueda sugerir "orientación", una orientación es un conjunto de bases de $\mathbb{E}$. Elegida una orientación, sus bases se llaman **positivas** y las de la otra clase, **negativas.** Hay por tanto dos **orientaciones opuestas.** Muchas veces se elige una orientación en $\mathbb{E}$ a través de una base preferente. El caso más obvio es $(\mathbb{R}^n, \varepsilon)$, siendo la base preferente la base estándar $\mathcal{E}$. Si $\mathcal{U}$ es otra base, $\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}_{\mathbb{R}^n})$ está formado por yuxtaposición de los u_i y las bases positivas son las que cumplen $\det(u_1, \ldots, u_n) > 0$.

En $\mathbb{R}^2$ con su **orientación estándar** (la que da la base $\mathcal{E}$), la base $\left((1,2)^\top, (3,4)^\top\right)$ cumple

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = -2 < 0,$$

luego es base negativa. Por otra parte, si $\mathbb{E} \subset \mathbb{R}^3$ es el espacio de soluciones de $x + y + z = 0$, son bases

$$\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right), \quad \left(\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right), \quad \left(\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right),$$

y muchas más. Ninguna parece preferible a las otras y ahí la elección de orientación es arbitraria.

En un espacio abstracto hay que elegir, con frecuencia arbitrariamente, una orientación, si bien en $\mathbb{R}^n$ se supone que es por defecto la que da la base estándar. Supongamos ahora que tenemos un espacio euclidiano $(\mathbb{E}, \omega)$ donde definimos $\mathbf{\Omega}(\mathbb{E}, \omega)$ como el conjunto de las bases ortonormales para ω. A continuación elegimos una orientación, que denotaremos por $\mathbf{\Omega}^+(\mathbb{E}, \omega)$, y $\mathbf{\Omega}^-(\mathbb{E}, \omega)$ será la orientación opuesta. La terna $(\mathbb{E}, \omega, \mathbf{\Omega}^+(\mathbb{E}, \omega))$ se llama **espacio euclidiano orientado**. Ahora sí podemos, *dependiendo solo de* $(\mathbb{E}, \omega, \mathbf{\Omega}^+(\mathbb{E}, \omega))$, definir el análogo al determinante, sin que haya dependencia de base con tal de que esté en $\mathbf{\Omega}^+(\mathbb{E}, \omega)$. La función es

$$\Delta : \mathbb{E}^n = \mathbb{E} \times \ldots \times \mathbb{E} \longrightarrow \mathbb{R}, \quad \Delta(x_1, \ldots, x_n) = \Delta_{\mathcal{U}}(x_1, \ldots, x_n), \text{ con cualquier } \mathcal{U} \in \mathbf{\Omega}^+(\mathbb{E}, \omega).$$

Como las bases de $\mathbf{\Omega}^+(\mathbb{E}, \omega)$ tienen matriz de cambio con determinante 1, queda claro que la definición es correcta. Se llama a Δ el **elemento de volumen** de $(\mathbb{E}, \omega, \mathbf{\Omega}^+(\mathbb{E}, \omega))$ y es obvio que si cambia ω u $\mathbf{\Omega}^+(\mathbb{E}, \omega)$, cambia Δ. Es también obvio que si $\mathbb{E} = \mathbb{R}^n$ con el producto euclidiano y la orientación estándar, el elemento de volumen es $\Delta(x_1, \ldots, x_n) = \det(x_1, \ldots, x_n)$. El elemento de volumen verifica la **identidad de Gram**, generalización del teorema 170,

$$\Delta(x_1, \ldots, x_n)\,\Delta(y_1, \ldots, y_n) = \det\left(\langle x_i, y_j\rangle\right)_{1 \le i,j \le n}. \tag{9.2}$$

La demostración es prácticamente la del teorema citado. Se toma una base ortogonal positiva auxiliar $\mathcal{U}$ y se calcula con ella

$$\langle x_i, y_j\rangle = \sum_{h=1}^{n} x_i^h y_j^h = \sum_{h=1}^{n} \left(x^\top\right)_h^i y_j^h = \left(x^\top y\right)_j^i; \text{ o sea, } \left(\langle x_i, y_j\rangle\right)_{1 \le i,j \le n} = x^\top y.$$

Las matrices x e y están formadas por yuxtaposición las columnas de coordenadas de los x_i e y_j en $\mathcal{U}$. Entonces,

$$\begin{aligned}\det\left(\langle x_i, y_j\rangle\right)_{1\leq i,j\leq n} &= \det\left(x^\top y\right) = \det(x)\det(y)\\ &= \Delta_{\mathcal{U}}(x_1,\ldots,x_n)\,\Delta_{\mathcal{U}}(y_1,\ldots,y_n) = \Delta(x_1,\ldots,x_n)\,\Delta(y_1,\ldots,y_n).\end{aligned}$$

Con este aparato es posible generalizar el producto vectorial a un espacio euclidiano $(\mathbb{E},\omega,\mathbf{\Omega}^+(\mathbb{E},\omega))$. Advertimos que un cambio en ω o de $\mathbf{\Omega}^+(\mathbb{E},\omega)$ por $\mathbf{\Omega}^-(\mathbb{E},\omega)$ lo alterará. Si queremos que el producto sea "tradicional"; es decir, que a dos vectores a y b le asigne otro $a\times b$, es necesario que sea $n=3$ y por eso hay quien dice que el producto vectorial es una peculiaridad de la dimensión 3. Si se está dispuesto a generalizar, es una función que multiplica $n-1$ factores,

$$\pi:\mathbb{E}^{n-1}=\mathbb{E}\times\ldots\times\mathbb{E}\longrightarrow\mathbb{E},\text{ que se suele denotar por }\pi(x_1,\ldots,x_{n-1})=x_1\times\ldots\times x_{n-1}.$$

La definición va a ser un tanto sutil porque se va a dar "en implicitas". Es el único vector $\pi(x_1,\ldots,x_{n-1})$ verificando que para todo y se cumple

$$\Delta(x_1,\ldots,x_{n-1},y)=\langle\pi(x_1,\ldots,x_{n-1}),y\rangle=\langle x_1\times\ldots\times x_{n-1},y\rangle.$$

Se puede dar una definición explícita. Se toma $\mathcal{U}$ base ortonormal positiva y con las coordenadas x_j^i de los x_j se define

$$x_1\times\ldots\times x_{n-1}=\begin{vmatrix} x_1^n & \cdots & x_{n-1}^1 & u_1\\ \vdots & \ddots & \vdots & \vdots\\ x_1^{n-1} & \cdots & x_{n-1}^{n-1} & u_{n-1}\\ x_1^n & \cdots & x_{n-1}^n & u_n\end{vmatrix}=\sum_{i=1}^n(x_1\times\ldots\times x_{n-1})^i\,u_i.$$

La definición es muy parecida a la del producto vectorial en $\mathbb{R}^3$. El determinante es de una "matriz" donde los x_j^i son números y los u_k son vectores y hay que entender que hay que desarrollar formalmente el "determinante" por la última columna, siendo entonces los coeficientes de los u_i las coordenadas de $x_1\times\ldots\times x_{n-1}$ en $\mathcal{U}$. Por ejemplo, en $\mathbb{R}^4$,

$$\begin{aligned}\begin{pmatrix}1\\2\\0\\2\end{pmatrix}\times\begin{pmatrix}2\\0\\1\\0\end{pmatrix}\times\begin{pmatrix}0\\0\\3\\4\end{pmatrix}&=\begin{vmatrix}1&2&0&u_1\\2&0&0&u_2\\0&1&3&u_3\\2&0&4&u_4\end{vmatrix}\\ &=-\begin{vmatrix}2&0&0\\0&1&3\\2&0&4\end{vmatrix}e_1+\begin{vmatrix}1&2&0\\0&1&3\\2&0&4\end{vmatrix}e_2-\begin{vmatrix}1&2&0\\2&0&0\\2&0&4\end{vmatrix}e_3+\begin{vmatrix}1&2&0\\2&0&0\\0&1&3\end{vmatrix}e_4.\\ &=-8e_1+16e_2+16e_3-12e_4=(-8,16,16,-12)^\top.\end{aligned}$$

La comprobación que la fórmula implícita equivale a la explícita es inmediata porque

$$\begin{aligned}\langle x_1\times\ldots\times x_{n-1},y\rangle&=\left\langle\begin{vmatrix} x_1^n & \cdots & x_{n-1}^1 & u_1\\ \vdots & \ddots & \vdots & \vdots\\ x_1^{n-1} & \cdots & x_{n-1}^{n-1} & u_{n-1}\\ x_1^n & \cdots & x_{n-1}^n & u_n\end{vmatrix},y^1u_1+\ldots+y^nu_n\right\rangle=\begin{vmatrix} x_1^n & \cdots & x_{n-1}^1 & y^1\\ \vdots & \ddots & \vdots & \vdots\\ x_1^{n-1} & \cdots & x_{n-1}^{n-1} & y^{n-1}\\ x_1^n & \cdots & x_{n-1}^n & y^n\end{vmatrix}\\ &=\Delta_{\mathcal{U}}(x_1,\ldots,x_{n-1},y)=\Delta(x_1,\ldots,x_{n-1},y).\end{aligned}$$

La fórmula explícita (el determinante formal) es la que se usa para cálculos explícitos y tiene un valor añadido: sirve para calcular el producto en *cualquier base ortonormal positiva* $\mathcal{U}$, si bien el resultado son las coordenadas en esa misma base. Esto no se ve en el producto vectorial "clásico" de $\mathbb{R}^3$ pero es cierto y tiene su importancia. Es curioso que si bien la definición explícita es la de los cálculos explícitos (valga la redundancia) *la implícita es la verdaderamente adecuada para demostrar las fórmulas generales*. Se tienen las propiedades fundamentales

1. El producto vectorial es antisimétrico y lineal en cada variable; o sea

$$\begin{gathered}x_1\times\ldots\times\left(x_j+x_j'\right)\times\ldots\times x_{n-1}=x_1\times\ldots\times x_j\times\ldots\times x_{n-1}+x_1\times\ldots\times x_j'\times\ldots\times x_{n-1}\\ x_1\times\ldots\times(\lambda x_j)\times\ldots\times x_{n-1}=\lambda(x_1\times\ldots\times x_j\times\ldots\times x_{n-1})\\ x_1\times\ldots\times x_i\times\ldots\times x_j\times\ldots\times x_{n-1}=-x_1\times\ldots\times x_j\times\ldots\times x_i\times\ldots\times x_{n-1}\end{gathered}$$

2. El producto vectorial es perpendicular a cada factor.

3. Se verifica que $x_1 \times \ldots \times x_{n-1} \neq 0$ si y solo si $(x_1, \ldots, x_{n-1})$ es independiente. De hecho, si $(x_1, \ldots, x_{n-1})$ es independiente se tiene que $(x_1, \ldots, x_{n-1}, x_1 \times \ldots \times x_{n-1})$ es base de $\mathbb{E}$.

Probemos, por ejemplo **3**. Si $x_1 \times \ldots \times x_{n-1} \neq 0$ será también $\|x_1 \times \ldots \times x_{n-1}\| \neq 0$ y tendremos con la definición implícita que

$$\Delta(x_1, \ldots, x_{n-1}, x_1 \times \ldots \times x_{n-1}) = \|x_1 \times \ldots \times x_{n-1}\|^2 \neq 0.$$

Como Δ se calcula con un determinante, $\Delta(y_1, \ldots, y_n) \neq 0$ equivale a que $(y_1, \ldots, y_n)$ sea independiente. Hemos probado que $x_1 \times \ldots \times x_{n-1} \neq 0$ implica que $(x_1, \ldots, x_{n-1}, x_1 \times \ldots \times x_{n-1})$ es base de $\mathbb{E}$ y, en particular la independencia de $(x_1, \ldots, x_{n-1})$. Recíprocamente, si $(x_1, \ldots, x_{n-1})$ es independiente, se completa hasta una base $(x_1, \ldots, x_{n-1}, x_n)$ de $\mathbb{E}$. Evidentemente, al calcularse Δ con un determinante, $\Delta(x_1, \ldots, x_{n-1}, x_n) \neq 0$ y, con la definición implícita, $\langle x_1 \times \ldots \times x_{n-1}, x_n \rangle \neq 0$, luego $x_1 \times \ldots \times x_{n-1} \neq 0$.

El producto vectorial de productos vectoriales tiene enunciado y demostración similar al teorema 173.

Teorema 175 (producto escalar de productos vectoriales) . *Tenemos que*

$$\langle x_1 \times \ldots \times x_{n-1}, y_1 \times \ldots \times y_{n-1} \rangle = \det\left(\langle x_i, y_j \rangle\right)_{1 \leq i,j \leq n-1}.$$

En particular, $\|x_1 \times \ldots \times x_{n-1}\|^2 = \Gamma(x_1, \ldots, x_{n-1})$.

Demostración. Si hay dependencia en $(x_1, \ldots, x_{n-1})$ se ve enseguida que la fórmula es cierta, pues sale $0 = 0$. Descartemos ese caso. Entonces $(x_1, \ldots, x_{n-1})$ es independiente y $\|x_1 \times \ldots \times x_{n-1}\| \neq 0$ como acabamos de ver. Hacemos $x_n = y_n = x_1 \times \ldots \times x_{n-1}$ y aplicamos la identidad de Gram (9.2). Por una parte,

$$\begin{aligned}
\det\left(\langle x_i, y_j \rangle\right)_{1 \leq i,j \leq n} &= \begin{vmatrix} \langle x_1, y_1 \rangle & \cdots & \langle x_1, y_{n-1} \rangle & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \langle x_{n-1}, y_1 \rangle & \cdots & \langle x_{n-1}, y_{n-1} \rangle & 0 \\ \langle x_n, y_1 \rangle & \cdots & \langle x_n, y_{n-1} \rangle & \|x_1 \times \ldots \times x_{n-1}\|^2 \end{vmatrix} \\
&= \begin{vmatrix} \langle x_1, y_1 \rangle & \cdots & \langle x_1, y_{n-1} \rangle \\ \vdots & \ddots & \vdots \\ \langle x_{n-1}, y_1 \rangle & \cdots & \langle x_{n-1}, y_{n-1} \rangle \end{vmatrix} \|x_1 \times \ldots \times x_{n-1}\|^2 .
\end{aligned}$$

El otro lado de la identidad de Gram es

$$\Delta(x_1, \ldots, x_n)\, \Delta(y_1, \ldots, y_n) = \|x_1 \times \ldots \times x_{n-1}\|^2 \langle y_1 \times \ldots \times y_{n-1}, x_1 \times \ldots \times x_{n-1} \rangle .$$

Cancelando $\|x_1 \times \ldots \times x_{n-1}\|^2 \neq 0$ queda la identidad del teorema. La última parte es trivial. ♣

9.4. Sumas, proyecciones y simetrías ortogonales

En lo sucesivo $(\mathbb{E}, \omega)$ será un espacio euclidiano. Se define el **ortogonal** de $\mathbb{F}$ como el subespacio

$$\mathbb{F}^{\perp} = \{x \in \mathbb{E} \mid \omega(x, y) = 0 \text{ para todo } y \in \mathbb{F}\}.$$

Es muy fácil ver que $\mathbb{F}$ es un subespacio y que los vectores de $\mathbb{F}$ y $\mathbb{F}^{\perp}$ son mutuamente perpendiculares. Si $x \in \mathbb{F} \cap \mathbb{F}^{\perp}$ se tiene con $x = y$ que $\omega(x, x) = 0$, luego $x = 0$ y $\mathbb{F} \cap \mathbb{F}^{\perp} = 0$. Todo esto vale aun sin dimensión finita.

Teorema 176 *Se tiene siempre que* $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^{\perp}$ *supuesto* $\mathbb{F}$ *(pero no necesariamente* $\mathbb{E}$*) de dimensión finita* m*. Si* $dim(\mathbb{E}) = n$ *entonces* $\dim\left(\mathbb{F}^{\perp}\right) = n - m$.

Demostración. Es suficiente que sea $\mathbb{E} = \mathbb{F} + \mathbb{F}^{\perp}$ porque $\mathbb{F} \cap \mathbb{F}^{\perp} = 0$. Sea $\mathcal{W} = (w_1, \ldots, w_m)$ una base ortonormal de $\mathbb{F}$, que existe por el teorema 167. Evidentemente,

$$x = \left[\sum_{j=1}^{m} \langle x, w_j \rangle w_j\right] + \left[x - \sum_{j=1}^{m} \langle x, w_j \rangle w_j\right]. \tag{9.3}$$

El primer corchete sin duda está en $\mathbb{F}$. El segundo está en $\mathbb{F}^\perp$ pues al multiplicar por cada w_i,

$$\left\langle x-\sum_{j=1}^{m}\langle x,w_j\rangle\, w_j, w_i\right\rangle = \langle x,w_i\rangle - \sum_{j=1}^{n}\langle x,w_j\rangle\,\langle w_j,w_i\rangle = \langle x,w_k\rangle - \sum_{j=1}^{n}\langle x,w_j\rangle\,\delta_{ji} = \langle x,w_i\rangle - \langle x,w_i\rangle = 0.$$

Tenemos cada $x\in\mathbb{E}$ como suma de un vector en $\mathbb{F}$ y otro en $\mathbb{F}^\perp$, luego $\mathbb{E}=\mathbb{F}+\mathbb{F}^\perp$. Una vez que $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$, $\dim\left(\mathbb{F}^\perp\right)=\dim\left(\mathbb{E}\right)-\dim\left(\mathbb{F}\right)$. ♣

La ventaja de tener en $\mathbb{E}$ un producto euclidiano ω es que *con un solo subespacio* $\mathbb{F}$ de dimensión finita, hay una descomposición en suma directa $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$, al haber un suplementario de $\mathbb{F}$ determinado por ω. La descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$ en suma directa se llama la **descomposición ortogonal** (dada por $\mathbb{F}$) y se hace con dos subespacios mutuamente perpendiculares. Por nuestra costumbre de trabajar en $\mathbb{R}^3$ parece imposible que sea $\mathbb{F}+\mathbb{F}^\perp\neq\mathbb{E}$, pero la dimensión infinita se la juega a la intuición. Véase el problema 345.

Sabemos desde si $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$, hay simetrías y proyecciones. Tenemos por tanto dos proyecciones $P:\mathbb{E}\to\mathbb{F}$ y $P^\perp:\mathbb{E}\to\mathbb{F}^\perp$ que se llaman **proyecciones ortogonales** (sobre $\mathbb{F}$ y $\mathbb{F}^\perp$) y **simetrías ortogonales** S y $S^\perp$ respecto a $\mathbb{F}$, aunque, insistimos, se necesita disponer de un producto euclidiano $\omega=\langle\bullet,\bullet\rangle$. Si $x=y+z$ con $y\in\mathbb{F}$ y $z\in\mathbb{F}^\perp$ tenemos por definición que

$$P\left(x\right)=y,\qquad P^\perp\left(x\right)=z,\qquad S\left(x\right)=y-z,\qquad S^\perp\left(x\right)=-y+z.$$

Los cuatro endomorfismos $P,P^\perp,S$ y $S^\perp$ se pueden construir con solo uno de ellos; por ejemplo P pues

$$S\left(x\right)=P\left(x\right)-\left[x-P\left(x\right)\right]=2P\left(x\right)-x,\qquad S^\perp\left(x\right)=-P\left(x\right)+\left[x-P\left(x\right)\right]=x-2P\left(x\right).$$

La fórmula (9.3) permite calcular P, pero se necesita primero construir una base ortogonal $\mathcal{V}$ u ortonormal $\mathcal{W}$ en $\mathbb{F}$ y entonces

$$P\left(x\right)=\sum_{j=1}^{m}\frac{\langle x,v_j\rangle}{\left\|v_j\right\|^2}v_j=\sum_{j=1}^{m}\langle x,w_j\rangle\,w_j. \tag{9.4}$$

Como el construir estas bases supone un trabajo, suele ser conveniente calcular $P^\perp$, si $\dim\left(\mathbb{F}^\perp\right)<\dim\left(\mathbb{F}\right)$ y luego aplicar $P=\mathrm{id}-P^\perp$. Hay dos casos muy sencillos: cuando $\mathbb{F}=\mathbb{D}$ es una recta dada como $\lg\left(v\right)$, o cuando $\mathbb{F}=\mathbb{H}$ es un hiperplano dado por $\langle x,v\rangle=0$. Entonces $\mathcal{V}=\lg\left(v\right)$ es una base de $\mathbb{D}$ o $\mathbb{H}^\perp$, los sumatorios solo tienen un término, y

$$P_{\mathbb{D}}\left(x\right)=\frac{\langle x,v\rangle}{\left\|v\right\|^2}v,\qquad P_{\mathbb{H}}\left(x\right)=x-\frac{\langle x,v\rangle}{\left\|v\right\|^2}v$$

son las proyecciones sobre la recta $\mathbb{D}$ y el hiperplano $\mathbb{H}$. Las fórmulas salen de

$$x=\frac{\langle x,v\rangle}{\left\|v\right\|^2}v+\left[x-\frac{\langle x,v\rangle}{\left\|v\right\|^2}v\right]\in\mathbb{D}\oplus\mathbb{H}=\mathbb{D}\oplus\mathbb{D}^\perp=\mathbb{H}^\perp\oplus\mathbb{H}.$$

La **simetría axial** respecto a $\mathbb{D}$ y la **simetría especular** respecto a $\mathbb{H}$ son

$$S_{\mathbb{D}}\left(x\right)=P_{\mathbb{D}}\left(x\right)-P_{\mathbb{D}}^\perp\left(x\right)=\frac{2\,\langle x,v\rangle}{\left\|v\right\|^2}v-x,\qquad S_{\mathbb{H}}\left(x\right)=P_{\mathbb{H}}\left(x\right)-P_{\mathbb{H}}^\perp\left(x\right)=x-\frac{2\,\langle x,v\rangle}{\left\|v\right\|^2}v.$$

La manera de salir de dudas en las fórmulas es que en $P_\bullet$ o $S_\bullet$ el subespacio que aparece como subíndice es el conjunto de puntos fijos. Simplemente en $n=3$, una simetría axial fija una recta y una especular fija un plano. Son las que solemos llamar "simetrías" en el lenguaje usual.

Si estamos en $\mathbb{R}^3$ con el producto estándar ε (esto es esencial) podemos calcular la proyección sobre cualquier subespacio, sin que importe si nos lo dan en implícitas o paramétricas. Solo hay dos posibilidades, que son $\mathbb{F}=\mathbb{D}$ (recta) o $\mathbb{F}=\mathbb{H}$ (plano). Si nos dan $\mathbb{D}$ en paramétricas $\mathbb{D}=\lg\left(u\right)$ o $\mathbb{H}$ en implícitas $\mathbb{H}=\{x\mid\langle x,u\rangle=0\}$, ya conocemos las fórmulas de $P_{\mathbb{D}}$ y $P_{\mathbb{H}}$, incluso con mucha más generalidad. Si tenemos $\mathbb{H}$ en paramétricas, definido por $\mathbb{H}=\lg\left(u,v\right)$, sabemos (aquí entra en juego el caso particular de $\mathbb{R}^3$)[12] que $u\times v$ es no nulo y ortogonal a $\mathbb{H}$. Sirve la fórmula anterior

$$P_{\mathbb{H}}\left(x\right)=x-\frac{\langle x,u\times v\rangle}{\left\|u\times v\right\|^2}\left(u\times v\right)=x-\frac{\det\left(x,u,v\right)}{\left\|u\times v\right\|^2}\left(u\times v\right)\overset{*}{=}x-\frac{\det\left(x,u,v\right)}{\left\|u\right\|^2\left\|v\right\|-\langle u,v\rangle^2}\left(u\times v\right).$$

[12] Esto es así porque suponemos que el lector ha ignorado la materia optativa sobre el producto vectorial y solo lo conoce en $\mathbb{R}^3$. No obstante, si supone $\mathbb{E}$ tridimensional y orientado tiene el el producto vectorial asociado y todo lo que ahora describimos funciona sin cambios.

(Se da el paso $\overset{*}{=}$ porque $\|u\times v\|^2 = \|u\|^2\|v\| - \langle u,v\rangle^2$). Nos queda el caso en que $\mathbb{D}$ esté en implícitas; o sea $\mathbb{D} = \{x \mid \langle x,u\rangle = \langle x,v\rangle = 0\}$. Entonces $\lg(u\times v) = \mathbb{D}$ y se tiene

$$P_{\mathbb{D}}(x) = \frac{\langle x, u\times v\rangle}{\|u\times v\|^2}(u\times v) = \frac{\det(x,u,v)}{\|u\times v\|^2}(u\times v) \overset{*}{=} \frac{\det(x,u,v)}{\|u\|^2\|v\| - \langle u,v\rangle^2}(u\times v).$$

Problema 337 *Plantear y resolver en* $\left(\mathbb{R}^3,\varepsilon\right)$ *un problema de cálculo numérico de* $P_{\mathbb{D}}(x)$ *y* $P_{\mathbb{H}}(x)$ *expresando* $\mathbb{D}$ *y* $\mathbb{H}$ *de todas las maneras posibles.*

En el siguiente problema se trabaja en $\mathbb{R}^3$ con ε. Hemos calculado ignorando el producto vectorial. No obstante, al generalizar a $\mathbb{R}^n$ ya no hay fórmulas simplificadas.

Problema 338 *En* $\mathbb{R}^3$ *se trabaja con el producto estándar* ε. *Dar las ecuaciones del hiperplano* $\mathbb{H}$ *ortogonal a la recta* $\mathbb{D}$ *generada por* $v = (1,1,1)^\top$. *Determinar las proyecciones del vector* $t = (1,2,3)^\top$ *sobre* $\mathbb{D}$ *y* $\mathbb{H}$ *así como sus imágenes bajo las simetrías* $S_{\mathbb{D}}$ *y* $S_{\mathbb{H}}$. *Extender los cálculos para* n *arbitrario.* ♦

Solución. La condición $\langle t,x\rangle = 0$ equivale a $x^1+x^2+x^3 = 0$. Entonces, $\|v\|^2 = 1^2+1^2+1^2 = 3$ y

$$P_{\mathbb{D}}(t) = \frac{\langle t,v\rangle}{\|v\|^2}v = \frac{1}{3}(1\cdot 1 + 2\cdot 1 + 3\cdot 1)\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}2\\2\\2\end{pmatrix},$$

$$S_{\mathbb{D}}(t) = \frac{2\langle t,v\rangle}{\|v\|^2}v - t = 2\frac{6}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}3\\2\\1\end{pmatrix},$$

$$S_{\mathbb{H}}(t) = t - \frac{2\langle t,v\rangle}{\|v\|^2}v = \begin{pmatrix}1\\2\\3\end{pmatrix} - 2\frac{6}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}-3\\-2\\-1\end{pmatrix} = -S_{\mathbb{L}}(t).$$

Esto último se podía adivinar porque, con toda generalidad,

$$S^\perp \circ S(x) = S^\perp(y-z) = -y-z = -x.$$

Para n arbitrario la única diferencia es que

$$\|v\|^2 = \sum_{i=1}^{n} 1^2 = n, \quad \langle t,v\rangle = 1+2+\ldots+n = \frac{n(n+1)}{2}, \quad \frac{\langle t,v\rangle}{\|v\|^2} = \frac{n+1}{2}$$

y las fórmulas definitivas quedan para el lector. ♦

Si en la descomposición $\mathbb{E} = \mathbb{F}\oplus\mathbb{F}^\perp$ ni $\mathbb{F}$ ni $\mathbb{F}^\perp$ son rectas e hiperplanos, el cálculo de las proyecciones se complica porque se necesita una base ortogonal de $\mathbb{F}$ o $\mathbb{F}^\perp$. Hay una manera de evitar este cálculo directo con fórmulas donde aparecen las matrices de Gram. Veremos también que se puede obtener $\|x - P(x)\|^2$ que, con definiciones que se tratan en otra sección, será el cuadrado de la distancia entre x y $\mathbb{F}$. En $\mathbb{E} = \mathbb{F}\oplus\mathbb{F}^\perp$ tomemos $x\in\mathbb{F}$ y descompongámoslo como $x = y+z$. Sea $\mathcal{B} = (b_1,\ldots,b_m)$ una base, posiblemente no ortogonal, de $\mathbb{F}$. Podemos escribir

$$x = y+z = P(x) + P^\perp(x) = \sum_{i=1}^{m}\lambda^i b_i + z.$$

Teorema 177 *Con estas notaciones los números* λ^i *y* $\|z\|^2$ *verifican la ecuación matricial*

$$\begin{pmatrix} 1 & \langle b_1,x\rangle & \cdots & \langle b_m,x\rangle \\ 0 & \langle b_1,b_1\rangle & \cdots & \langle b_1,b_m\rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \langle b_m,b_1\rangle & \cdots & \langle b_m,b_m\rangle \end{pmatrix}\begin{pmatrix} \|z\|^2 \\ \lambda^1 \\ \vdots \\ \lambda^m \end{pmatrix} = \begin{pmatrix} \|x\|^2 \\ \langle b_1,x\rangle \\ \vdots \\ \langle b_m,x\rangle \end{pmatrix}. \tag{9.5}$$

Sus expresiones explicitas, con la regla de Cramer, son

$$\lambda^j = \frac{\begin{vmatrix} \langle b_1,b_1\rangle & \cdots & \langle b_1,b_{j-1}\rangle & \langle b_1,x\rangle & \langle b_1,b_{j+1}\rangle & \cdots & \langle b_1,b_m\rangle \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle b_m,b_1\rangle & \cdots & \langle b_m,b_{j-1}\rangle & \langle b_m,x\rangle & \langle b_m,b_{j+1}\rangle & \cdots & \langle b_m,b_m\rangle \end{vmatrix}}{\Gamma(b_1,\ldots,b_m)}, \quad \|z\|^2 = \frac{\Gamma(x,b_1,\ldots,b_m)}{\Gamma(b_1,\ldots,b_m)}.$$

Demostración. Dado que $\langle y,z\rangle = 0$ se tiene

$$\left\|x^2\right\| = \left\langle \sum_{i=1}^m \lambda^i b_i, x \right\rangle + \langle z,x\rangle = \sum_{i=1}^m \lambda^i \langle b_i,x\rangle + \langle z, y+z\rangle = \sum_{i=1}^m \lambda^i \langle b_i,x\rangle + \|z\|^2.$$

Añadimos m ecuaciones más multiplicando $x = \sum_{i=1}^m \lambda^i b_i + z$ por las b_j, y queda el sistema

$$\begin{cases} \left\|x^2\right\| = \lambda^1 \langle b_1,x\rangle + \ldots + \lambda^m \langle b_1,x\rangle + \|z\|^2 \\ \langle b_1,x\rangle = \lambda^1 \langle b_1,b_1\rangle + \ldots + \lambda^m \langle b_1,b_m\rangle \\ \vdots\vdots \\ \langle b_m,x\rangle = \lambda^1 \langle b_m,b_1\rangle + \ldots + \lambda^m \langle b_m,b_m\rangle \end{cases},$$

que tiene la forma matricial del enunciado. La aplicación de la regla de Cramer es inmediata. ♣

Advertimos de nuevo que los determinantes son muy laboriosos de calcular y que para m y n grande pueden no ser las mejores fórmulas. Si nos interesa solo la proyección $y = P(x)$, o sea, los λ^i, tenemos un sistema un poco más pequeño que (9.5), que es

$$G(b_1,\ldots,b_m)\begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^m \end{pmatrix} = \begin{pmatrix} \langle b_1,b_1\rangle & \cdots & \langle b_1,b_m\rangle \\ \vdots & \ddots & \vdots \\ \langle b_m,b_1\rangle & \cdots & \langle b_m,b_m\rangle \end{pmatrix}\begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^m \end{pmatrix} = \begin{pmatrix} \langle b_1,x\rangle \\ \vdots \\ \langle b_m,x\rangle \end{pmatrix}. \tag{9.6}$$

En el teorema se ha pedido que $(b_1,\ldots,b_m)$ sea base se $\mathbb{F} = \lg(b_1,\ldots,b_m)$, pero aunque solo sea una sucesión generadora se puede escribir el sistema (9.6) que será compatible, pues $P(x)$ existe y será expresable al menos de una manera como $P(x) = \lambda^1 b_1 + \ldots + \lambda^m b_m$ siendo $\left(\lambda^1,\ldots,\lambda^m\right)$ solución de ese sistema. Lo normal sin embargo es que $(b_1,\ldots,b_m)$ sea base de $\mathbb{F}$ y los λ^i sean únicos. Si no se dice lo contrario, $(b_1,\ldots,b_m)$ se supondrá base de $\mathbb{F}$. Hay que advertir también que $\mathbb{F}$ debe tener dimensión finita, lo que asegura el poder descomponer $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^\perp$ (teorema 176), pero puede ser $\dim(\mathbb{E}) = \infty$.

Si particularizamos al espacio $(\mathbb{R}^n,\varepsilon)$ y $b = (b_1,\ldots,b_m) \in \mathbb{R}^{n\times m}$, este sistema es $b^\top b\lambda = b^\top x$ ya que $b^\top b = G(b_1,\ldots,b_m)$ y $(\langle b_1,x\rangle,\ldots,\langle b_m,x\rangle) = b^\top x$. Como $b^\top b$ es invertible y x es conocido (la incógnita es λ) resolvemos $b^\top b\lambda = b^\top x$ y tenemos $\lambda = \left(b^\top b\right)^{-1} b^\top x$.[13]

Problema 339 *Calcular en $\mathbb{R}^4$ con el producto estándar la proyección de $x = (1,2,1,2)^\top$ en el subespacio generado por $p = (1,2,3,4)$ y $q = (4,3,2,1)$.*

Problema 340 *En $\mathbb{R}^4$ con ε como producto euclidiano calcular la proyección P sobre el plano $\mathbb{F}$ generado por $b_1 = (1,1,1,1)$ y $b_2 = (1,0,1,0)$, incluyendo también $\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(P)$.*

Problema 341 *Se considera en $\mathbb{R}^3$ el producto escalar dado por ω con matriz en la base estándar*

$$s = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

y $v = (0,1,1)^\top$. Determinar la imagen de $t = (2,0,1)^\top$ por las simetrías axial y especular respecto a la recta $\mathbb{D}$ generada por v y al plano $\mathbb{H}$ ortogonal a v.

[13] Muchos libros orientados al cálculo presentan una sucesión $(b_1,\ldots,b_m)$ de m vectores independientes en $\mathbb{R}^m$, los yuxtaponen formando una matriz $n\times m$ (¡ojo al orden de m y n!) y dan como dato $x \in \mathbb{R}^n$. Se tiene que $b^\top b \in \mathbb{R}^{m\times m}$ y $b^\top x \in \mathbb{R}^m$ por lo que el sistema lineal $b^\top b\lambda = b^\top x$ con incógnita $\lambda = \left(\lambda^1,\ldots,\lambda^m\right)$ tiene sentido y solución $\lambda = \left(b^\top b\right)^{-1} b^\top x$. Estos libros acaban por donde nosotros empezamos, que es probando que $y = \lambda^1 b_1 + \ldots + \lambda^m b_m$ es la proyección ortogonal de x sobre $\lg(b_1,\ldots,b_m) = \mathbb{F}$, supuesto que el producto euclidiano es el estándar.

Problema 342 *Damos a* $\mathbb{E}=\mathbb{R}_n[X]$ *el producto euclidiano* $\omega(P(X),Q(X))=\int_0^1 P(X)\,Q(X)\,dX$. *Probar que la base estándar* $\mathcal{E}$ *no es ortogonal ni ortonormal porque* $\omega\left(X^i,X^j\right)=1/(i+j+1)$.

Problema 343 *Damos a* $\mathbb{E}=\mathbb{R}[X]$ *el producto euclidiano* $\omega(P(X),Q(X))=\int_0^1 P(X)\,Q(X)\,dX$, *Se considera el subespacio* $\mathbb{F}$ *generado por* $\left(X,X^2\right)$.

1. *Determinar la proyección ortogonal de* $1+X^3$ *sobre* $\mathbb{F}$.
2. *Hallar la imagen del simétrico de* X^5 *respecto del hiperplano ortogonal a* X^4

Problema 344 *En* $\mathbb{R}^{n\times n}$ *damos el producto euclidiano* $\langle a,b\rangle=\operatorname{tr}\left(a^\top b\right)$. *Pedimos*

1. *Probar que los subespacios* $\mathbb{S}$ *y* $\mathbb{A}$ *de las matrices simétricas y antisimétricas son mutuamente perpendiculares y se tiene* $\mathbb{S}^\perp=\mathbb{A}$ *y* $\mathbb{A}=\mathbb{S}^\perp$, *luego* $\mathbb{R}^{n\times n}=\mathbb{S}\oplus\mathbb{A}$. *Las proyecciones sobre* $\mathbb{S}$ *y* $\mathbb{A}$ *son respectivamente* $P(a)=\left(a+a^\top\right)/2$ *y* $Q(a)=\left(a-a^\top\right)/2$.
2. *Si* $\mathbb{F}$ *es la recta generada por* I, *dar las fórmulas de* $P,P^\perp,Q,Q^\perp$ *para* $\mathbb{F}\oplus\mathbb{F}^\perp$.

Problema 345 *Sea* $\mathbb{E}$ *el espacio de las funciones continuas de* $[0,1]$ *en* $\mathbb{R}$ *y* $\mathbb{F}$ *el subespacio de las que cumplen* $f(0)=0$. *El producto es* $\omega(f(t),g(t))=\int_0^1 f(t)\,g(t)\,dt$. *Probar que* $\mathbb{F}^\perp=0$ *luego, como* $\mathbb{F}\neq\mathbb{E}$, *es imposible tener* $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$. *Indicación: si* $f(t)\in\mathbb{F}^\perp$, $tf(t)\in\mathbb{F}$ *y...*

Problema 346 *En* $\mathbb{E}$, *posiblemente de dimensión infinita, enunciamos propiedades de subespacios*

$$(\mathbf{a})\ \left(\mathbb{F}^\perp\right)^\perp=\mathbb{F},\quad (\mathbf{b})\ \mathbb{F}=\mathbb{G}\iff\mathbb{F}^\perp=\mathbb{G}^\perp,\quad (\mathbf{c})\ \mathbb{F}\subset\mathbb{G}\iff\mathbb{F}^\perp\supset\mathbb{G}^\perp.$$

¿Cuáles son ciertas? ¿Son parcialmente ciertas? ¿Y si $\dim(\mathbb{E})<\infty$*?*

Problema 347 *Sean* $\mathbb{F}_1$ *y* $\mathbb{F}_2$ *subespacios de* $\mathbb{E}$. *Consideramos las propiedades*

$$(\mathbf{a})\ (\mathbb{F}_1+\mathbb{F}_2)^\perp=\mathbb{F}_1^\perp\cap\mathbb{F}_2^\perp,\quad (\mathbf{b})\ (\mathbb{F}_1\cap\mathbb{F}_2)^\perp=\mathbb{F}_1^\perp+\mathbb{F}_2^\perp.$$

¿Cuáles son ciertas? ¿Son parcialmente ciertas? ¿Y si $\dim(\mathbb{E})<\infty$*?*

9.5. La ortogonalización de Gram-Schmidt

En $\mathbb{E}$, que puede tener dimensión infinita, sea $X=(x_1,\ldots,x_m)$ una sucesión de vectores independientes. **Ortogonalizar** X es construir una sucesión $V=(v_1,\ldots,v_m)$ de vectores *ortogonales* de forma que para $k=1,2,\ldots,m$ se tenga $\lg(x_1,\ldots,x_k)=\lg(v_1,\ldots,v_k)$. La ortogonalización de Gram-Schmidt es el método más usual para ortogonalizar $(x_1,\ldots,x_m)$. Antes de describirla, nos planteamos si $(x_1,\ldots,x_m)$ puede tener más de una ortogonalización. La respuesta es sí, pero la diferencia es pequeña.

Teorema 178 *Sean* $U=(u_1,\ldots,u_m)$ *y* $V=(v_1,\ldots,v_m)$ *ortogonalizaciones del mismo* X. *Se cumple entonces que existen* $\lambda_1,\ldots,\lambda_m$ *no nulos tales que* $u_j=\lambda_j v_j$ *para* $j=1,\ldots,,m$.

Demostración. Como $\lg(x_1,\ldots,x_k)=\lg(u_1,\ldots,u_k)=\lg(v_1,\ldots,v_k)$, hay ecuaciones *vectoriales*

$$\begin{cases} u_1=\lambda_1^1 v_1\\ u_2=\lambda_2^1 v_1+\lambda_2^2 v_2\\ \vdots\\ u_k=\lambda_k^1 v_1+\lambda_k^2 v_2+\ldots+\lambda_k^h v_h+\ldots+\lambda_k^{k-1}v_{k-1}+\lambda_k^k v_k\\ \vdots\\ u_m=\lambda_m^1 v_1+\lambda_m^2 v_2+\ldots+\lambda_m^h v_h+\ldots+\lambda_m^{m-1}v_{k-1}+\lambda_m^m v_m \end{cases}$$

Probamos que todos los λ_j^i con $i<j$ son nulos. Cada v_k es ortogonal a $v_1,\ldots,v_{k-1}$ luego también a $u_1,\ldots,u_{k-1}$, porque están en el mismo subespacio, y tenemos $\langle u_k,v_h\rangle=0$ si $h<k$. Multiplicando por v_h las ecuaciones de u_k con $h<k$, por la ortogonalidad de $(v_1,\ldots,v_m)$,

$$0=\langle u_k,v_h\rangle=\lambda_k^h\left\|v_h\right\|^2,$$

y como $\left\|v_h\right\|^2\neq 0$, resulta $\lambda_k^h=0$ si $h<k$. Las ecuaciones se reducen a $u_k=\lambda_k^k v_k$ como queríamos. ♣

Teorema 179 (Ortogonalización de Gram-Schmidt) *Sea $X = (x_1, \ldots, x_r)$ una sucesión de vectores independiente que definen subespacios $\mathbb{F}_k = \lg(x_1, \ldots, x_k)$, $k = 1, \ldots, , r$. Sean P_k y $P_k^\perp$ las proyecciones sobre $\mathbb{F}_k$ y $\mathbb{F}_k^\perp$. La sucesión $V = (v_1, \ldots, v_r)$ de vectores dados por*

$$v_1 = x_1, \qquad v_k = P_{k-1}^\perp(x_k) = x_k - P_{k-1}(x_k) \text{ para } 2 \leq k \leq r, \tag{9.7}$$

es una ortogonalización de $X = (x_1, \ldots, x_r)$.

Recomendamos hacer un dibujo en $\mathbb{R}^3$ con $\omega = \varepsilon$, donde la perpendicularidad es fácilmente visualizable, representando $X = (x_1, x_2, x_3)$ como tres vectores que forman un triedro no ortogonal.

1. Empezamos con $v_1 = x_1$ y tenemos $\mathbb{F}_1 = \lg(x_1) = \lg(v_1)$, que es una recta.
2. Descomponemos $\mathbb{E} = \mathbb{F}_1 \oplus \mathbb{F}_1^\perp$ y con ella $x_2 = u_2 + v_2$. Tomamos $v_2 \in \mathbb{E}_1^\perp$.
3. Descomponemos $\mathbb{E} = \mathbb{F}_2 \oplus \mathbb{F}_2^\perp$ y con ella $x_3 = u_3 + v_3$. Tomamos $v_3 \in \mathbb{E}_2^\perp$.

Si $\mathbb{F}_1, \mathbb{F}_1^\perp, \mathbb{F}_2$ y $\mathbb{F}_2^\perp$ se han dibujado correctamente se verá que (v_1, v_2, v_3) es un triedro ortogonal en el espacio.

Demostración. Para cada $k = 1, \ldots, r$ comprobamos que $V_k = (v_1, \ldots, v_k)$ es una ortogonalización de $X_k = (x_1, \ldots, x_k)$ de modo inductivo. Así, para $k = r$ se tendrá el teorema.

El teorema es cierto si $k = 1$. Supongamos el resultado cierto hasta k y probémoslo para $k+1$. Por la definición, $v_{k+1} = P_k^\perp(x_{k+1}) \in \mathbb{F}_k^\perp = \lg(x_1, \ldots, x_k)^\perp \overset{*}{=} \lg(v_1, \ldots, v_k)^\perp$, usándose en $\overset{*}{=}$ la hipótesis inductiva. Por tanto $\langle v_{k+1}, v_j \rangle = 0$ para $j = 1, \ldots, k$, y a esto añadimos que, por hipótesis inductiva, $\langle v_i, v_j \rangle = 0$ cuando $i, j \leq k$ e $i \neq j$. Queda probar que $\lg(v_1, \ldots, v_k, v_{k+1}) = \lg(x_1, \ldots, x_k, x_{k+1})$. Por definición $v_{k+1} = x_{k+1} - P_k(x_{k+1})$ y $P_k(x_{k+1}) \in \mathbb{F}_k = \lg(x_1, \ldots, x_k)$, mostrando que v_{k+1} es combinación de $(x_1, \ldots, x_k, x_{k+1})$. La otra parte es similar pues $x_{k+1} = v_{k+1} + P_k(x_{k+1})$ y $P_k(x_{k+1}) \in \mathbb{F}_k = \lg(x_1, \ldots, x_k)$ que también es $\lg(v_1, \ldots, v_k)$, luego $x_{k+1} \in \lg(v_1, \ldots, v_k, v_{k+1})$. ♣

Con las fórmulas de $P_k^\perp(x) = x - P_k(x)$ en (9.4) obtenemos

$$v_1 = x_1, \quad v_j = x_j - \frac{\langle x_j, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle x_j, v_2 \rangle}{\|v_2\|^2} v_2 - \ldots - \frac{\langle x_j, v_{j-2} \rangle}{\|v_{j-2}\|^2} v_{j-2} - \frac{\langle x_j, v_{j-1} \rangle}{\|v_{j-1}\|^2} v_{j-1}. \tag{9.8}$$

Hay que observar que hay una gran cantidad de sustituciones porque la fórmula de v_j no aparece directamente en función de $x_1, \ldots, x_j$ sino de $v_1, \ldots, , v_{j-1}, x_j$ y hay que sustituir $v_1, \ldots, , v_{j-1}$ por sus valores previamente calculados.

Problema 348 *En $\mathbb{R}^3$ se considera ε el producto euclidiano estándar y los vectores $x_1 = (1,0,0)^\top$, $x_2 = (1,1,0)^\top$ y $x_3 = (p,q,r)^\top$ .con $r \neq 0$. Aplicar el proceso de Gram-Schmidt a $X = (x_1, x_2, x_3)$.* ♦

Solución. Se tiene $v_1 = x_1 = (1,0,0)$ y $\|v_1\| = 1$. A continuación,[14]

$$v_2 = x_2 - \frac{\langle x_2, v_1 \rangle}{\|v_1\|^2} v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{1^2} \left\langle \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Claramente, $\|v_2\| = 1$. Finalmente,

$$\begin{aligned} v_3 &= x_3 - \frac{\langle x_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle x_3, v_2 \rangle}{\|v_2\|^2} v_2 \\ &= \begin{pmatrix} p \\ q \\ r \end{pmatrix} - \left\langle \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \left\langle \begin{pmatrix} p \\ q \\ r \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ r \end{pmatrix}. \blacklozenge \end{aligned}$$

Problema 349 *En $\mathbb{R}^2$ y $\mathbb{R}^3$ nos dan productos euclidianos ω, η con matrices en $\mathcal{E}$*

$$a = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix}.$$

Determinar las bases ortonormales que resultan al aplicar el procedimiento de Gram-Schmidt a la base estándar. ♦

[14] En cálculos concretos, por razones de espacio usaremos a veces $x \cdot y$ en vez de $\langle x, y \rangle$.

Solución (del primer caso). Se hace

$$v_1 = e_1, \quad \|v_1\|^2 = \omega(e_1, e_1) = 1,$$

$$v_2 = e_2 - \frac{\langle e_2, v_1\rangle}{\|v_1\|^2} v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \frac{1}{1}\left[(0,1)\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Hemos seguido mecánicamente los pasos pero ahorra escribir si se observa que el corchete $[\bullet]$ es $\omega(e_2, e_1) = \omega_{21} = 1$. La ortogonalización es $\left((1,0)^\top, (-1,1)^\top\right)$. El segundo caso queda para el lector. ♦

Problema 350 *Aplicar el procedimiento de Gram-Schmidt, supuesto* $(\mathbb{E}, \omega) = \left(\mathbb{R}^3, \varepsilon\right)$, *a*

$$(x_1, x_2, x_3) = \left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right).$$

Problema 351 *Probar que si en* $(\mathbb{R}^m, \varepsilon)$ *se define* $X = (x_1, \ldots, x_m)$ *por* $x_k = e_1 + \ldots + e_k$, *se tiene que* V *construido con el teorema 179 es la base estándar.Suele ser más conveniente en muchos casos utilizar la ortogonalización* $V = (v_1, \ldots, v_r)$ *de* $X = (x_1, \ldots, x_r)$ *que su ortonormalización* $W = (w_1, \ldots, w_r)$ *con* $w_j = v_j / \|v_j\|$. *Sin embargo* W *(o sea,* V *normalizada) tiene una propiedad de interés teórico y es que una sencilla condición la hace única.*

Problema 352 *Hay una sola ortogonalización* normalizada W *de* $X = (x_1, \ldots, x_r)$ *tal que para todo* j *sea* $\langle w_j, x_j\rangle > 0$.

9.6. Factorizaciones

Vamos a probar como ciertas matrices a se pueden factorizar como producto de otras dos con propiedades especiales. Esto tendra su importancia a la hora de resolver sistemas $ax = y$ numéricamente. Van en este capítulo porque utilizaran la posibilidad de ortogonalizar u ortonormalizar sucesiones de vectores. El espacio será $\mathbb{R}^m$ o $\mathbb{R}^n$ para que la matriz principal sea siempre $m \times n$. y el producto euclidiano será el estándar. *La materia de esta sección es opcional.*

9.6.1. La factorización QR

Sea $a \in \mathbb{R}^{m\times n}$ tal que la sucesión $(a_1, \ldots, a_n)$ de sus columnas sea independiente en $\mathbb{R}^m$, luego $m \geq n$. Apliquemos con el producto estándar ε de $\mathbb{R}^m$ el método de Gram-Scmidt a $(a_1, \ldots, a_n)$ para obtener una sucesión ortogonal $(u_1, \ldots, u_n)$,

$$u_1 = a_1, \quad u_j = a_j - \sum_{k=1}^{j-1} \frac{\langle a_j, u_k\rangle}{\|u_k\|^2} u_k, \ j = 2, \ldots, n.$$

Podemos poner las de $(a_1, \ldots, a_n)$ en función de $(u_1, \ldots, u_n)$ o, mejor todavía, en términos de $(q_1, \ldots, q_n)$ obtenida tras normalizar las u_j, luego $q_j = u_j / \|u_j\|$. Las fórmulas son

$$a_1 = \|u_1\| q_1, \quad a_j = \sum_{k=1}^{j-1} \langle a_j, q_k\rangle q_k + \|u_j\| q_j, \ j = 2, \ldots, n.$$

Obsérvese que al ser $(q_1, \ldots, q_n)$ familia ortonormal de n vectores en $\mathbb{R}^m$, la matriz $q = (q_1, \ldots, q_n)$ verifica $q^\top q = I_n$. Definiendo

$$r_j^k = \langle a_j, q_k\rangle \text{ si } 1 \leq k < j, \quad r_j^j = \|u_j\|, \quad r_j^k = 0 \text{ si } j < k \leq n.$$

reescribimos (permutamos para mejor ver un producto matricial $r_j^k q_k^i$ a $q_k^i r_j^k$)

$$a_1^i = \|u_1\| q_1^i, \quad a_j^i = \sum_{k=1}^{j-1} q_k^i r_j^k + q_j^i r_j^j, \ j = 2, \ldots, n. \tag{9.9}$$

Esto es como decir que $q \in \mathbb{R}^{m\times n}$ y $r = \left(r_j^k\right) \in \mathbb{R}^{n\times n}$ cumplen $a = qr$. La matriz r es triangular superior e invertible puesto que los términos de su diagonal son $r_j^j = \|u_j\| > 0$.

Teorema 180 *Dada $a \in \mathbb{R}^{m\times n}$ tal que sus columnas a_j sean independientes, existe una y solo una factorización $a = qr$, siendo $q \in \mathbb{R}^{m\times n}$ tal que $q^\top q = I_n$ y $r \in \mathbb{R}^{n\times n}$ triangular superior con todo $r_i^i > 0$.*

Demostración. Acabamos de mostrar cómo construir al menos un par q, r. La unicidad se sigue del problema 352 puesto que la ecuación (9.9) nos da $\langle a_j, q_j\rangle = r_j^j > 0$. ♣

La factorización descrita es la **factorización QR**. Es muy útil para resolver numéricamente sistemas de ecuaciones $ax = y$ donde y toma muchos valores y $a \in \mathbb{R}^{m\times n}$ tiene sus m columnas independientes, siendo en la práctica $m = n$. Se factoriza $a = qr$ y se multiplica $qrx = y$ por $q^\top$ con lo que hay que resolver $rx = q^\top y$. Las operaciones $q \to q^\top$ y $q^\top y$ son sencillas para un ordenador y resolver sistemas $rx = z$ (la incógnita es x) lo es también por el escaso número de multiplicaciones. Es por tanto económico desde el punto de vista computacional el factorizar $a = qr$, obtener los diversos $z = q^\top y$, y resolver para y los sistemas $rx = z = q^\top y$. Subrayamos que esta ventaja es para un ordenador porque para él las multiplicaciones son igual de difíciles y solo influye su número para medir la eficacia del procedimiento. Por ejemplo, si

$$a = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \text{ se tiene } q = \begin{pmatrix} 0 & \frac{1}{3}\sqrt{2}\sqrt{3} & \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{2}\sqrt{3} & \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{2}\sqrt{3} & -\frac{1}{3}\sqrt{3} \end{pmatrix},\ r = \begin{pmatrix} \sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 0 & \frac{1}{2}\sqrt{2}\sqrt{3} & \frac{1}{6}\sqrt{2}\sqrt{3} \\ 0 & 0 & \frac{2}{3}\sqrt{3} \end{pmatrix}.$$

Evidentemente, si nos piden resolver $ax = (2, 4, 6)^\top$ es mucho más fácil calcular

$$a^{-1} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix},\quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix}$$

que hacer

$$\begin{pmatrix} 0 & \frac{1}{3}\sqrt{2}\sqrt{3} & \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{2}\sqrt{3} & \frac{1}{3}\sqrt{3} \\ \frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{2}\sqrt{3} & -\frac{1}{3}\sqrt{3} \end{pmatrix}^T \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 5\sqrt{2} \\ \sqrt{2}\sqrt{3} \\ 0 \end{pmatrix}$$

y resolver

$$\begin{pmatrix} \sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 0 & \frac{1}{2}\sqrt{2}\sqrt{3} & \frac{1}{6}\sqrt{2}\sqrt{3} \\ 0 & 0 & \frac{2}{3}\sqrt{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5\sqrt{2} \\ \sqrt{2}\sqrt{3} \\ 0 \end{pmatrix},\quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix}$$

con tanta raíz cuadrada. Además, hay que calcular (lo ha hecho el ordenador) q y r. Pero lo que cuenta es que el número de operaciones para llegar a la solución es mucho menor, si hay muchos y y n es grande, aunque haya que factorizar $a = qr$.

9.6.2. La factorización de Cholesky

Sea $s \in \mathbb{R}^{n\times n}$ simétrica y definida positiva. Vimos en el teorema 159 que haciendo operaciones columna/fila podíamos transformar s en una matriz diagonal d y que si las operaciones columna correspondientes (pero no las fila) se hacían sobre I, se obtenía una matriz c invertible tal que $s = c^\top dc$. Si no pedimos que s sea definida positiva, podemos encontrarnos con uno o varios $s_i^i = 0$, lo que obliga a operaciones de permuta columna/fila. Sin embargo, al ser definida positiva, en cualquier matriz intermedia a para pasar de s hasta d diagonal, será siempre $a_i^i \neq 0$ para todo i,[15] lo que nos permitirá *restringirnos a un solo tipo de operaciones*: las que restan a la cada columna o fila una combinación lineal de las columnas o filas anteriores. Esto nos permite ver que al hacer ese tipo de operaciones columna en I llegaremos a c triangular superior con todo 1 en la diagonal.

Teorema 181 *Si s de definida positiva hay una factorización, que además es única, en la forma $s = c^\top dc$ con c triangular superior con todo 1 en la diagonal y d diagonal con todo $d_i^i > 0$.*

Demostración. Antes del teorema construimos la factorización. Supongamos $s = c_1^\top d_1 c_1 = c_2^\top d_2 c_2$ con las características expuestas. Tenemos

$$\left(c_2^\top\right)^{-1} c_1^\top d_1 = d_2 c_2 c_1^{-1},\quad \left(c_2^{-1}\right)^\top c_1^\top d_1 = d_2 c_2 c_1^{-1},\quad \left(c_1 c_2^{-1}\right)^\top d_1 = d_2 c_2 c_1^{-1}$$

[15] Tendremos $s = b^\top ab$ con b invertible y para $x = b^{-1}e_i$ se cumplirá $0 < x^\top sx = a_i^i$

La matriz $\left(c_1c_2^{-1}\right)^\top$ es triangular inferior, luego $\left(c_1c_2^{-1}\right)^\top d_1$ también lo será, Análogamente, $c_2c_1^{-1}$ es triangular superior y $d_2c_2c_1^{-1}$ también lo será. Una matriz a la vez triangular superión e inferior ha de ser diagonal, y tenemos $d_3 = \left(c_1c_2^{-1}\right)^\top d_1 = d_2c_2c_1^{-1}$, siendo d_3 diagonal. Las matriz d_2 es invertible, $c_2c_1^{-1} = d_2^{-1}d_3$ que es diagonal (producto de dos matrices diagonales. Finalmente, como c_1 y c_2 tienen todo 1 en la diagonal, $c_2c_1^{-1}$ tiene también esa propiedad, y llegamos al final porque $c_2c_1^{-1}$ es diagonal con todo 1 en la diagonal, luego $c_2c_1^{-1} = I$, $c_1 = c_2$ y $d_1 = d_2$. ♣

Teorema 182 (Factorización de Cholesky) *Si s de definida positiva hay una factorización en la forma $s = a^\top a$ con a triangular e invertible. Si hay otra factorización $s = b^\top b$ con b triangular e invertible, existe una matriz diagonal d con todo ± 1 en la diagonal tal que $b = da$.*

Demostración. Con el teorema anterior factorizamos $s = c^\top dc$ y como todo $d_i^i > 0$ definimos D diagonal con $D_i^i = \sqrt{d_i^i}$. Enseguida se ve que $a = Dc$ cumple lo requerido. Si $s = a^\top a = b^\top b$ tenemos $ba^{-1} = \left(ab^{-1}\right)^\top$ con ba^{-1} triangular superior y $\left(ab^{-1}\right)^\top$ triangular inferior, luego $ba^{-1} = d$ diagonal y $b = da$. Entonces $s = a^\top a = a^\top d^2 a$ y, al ser a y $a^\top$ invertibles resulta $d^2 = I$ y se acaba la demostración. ♣

Claramente se pedimos que en las factorización pedimos que los términos en la diagonal de las matrices triangulares sean > 0, la factorización es única. Como ejemplo,

$$s = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & \sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \sqrt{2} \end{pmatrix} = a^\top a$$

La utilidad de la factorización de Cholesky es muy parecida a la factorización QR. Si hemos de resolver el sistema $sx = y$ para muchos casos de y pero con la misma s, planteamos $a^\top ax = y$. Con el cambio $z = ax$ resolvemos $a^\top z = y$ (incógnita z) que es fácil para el ordenador ya que $a^\top$ es triangular, y luego ya con z conocida, $ax = z$ que tanbién es fácil porque a es triangular.

9.7. Distancias y desigualdades

Si en $\mathbb{E}$ tenemos un producto euclidiano, tenemos una **distancia** $d(x, y) = \|x - y\|$. La definición es muy razonable porque en $\mathbb{R}^2$ con el producto estándar,

$$d(x, y) = \sqrt{\langle x - y, x - y\rangle} = \sqrt{(x^1 - y^1)^2 + (x^2 - y^2)},$$

que es la fórmula de la distancia entre puntos del plano. Hay una fórmula análoga para $n = 3$. Veremos que muchas fórmulas y propiedades evidentes de $\mathbb{R}^2$ y $\mathbb{R}^3$ se trasladan a un espacio euclidiano $\mathbb{E}$ con un producto ω arbitrario, incluso si no tiene dimensión finita. En apariencia, si $\mathbb{E}$ es un espacio de funciones, esto no pasa de ser un juego, pero no, porque numerosas desigualdades en que hay integrales son producto de la visión "geométrica" de este $\mathbb{E}$ tan distante de $\mathbb{R}^3$.

Es muy intuitivo que si tenemos $x \in \mathbb{E}$ y lo proyectamos sobre $\mathbb{F}$ en la descomposición $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^\perp$, la proyección tendrá menos longitud que x. Importantes desigualdades se basan en esta idea tan sencilla.

Teorema 183 *Dado $x \in \mathbb{E}$ y un subespacio $\mathbb{F}$ de dimensión finita, la proyección $P(x) \in \mathbb{F}$ cumple*

1. *$\|P(x)\| \leq \|x\|$, habiendo igualdad si y solo si $x \in \mathbb{F}$.*
2. *La proyección $P(x)$ es un elemento de $\mathbb{F}$ a distancia mínima de x; o sea, tal que $\|x - P(x)\| \leq \|x - t\|$ para cualquier $t \in \mathbb{F}$. Además, $P(x)$ es el* único *vector de $\mathbb{F}$ con esta propiedad.*

Demostración. Descomponemos $x = P(x) + z$ con $z \in \mathbb{F}^\perp$. Al ser $\langle P(x), z\rangle = 0$ se sigue que

$$\|x\|^2 = \|P(x)\|^2 + \|z\|^2 + 2\langle P(x), z\rangle = \|P(x)\|^2 + \|z\|^2 \geq \|P(x)\|^2,$$

con igualdad al final si y solo si $z = 0$, equivalente a $x = P(x) \in \mathbb{F}$.

Dado $t \in \mathbb{F}$ escribimos $x - t = (x - P(x)) + (P(x) - t)$. Sucede que $x - P(x) \in \mathbb{F}^\perp$ y $P(x) - t \in \mathbb{F}$. Por tanto, al ser $\langle x - P(x), P(x) - t\rangle = 0$ resulta

$$\|x - t\|^2 = \|x - P(x)\|^2 + \|P(x) - t\|^2 \geq \|x - P(x)\|^2.$$

Si t consigue la igualdad (está a mínima distancia) ha de ser porque $P(x) = t$. ♣

La desigualdad $\|P(x)\| \leq \|x\|$ se llama la **desigualdad de Bessel**. Se suele presentar del siguiente modo: se toma una base ortogonal u ortonormal $\mathcal{V} = (v_1, \ldots, v_n)$ de $\mathbb{F}$. Si, por ejemplo, $\mathcal{V}$ es ortogonal,

$$P(x) = \sum_{j=1}^{n} \frac{\langle x, v_j \rangle}{\|v_j\|^2} v_j, \quad \|P(x)\|^2 = \sum_{j=1}^{n} \frac{\langle x, v_j \rangle^2}{\|v_j\|^2}, \quad \|x\|^2 \geq \sum_{j=1}^{n} \frac{\langle x, v_j \rangle^2}{\|v_j\|^2}$$

y la última desigualdad es la de Bessel. Si $\mathcal{V}$ es ortonormal se sustituyen todos los $\|v_j\|$ por 1.

Probamos la **desigualdad de Schwarz.** Sea $y \neq 0$ y $\mathbb{F} = \lg(y)$. Entonces $P(x) = \left(\langle x, y \rangle / \|y\|^2\right) y$ y la desigualdad de Bessel adopta las formas equivalentes

$$\|x\| \geq \|P(x)\|, \qquad \|x\|^2 \geq \frac{\langle x, y \rangle^2}{\|y\|^4} \|y\|^2 = \frac{\langle x, y \rangle^2}{\|y\|^2}, \qquad \|x\|^2 \|y\|^2 \geq \langle x, y \rangle^2.$$

La **desigualdad de Schwarz** se suele enunciar del modo siguiente

Teorema 184 *Dados $x, y \in \mathbb{E}$ se tiene que $\langle x, y \rangle^2 \leq \|x\|^2 \|y\|^2$ con igualdad si y solo si (x, y) es dependiente. De modo equivalente, $|\langle x, y \rangle| \leq \|x\| \|y\|$ con igualdad si y solo si hay dependencia lineal.*

Demostración. El caso $y \neq 0$ ya ha sido probado, y si $y = 0$, es obvia. Es obvio también que se cumple si hay dependencia lineal pues si $\alpha x + \beta y = 0$ y, digamos, $\beta \neq 0$, vemos para $y = -(\alpha/\beta) x$ que

$$\langle x, y \rangle^2 = \frac{\alpha^2}{\beta^2} \left(\|x\|^2\right)^2, \qquad \|x\|^2 \left\| -\frac{\alpha}{\beta} x \right\|^2 = \left(-\frac{\alpha}{\beta}\right)^2 \|x\|^4.$$

Solo queda ver qué pasa si la desigualdad es una igualdad. Si $y = 0$ hay dependencia lineal. Si es $y \neq 0$, la igualdad $\langle x, y \rangle^2 = \|x\|^2 \|y\|^2$ implica que

$$\|P(x)\|^2 = \left\| \frac{\langle x, y \rangle}{\|y\|^2} y \right\|^2 = \frac{\langle x, y \rangle^2}{\|y\|^4} \|y\|^2 = \|x\|^2$$

y por **1** en el teorema 183, $x \in \mathbb{F} = \lg(y)$ y hay dependencia lineal.

Si $a, b \in \mathbb{R}$ y $b \geq 0$, la condición $a^2 \leq b^2$ equivale a $|a| \leq b$. La desigualdad de Schwarz tiene la forma equivalente $|\langle x, y \rangle| \leq \|x\| \|y\|$. ♣

Muchos textos tratan la desigualdad de Schwarz antes de hablar de proyecciones ortogonales y dan demostraciones ligeramente diferentes que vamos a exponer. La desigualdad es obvia si x o y son cero. Supongamos entonces $y \neq 0$. La función

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(\lambda) = \|x - \lambda y\|^2 = \|x\|^2 - 2\lambda \langle x, y \rangle + \lambda^2 \|y\|^2$$

es ≥ 0. Por tanto, su mínimo es ≥ 0. Esto se traduce en cualquiera de las condiciones

1. El valor mínimo λ_0, solución de $f'(\lambda) = 2\lambda \|y\|^2 - 2\langle x, y \rangle = 0$, ha de dar un valor mínimo ≥ 0; o sea, para $\lambda_0 = \langle x, y \rangle / \|y\|^2$,

$$f(\lambda_0) = \|x\|^2 - 2\langle x, y \rangle \frac{\langle x, y \rangle}{\|y\|^2} + \left(\frac{\langle x, y \rangle}{\|y\|^2}\right)^2 \|y\|^2 = \|x\|^2 - \frac{\langle x, y \rangle^2}{\|y\|^2} \geq 0. \tag{9.10}$$

   Quitando denominadores resulta la desigualdad.

2. Si elegimos $\lambda_0 = \langle x, y \rangle / \|y\|^2$ (aunque no sepamos que es el mínimo) observamos que si es $f(\lambda) \geq 0$ para todo λ, será en particular $f(\lambda_0) \geq 0$ y sale la desigualdad en (9.10).

3. El grafo de $\mu = f(\lambda)$ es de la forma $y = ax^2 + bx + c$ que es una parábola. El que sea $f(\lambda) \geq 0$ equivale a decir que no puede haber dos soluciones distintas de $ax^2 + bx + c = 0$. Esto impone que sea
$$b^2 - 4ac = 4\langle x, y \rangle^2 - 4\|y\|^2 \|x\|^2 \leq 0$$
   que nos lleva a la desigualdad de Schwarz.

4. Todo lo dicho se aplica a $\omega(x,y)=\langle x,y\rangle$ que es definida positiva. Sin embargo, las ideas son aplicables a σ que sea tan solo semidefinida positiva (es decir, $\sigma(x,x)\geq 0$ pero puede ser $\sigma(x,x)=0$ con $x\neq 0$). En efecto, consideramos

$$f:\mathbb{R}\to\mathbb{R},\qquad f(\lambda)=\sigma(x-\lambda y,x-\lambda y)=\sigma(x,x)-2\lambda\sigma(x,y)+\lambda^2\sigma(y,y)$$

que es ≥ 0. Si $\sigma(y,y)=0$ (puede ser $y\neq 0$), queda $\sigma(x,x)-2\lambda\sigma(x,y)\geq 0$ para todo λ lo que implica $\sigma(x,y)=0$ y $\sigma(x,y)^2\leq\sigma(x,x)\,\sigma(y,y)$ es $0\leq 0$. Si $\sigma(y,y)\neq 0$ valen los razonamientos de **1-3**. Siempre hay desigualdad pero de la igualdad no se deduce dependencia lineal.

Teorema 185 *Dados $x,y\in\mathbb{E}$ se tienen las* ***desigualdades triangulares***

$$\|x+y\|\leq\|x\|+\|y\|,\qquad |\|x\|-\|y\||\leq\|x-y\|.$$

Supuestos x e y no nulos, la igualdad en cualquiera de ellas equivale a la proporcionalidad $x=\lambda y$ o $y=\lambda x$ con $\lambda>0$.[16] *(La condición es redundante, pues si $x=\lambda y$ con $\lambda>0$, $y=(1/\lambda)\,x$ con $1/\lambda>0$.)*

Demostración. Usaremos la desigualdad de Schwarz en la forma

$$\langle x,y\rangle\leq|\langle x,y\rangle|\leq\|x\|\,\|y\|,\qquad -\langle x,y\rangle\geq-|\langle x,y\rangle|\geq-\|x\|\,\|y\|$$

y que, entre números ≥ 0, hay una equivalencia entre $a\leq b$ y $a^2\leq b^2$. Basta comprobar que

$$\|x+y\|^2=\|x\|^2+\|y\|^2+2\langle x,y\rangle\leq\|x\|^2+\|y\|^2+2\|x\|\,\|y\|=(\|x\|+\|y\|)^2,$$

$$\|x-y\|^2=\|x\|^2+\|y\|^2-2\langle x,y\rangle\geq\|x\|^2+\|y\|^2-2\|x\|\,\|y\|=(\|x\|-\|y\|)^2=|\|x\|-\|y\||^2.$$

Dejamos para el lector comprobar que si $x=\lambda y$ o $y=\lambda x$ con $\lambda>0$ las dos desigualdades son igualdades. Vamos con la recíproca suponiendo $\|x+y\|=\|x\|+\|y\|$ con $x,y\neq 0$. La desigualdad que aparece al deducir $\|x+y\|\leq\|x\|+\|y\|$ ha de ser igualdad; o sea $\langle x,y\rangle=|\langle x,y\rangle|=\|x\|\,\|y\|$. De $|\langle x,y\rangle|=\|x\|\,\|y\|$ se sigue con 184 que $x=\lambda y$ para cierto $\lambda\in\mathbb{R}$. De $\langle x,y\rangle=|\langle x,y\rangle|$ obtenemos por sustitución $\lambda\|x\|^2=|\lambda|\,\|x\|^2$ y $\lambda>0$. El trabajo con la segunda desigualdad es análogo. ♣

Si vemos $a,b,c\in\mathbb{E}$ como los vértices de un triángulo, aplicando las desigualdades triangulares a $x=a-b$ e $y=b-c$ obtenemos

$$\|a-c\|=\|(a-b)+(b-c)\|\leq\|a-b\|+\|b-c\|,\qquad d(a,c)\leq d(a,b)+d(b,c),$$

$$|\|a-b\|-\|b-c\||\leq\|(a-b)-(b-c)\|=\|a-c\|,\qquad |d(a,b)-d(b,c)|\leq d(a,c).$$

Exponen algo muy intuitivo: en un triángulo la suma de longitudes de dos lados supera al tercero y la diferencia de longitudes (en valor absoluto) es menor que la del tercero. De ahí deriva el nombre de desigualdades triangulares.

Problema 353 *Probar con la desigualdad de Schwarz que si $a_1,a_2,\ldots,a_n,b_1,b_2,\ldots,b_n\in\mathbb{R}$ se tiene que*

$$\left(\sum_{i=1}^n a_ib_i\right)^2\leq\sum_{i=1}^n(a_i)^2\sum_{i=1}^n(b_i)^2,\qquad \left(\sum_{i=1}^n a_i\right)^2\leq m\sum_{i=1}^n(a_i)^2,$$

obteniendo como caso derivado de la segunda fórmula que "el cuadado de la media es menor que la media de los cuadrados"; esto es,

$$\left(\frac{1}{m}\sum_{i=1}^n a_i\right)^2\leq\frac{1}{m}\sum_{i=1}^n(a_i)^2,\ \textit{con el caso particular}\ \left(\frac{a+b}{2}\right)^2\leq\frac{a^2+b^2}{2}.$$

Probar también que

$$\left(\sum_{i=1}^n(a_i+b_i)^2\right)^{1/2}=\left(\sum_{i=1}^n(a_i)^2\right)^{1/2}+\left(\sum_{i=1}^n(b_i)^2\right)^{1/2}.$$

[16] En un espacio con $\Bbbk=\mathbb{R}$, donde $\lambda>0$ tiene sentido, la **semirrecta abierta** generada por $v\neq 0$ es el conjunto $\{\lambda v\mid\lambda>0\}$. Con esto se puede decir que la igualdad de cualquiera de las dos desigualdades equivale a que x e y estén en una misma semirrecta. La definición de **semirrecta cerrada** es análoga con $\lambda\geq 0$ (contiene a 0).

Problema 354 *Para funciones continuas de* $[0,1]$ *en* $\mathbb{R}$*, ¿de donde sale* $\left(\int_0^1 f(t)\,dt\right)^2 \leq \left(\int_0^1 f(t)^2\,dt\right)$*?*

Problema 355 *Sea* $f:[a,b]\to\mathbb{R}$ *continua y* estrictamente positiva. *Probar la desigualdad*

$$(a-b)^2 \leq \int_a^b f(t)\,dt \cdot \int_a^b \frac{1}{f(t)}dt.$$

¿Cuándo hay igualdad?

Volvemos al asunto de las distancias. Si estamos en $\mathbb{E}$ de dimensión infinita y $\mathbb{F}$ es un subespacio de dimensión finita m, como tenemos la descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$ del teorema 176, podemos definir la **distancia del punto** x **al subespacio** $\mathbb{F}$ como $d(x,\mathbb{F})=d(x,P(x))=\|x-P(x)\|$; o sea, la distancia de x a su proyección sobre $\mathbb{F}$. Esto es muy intuitivo pero plantea dos cuestiones. La primera es que si $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$ no existe (y en el problema 345 hay un ejemplo de ello) no tenemos P y $d(x,\mathbb{F})$ *no está definida*. Lo señalamos, pero para nosotros no es muy grave, porque en este curso nos centramos en espacios de dimensión finita. Volveremos sobre esto más adelante (problema 357 y lo que antecede). Otra cuestión es que para calcular $P(x)$ hay que disponer de una base ortogonal de $\mathbb{F}$ (que se puede conseguir con el procedimiento de Gram-Schmidt pero supone un trabajo). Sin embargo ya tenemos una fórmula si se dispone al menos de una base $\mathcal{B}=(b_1,\ldots,b_m)$ de $\mathbb{F}$ del tipo que sea. Se ha hecho el trabajo en el teorema 177 porque al descomponer $x=y+z$ de acuerdo con $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$ se tiene $d(x,\mathbb{F})^2=\|x-P(x)\|^2=\|z\|^2$ y ya se vio en ese teorema que

$$d(x,\mathbb{F})^2=\|z\|^2=\frac{\Gamma(x,b_1,\ldots,b_m)}{\Gamma(b_1,\ldots,b_m)}. \tag{9.11}$$

Siempre cabe la posibilidad, que quizás no sea tan desventajosa, de calcular primero la proyección $P(x)$ en la forma $P(x)=\sum_{i=1}^m \lambda^i b_i$ sabiendo que $\left(\lambda^1,\ldots,\lambda^m\right)$ es solución de

$$G(b_1,\ldots,b_m)\begin{pmatrix}\lambda^1\\ \vdots\\ \lambda^m\end{pmatrix}=\begin{pmatrix}\langle b_1,b_1\rangle & \cdots & \langle b_1,b_m\rangle\\ \vdots & \ddots & \vdots\\ \langle b_m,b_1\rangle & \cdots & \langle b_m,b_m\rangle\end{pmatrix}\begin{pmatrix}\lambda^1\\ \vdots\\ \lambda^m\end{pmatrix}=\begin{pmatrix}\langle b_1,x\rangle\\ \vdots\\ \langle b_m,x\rangle\end{pmatrix}$$

y luego calcular $d(x,\mathbb{F})^2=\|x-P(x)\|^2$. Hay una simplificación importante si estamos en $\mathbb{R}^n$ con el producto estándar ε y nos dan $\mathbb{F}\subset\mathbb{R}^n$ en forma implícita por las ecuaciones *independientes*

$$\begin{pmatrix}a_1^1 & a_2^1 & \cdots & a_{n-1}^1 & a_n^1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ a_1^m & a_2^m & \cdots & a_{n-1}^m & a_n^m\end{pmatrix}\begin{pmatrix}x^1\\ \vdots\\ x^n\end{pmatrix}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}.$$

Los vectores $b_i=\left(a^i\right)^\top\in\mathbb{R}^n$ forman base de $\mathbb{F}^\perp$ y la proyección $y=Q(x)$ de x sobre $\mathbb{F}^\perp$ cumple que $d(x,\mathbb{F})^2=\|y\|^2$.

Problema 356 *Se considera en* $\mathbb{R}^5$ *con el producto estándar* ε *el subespacio* $\mathbb{F}$ *de ecuaciones*

$$\begin{cases}x^1+2x^3-x^4-x^5=0\\ x^1+x^2+x^3-3x^5=0\end{cases}.$$

Calcular la distancia de $x=(1,0,0,0,2)^\top$ *a* $\mathbb{F}$.

El siguiente problema muestra que, si no se tiene una descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$, la distancia $d(x,\mathbb{F})$, *que habría que definir,* puede tener propiedades extrañas. La definición más razonable es

$$d(x,\mathbb{F})=\text{ínf}\left\{d(x,y)\mid y\in\mathbb{F}\right\}.$$

Recordemos que en el problema 345 se probó que si $\mathbb{E}$ es el espacio de las funciones continuas de $[0,1]$ en $\mathbb{R}$ y $\mathbb{F}$ el subespacio de las que cumplen $f(0)=0$, con el producto $\omega(f(t),g(t))=\int_0^1 f(t)\,g(t)\,dt$ se tenía que $\mathbb{F}^\perp=0$. Al ser $\mathbb{F}\neq\mathbb{E}$, es imposible tener $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$.

Problema 357 *Probar que, aunque la función constante* 1 *no está en* $\mathbb{F}$, $d(1,\mathbb{F})=0$. ♦

Solución. Basta con que encontremos una sucesión de funciones f_n en $\mathbb{F}$ de modo que $d(1,f_n)\to 0$ si $n\to\infty$. Describimos f_n por su grafo. Está formado por el segmento que une $(0,0)$ con $(1/n,1)$ y el segmento horizontal que une $(1/n,1)$ con $(1,1)$. Es fácil ver que $(1-f_n(t))\geq(1-f_n(t))^2$ y que por tanto

$$\|1-f_n\|^2=\int_0^1(1-f_n(t))^2\,dt\leq\int_0^1(1-f_n(t))\,dt=\frac{1}{2n}$$

de modo que, como $f_n\in\mathbb{F}$ porque $f_n(0)=0$, se llega a que $d(1,\mathbb{F})=0$. ♦

Teorema 186 *Se tiene para* $(x_1,\ldots,x_r)$ *vectores independientes de* $\mathbb{E}$ *que*

$$\Gamma(x_1,\ldots,x_r)\leq\|x_1\|^2\cdots\|x_k\|^2\,\Gamma(x_{k+1},\ldots,x_r),\ \textit{y en particular}\ \Gamma(x_1,\ldots,x_r)\leq\|x_1\|^2\cdots\|x_r\|^2,$$

que es la ***desigualdad de Hadamard****. Es una igualdad si y solo si los vectores son ortogonales.*

Demostración. Hay una inducción más o menos explícita. Sea z_k la proyección sobre $\lg(x_{k+1},\ldots,x_r)^\perp$ de x_k. Con (9.11) deducimos que para $k=1,\ldots,r-1$,

$$\|z_k\|^2=d(x_k,\lg(x_{k+1},\ldots,x_r))^2=\frac{\Gamma(x_k,x_{k+1},\ldots,x_r)}{\Gamma(x_{k+1},\ldots,x_r)}\leq\|x_k\|^2,$$

con la desigualdad justificada por la de Bessel. Usando estas desigualdades,

$$\begin{aligned}\Gamma(x_1,\ldots,x_r)&=\|x_1\|^2\,\Gamma(x_2,x_3,\ldots,x_r)\leq\|x_1\|^2\|x_2\|^2\,\Gamma(x_3,x_4,\ldots,x_r)\\&\leq\ldots\leq\|x_1\|^2\|x_2\|^2\cdots\|x_{r-1}\|^2\,\Gamma(x_r)=\|x_1\|^2\|x_2\|^2\cdots\|x_{r-1}\|^2\|x_r\|^2,\end{aligned}$$

que es la desigualdad de Hadamard. Si $\Gamma(x_1,\ldots,x_r)=\|x_1\|^2\cdots\|x_r\|^2$, todas las desigualdades intermedias han de ser igualdades y $\|z_k\|^2=\|x_k\|^2$ para todo k y, por la desigualdad de Bessel, $z_k=x_k$ para todo k. Como $z_k\in\lg(x_{k+1},\ldots,x_r)^\perp=\lg(z_{k+1},\ldots,z_r)^\perp$, los $x_k=z_k$ son ortogonales entre sí. ♣

Hay otra demostración de la desigualdad de Hadamard para matrices quizás más sencilla.

Problema 358 *Sea* $a\in\mathbb{R}^{n\times n}$ *con columnas* $a_1,\ldots,a_n$. *Probar la* ***desigualdad de Hadamard***

$$\det(a)^2=\det(a_1,\ldots,a_n)^2\leq\|a_1\|^2\|a_2\|^2\cdots\|a_n\|^2,$$

habiendo igualdad si y solo si las columnas forman un conjunto ortogonal para el producto estándar. Indicación: por el teorema 179 obtenemos un conjunto $\{v_1,\ldots,v_m\}$ *de vectores ortogonales tales que*

$$a_k=v_k+\sum_{i=1}^{k-1}h_k^iv_i=v_k+P_k(a_k),\qquad 1\leq k\leq n,$$

siendo para cada k, $\lg(a_1,\ldots,a_k)=\lg(v_1,\ldots,v_k)$. *Sustituimos y usamos el problema 327.*

Cerramos la sección con un breve comentario en relación con el Análisis. Las desigualdades que hemos visto dicen cosas fáciles de intuir si $\mathbb{E}$ es de dimensión finita, pero donde aparecen aplicaciones sorprendentes es cuando $\mathbb{E}$ es de dimensión infinita, lo que significa en la práctica que $\mathbb{E}$ es un espacio de funciones. Si por ejemplo $\mathbb{E}$ es el espacio de funciones continuas $f:[-\pi,\pi]\to\mathbb{R}$, tenemos en $\mathbb{E}$ un producto euclidiano

$$\langle f(t),g(t)\rangle=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\,g(t)\,dt.$$

El factor $1/2\pi$ se introduce para que las funciones 1, $\cos nt$ y $\sin nt$ con $n\in\mathbb{N}$ formen un conjunto ortonormal. Esto se tiene verificando igualdades, que aceptaremos

$$\int_{-\pi}^{\pi}1^2dt=\int_{-\pi}^{\pi}\cos^2 ntdt=\int_{-\pi}^{\pi}\sin^2 ntdt=2\pi,\quad\int_{-\pi}^{\pi}\cos nt\sin ntdt=\int_{-\pi}^{\pi}\cos ntdt=\int_{-\pi}^{\pi}\sin ntdt=0.$$

Tomemos $f(t) = t$. Calculamos integrando por partes

$$2\pi \langle t, \operatorname{sin} nt\rangle = \int_{-\pi}^{\pi} t \operatorname{sin} nt dt = -\frac{1}{n}\int_{-\pi}^{\pi} t d(\cos nt) = -\frac{1}{n}\left[t\cos nt\right]_{t=-\pi}^{t=\pi} - \frac{1}{n}\int_{-\pi}^{\pi}\cos nt dt$$
$$= \frac{(-1)^{n+1}2\pi}{n} - \frac{1}{n^2}\left[\operatorname{sin} nt\right]_{t=-\pi}^{t=\pi} = \frac{(-1)^{n+1}2\pi}{n}.$$

Si $\mathbb{F}_k$ es el subespacio generado por $\operatorname{sin} t, \operatorname{sin} 2t, \ldots, \operatorname{sin} kt$, se tiene con la desigualdad de Bessel que $\|P(f(t))\| \leq \|f(t)\|$ (el $x \in \mathbb{E}$ general es en este caso la función $f(t) = t$). Pero

$$\|f(t)\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} t^2 dt = \frac{1}{3}\pi^2, \quad \langle f(t), \operatorname{sin} nt\rangle = \frac{(-1)^{n+1}}{n}$$

luego los coeficientes x^i en la fórmula de la proyección $P(x) = \sum_{i=1}^{k}\langle x, w_i\rangle w_i$ son $(-1)^{n+1}/n$. Queda

$$\|x\|^2 \geq \sum_{i=1}^{k}\langle x, w_i\rangle^2 \text{; que en nuestro caso es } \frac{1}{3}\pi^2 \geq \sum_{i=1}^{k}\frac{1}{n^2}.$$

En un primer curso de Análisis habrá entrado el lector en contacto con la serie $\sum_{n=0}^{\infty} 1/n^2$ y habrá visto que, para probar que es convergente se necesita que haya un constante C tal que $\sum_{n=0}^{k} 1/n^2 \leq C$ en suyo caso la suma es $\leq C$. Pues bien, nuestro trabajo muestra que puede tomarse $C = \pi^2/3 \approx 3{,}29$.

No seguiremos este camino, que es más propio de los cursos de Análisis, pero aprovechamos para señalar la misteriosa aparición del número π en un mundo en apariencia muy distante de la Geometría. El lector debe saber también que aunque en este capítulo estamos considerando siempre $\mathbb{k} = \mathbb{R}$, el caso $\mathbb{k} = \mathbb{C}$ es igual de importante (quizás más), y que los conceptos que aquí han aparecido y aparecerán (productos euclidianos, bases ortonormales, desigualdades, matrices ortogonales, isometrías, etc.) tienen un análogo si $\mathbb{E}$ es un espacio *complejo, y para el Análisis, posiblemente de dimensión infinita.*

9.8. Cuestiones diversas

9.8.1. Ángulos no orientados

No hemos definido la noción de **ángulo**, con más complicaciones técnicas de lo que parece. Se intuye que si se preservan distancias se preservan ángulos y que la recíproca no es cierta (pensar en $L(x) = 2x$). Pero ¿qué es un ángulo y cómo se mide? *No vamos a tratar* los ángulos como "zonas del plano entre dos semirrectas que confluyen en un punto V que se llama vértice". Trataremos los ángulos como "números"[17] pero no va a ser lo mismo el ángulo entre dos vectores que el ángulo de una rotación. Si pensamos tan solo en vectores x, y, no es (intuitivamente) lo mismo considerar el ángulo entre vectores $\angle(x, y)$ de modo que influya el orden de los vectores o que no influya.[18] Digamos brevemente sin entrar en los complejos detalles que, en un espacio euclidiano, se pueden definir **ángulos no orientados** (será $\angle(x, y) = \angle(y, x)$) y que, *exclusivamente* si $\dim \mathbb{E} = 2$ y hemos añadido una estructura llamada **orientación**, podemos tratar **ángulos orientados** (será $\angle(x, y) = -\angle(y, x)$). Vamos ahora a tratar tan solo los ángulos en un espacio euclidiano $\mathbb{E}$ *de cualquier dimensión finita* pero sin orientación, que es el caso más sencillo y que, aunque no vale para definir el ángulo de una rotación, sí que sirve para dar una idea de "cuánto mueve una rotación". Repetimos que, aunque muchos ejemplos y problemas van en dimensión 2, trabajamos en dimensión arbitraria.

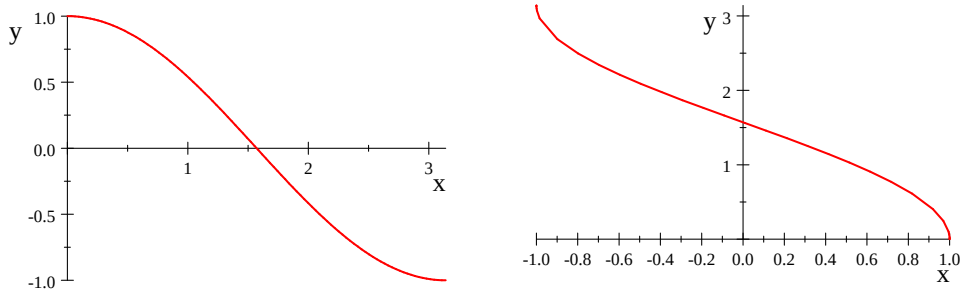
El lector tiene que saber que la *restricción* de la función coseno a $[0, \pi]$, el intervalo cerrado de extremos 0 y π es una función estrictamente decreciente que lleva biyectivamente $[0, \pi]$ en $[-1, 1]$. Su inversa, de $[-1, 1]$ en $[0, \pi]$ es por definición la función **arco coseno**, luego

$$\operatorname{arc\,cos} : [-1, 1] \longrightarrow [0, \pi], \qquad \cos\theta = t \iff \operatorname{arc\,cos} t = \theta,$$

siendo válida la equivalencia si se cuenta implícitamente con $t \in [-1, 1]$, $\theta \in [0, \pi]$. Conviene tomar nota

[17] Se matiza hablando de la **medida del ángulo**.
[18] Si $x = (1, 0)$ e $y = (0, 1)$ podemos pensar que $\angle(x, y) = \pi/2$ y $\angle(y, x) = -\pi/2$. ¿Es esto válido?

de los grafos del coseno y arco coseno



Sean $x, y \in \mathbb{E}$ *no nulos*. La desigualdad de Schwartz nos da $|\langle x, y\rangle| \leq \|x\| \, \|y\|$, equivalente a

$$-1 \leq \left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle \leq 1$$

Se define el **ángulo (no orientado) entre los vectores** x e y como el *único número* $\theta = \angle(x, y)$ *en* $[0, \pi]$ dado por

$$\cos\theta = \cos\left(\angle(x, y)\right) = \left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle = \frac{\langle x, y\rangle}{\|x\| \, \|y\|} \quad \text{equivalente a} \quad \theta = \angle(x, y) = \arccos \frac{\langle x, y\rangle}{\|x\| \, \|y\|}.$$

Describimos la forma de visualizar $\theta = \angle(x, y)$, *cuando estamos en* $\mathbb{R}^2$ con el producto estándar. Se normalizan x e y a $u = x/\|x\|$ y $v = y/\|y\|$ y se traza el círculo $\mathcal{C}$ de centro el origen y radio 1, siendo obvio que $u, v \in \mathcal{C}$. La longitud de $\mathcal{C}$ es $2\pi \approx 6{,}28$. Los puntos u y v dividen a $\mathcal{C}$ en dos arcos, que llamamos $\mathcal{C}_1$ y $\mathcal{C}_2$ y, excepto si $v = -u$, hay uno más corto, digamos $\mathcal{C}_1$, cuya longitud será menor que π y esta longitud será el ángulo. Desde luego, si $u = -v$, las longitudes de $\mathcal{C}_1$ y $\mathcal{C}_2$ son π, que es el ángulo. Con esto se ve sin esfuerzo que $\angle(e_1, e_2) = \pi/2$ y $\angle(e_1, e_1 + e_2) = \pi/4$.

El lector puede hacer unas comprobaciones muy sencillas (visuales y con cálculos)

1. El ángulo está en $[0, \pi]$ y no varía al sustituir x e y por λx y μy *siendo* $\lambda, \mu > 0$. Se puede decir por tanto que se mide el ángulo de las semirrectas que determinan x e y.

2. $\angle(x, y) = \angle(y, x)$ luego no depende del orden de los vectores. La condición $\langle x, y\rangle = 0$ equivale a que sea $\angle(x, y) = \pi/2$.

3. Si x, y son unitarios y proyectamos x sobre la recta que genera y, la proyección es $P(x) = \langle x, y\rangle y$, luego $\langle x, y\rangle \geq 0$ equivale a decir que $P(x)$ apunta en la misma dirección que y y $\langle x, y\rangle \leq 0$ equivale a que apunten $P(x)$ e y en sentido opuesto. El ángulo $\theta = \angle(x, P(x))$ viene dado por

$$\cos\theta = \frac{\langle x, \langle x, y\rangle y\rangle}{\|\langle x, y\rangle y\|} = \frac{\langle x, y\rangle^2}{|\langle x, y\rangle|} = |\langle x, y\rangle|.$$

   Dado que $|\langle x, y\rangle| \geq 0$, siempre será $\theta \leq \pi/2$. Si es $\langle x, y\rangle \geq 0$, que es cuando y y $P(x)$ apuntan en la misma dirección se tiene como condición equivalente $\angle(x, P(x)) = \angle(x, y)$.

4. Dada $\theta \in [0, \pi]$ hay en $[0, 2\pi]$ dos soluciones φ de $\cos\theta = \cos\varphi$, simétricas[19] respecto a π, que son $\varphi = \theta$ y $\varphi = 2\pi - \theta$. Si no se restringe el intervalo donde esté φ, $\cos\theta = \cos\varphi$ *no implica* $\theta = \varphi$.

5. Si tenemos $x, z \neq 0$ con z ortogonal a x, los vectores $y_1 = x + z$ e $y_2 = x - z$ forman el mismo ángulo con x.

Problema 359 *Hacer las comprobaciones que faltan en los puntos anteriores. Si en* ***5*** *se tiene* $\|x\| = \|z\| = 1$, *¿qué ángulo es ese?*

Problema 360 *Calcular el ángulo entre* ***(a)*** $x = (1, 0)$ *e* $y = (1, 1)$; ***(b)*** $x = (1, 1)$ *e* $y = (-1, 1)$; ***(c)*** x *y* $-x$; *y* ***(d)*** $x = (\cos\alpha, \sin\alpha)$ *e* $y = (\cos\beta, \sin\beta)$ *con* $0 \leq \alpha \leq \beta \leq 2\pi$. *Se supone en todos los casos que estamos en* $\mathbb{R}^2$ *con el producto estándar. Si el producto es otro, digamos* ω, *¿cuánto vale* $\angle(x, -x)$? *¿es cierto que* $\angle(e_1, e_2) = \pi/2$ *para cualquier* ω?

[19] Repetimos, aunque ya se citó al hablar de cuádricas, que la simetría en $\mathbb{E}$ respecto de $c \in \mathbb{E}$ es $Z(x) = 2c - x$. Los números θ, φ son simétricos respecto a $c = \pi/2$ si y solo si $\varphi = \pi - \theta$.

Problema 361 *Dado $\theta \in [0,\pi]$, $\theta \neq 0,\pi$, probar que se puede elegir un producto escalar ω en $\mathbb{E}=\mathbb{R}^2$ tal que los vectores e_1, e_2 de la base estándar formen ángulo θ.*

9.8.2. Semiespacios y semiplanos

Sea $\mathbb{H}$ un hiperplano del espacio euclidiano $\mathbb{E}$ de dimensión finita n, luego, por definición, $\dim(\mathbb{H}) = n-1$. Claramente $\dim\left(\mathbb{H}^{\perp}\right)=1$, luego $\mathbb{H}^{\perp}$ es una recta donde fijaremos $v\neq 0$ y $\mathbb{H}=\{x\in\mathbb{E}\mid\langle x,v\rangle=0\}$. El subconjunto $\mathbb{E}-\mathbb{H}$ tiene dos partes que llamaremos **semiplanos** o **semiespacios**[20] y vienen dados por

$$\mathbb{H}_v^+=\{x\in\mathbb{E}\mid\langle x,v\rangle>0\} \qquad \text{y} \qquad \mathbb{H}_v^-=\{x\in\mathbb{E}\mid\langle x,v\rangle<0\}.$$

Se puede decir, y ayuda a la intuición, que $x\in\mathbb{H}_v^+$ equivale a que v y x formen ángulo $<\pi/2$ y que $x\in\mathbb{H}_v^-$ equivale a que v y x formen ángulo $>\pi/2$. Los puntos $x\in\mathbb{H}$ forman con v ángulo de $\pi/2$. Hay que subrayar que hay infinitas maneras de elegir $v\neq 0$ en $\mathbb{H}^{\perp}$. Si se cambia v por $u=\lambda v$ con $\lambda>0$, se tiene $\mathbb{H}_v^+=\mathbb{H}_u^+$ y $\mathbb{H}_v^-=\mathbb{H}_v^-$ pero si se cambia v por $u=\lambda v$ con $\lambda<0$ se invierte la situación porque $\mathbb{H}_v^+=\mathbb{H}_u^-$ y $\mathbb{H}_v^-=\mathbb{H}_v^+$. Informalmente diremos que se puede dividir el plano o espacio en semiplanos y semiespacios, pero no hay un "lado preferente". Solo si hubiera una "característica preferente" para tomar $v\in\mathbb{H}^{\perp}$ se tendría un "lado preferente" de $\mathbb{H}$.

Problema 362 *Sea $v\neq 0$ y $\mathbb{H}=\{x\in\mathbb{E}\mid\langle x,v\rangle=0\}$. Se considera la simetría S respecto de $\mathbb{H}$. probar que si x está en un semiplano o semiespacio, $S(x)$ está al otro lado de $\mathbb{H}$.*

Problema 363 *Dado x en el hiperplano $\mathbb{H}$ y v no nulo ortogonal a $\mathbb{H}$, probar que $x+v$ y $x-v$ están en distintos semiespacios y $\angle(x,x+v)=\angle(x,x-v)$.*

9.8.3. Aproximaciones óptimas

Esta sección es optativa. Si $\mathbb{F}$ es un subespacio de $\mathbb{E}$ y $x\in\mathbb{E}$ diremos que $y\in\mathbb{E}$ es **aproximación óptima** o **mejor aproximación a x en $\mathbb{F}$** si $y\in\mathbb{F}$ y para todo $z\in\mathbb{F}$ se tiene $\|x-y\|\leq\|x-z\|$. Es pues un punto de $\mathbb{F}$ cuya distancia a x es menor que la de cualquier otro punto de $\mathbb{F}$. Sabemos por el teorema 183 que si $\mathbb{F}$ tiene dimensión finita (aunque no la tenga $\mathbb{E}$) la proyección $P(x)$ de x en $\mathbb{F}$ es la aproximación óptima. Si $\mathbb{F}$ tiene dimensión infinita, el tema es mucho más complicado, pero se pueden probar algunas cosas como que la aproximación óptima es única, si existe.

Teorema 187 *Si $y\in\mathbb{F}$ es la aproximación óptima de x se tiene que $x-y$ es perpendicular a $\mathbb{F}$. La aproximación óptima, si existe, es única.*

Demostración. Sea $v\neq 0$ en $\mathbb{F}$. Para $\lambda\in\mathbb{R}$, la función $\phi(\lambda)=\|x-y+\lambda v\|^2-\|x-y\|^2$ es ≥ 0 porque y es una aproximación óptima. Calculamos

$$\phi'(\lambda)=\frac{d}{d\lambda}\left(\|x-y\|^2+2\lambda\langle x-y,v\rangle+\lambda^2\|v\|^2\right)=2\langle x-y,v\rangle+2\lambda\|v\|^2, \quad \phi''(\lambda)=2\|v\|^2>0,$$

que nos da un mínimo para

$$\lambda_0=-\frac{\langle x-y,v\rangle}{\|v\|^2}.$$

Si no fuese $\langle x-y,v\rangle=0$ llegaríamos a la contradicción

$$\phi(\lambda_0)=\left(-\frac{\langle x-y,v\rangle}{\|v\|^2}\right)^2\|v\|^2+2\left(-\frac{\langle x-y,v\rangle}{\|v\|^2}\right)\langle x-y,v\rangle=-\frac{\langle x-y,v\rangle^2}{\|v\|^2}<0.$$

Si y_1,y_2 son aproximaciones óptimas a x se tiene

$$\|x-y_1\|^2=\|x-y_2+y_2-y_1\|^2\overset{*}{=}\|x-y_2\|^2+\|y_2-y_1\|^2 \quad \text{y} \quad \|x-y_1\|^2=\|x-y_2\|^2,$$

de donde deducimos que $\|y_2-y_1\|^2=0$ e $y_2=y_1$. En $\overset{*}{=}$ se usó la primera parte. ♣

Si tenemos una descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{G}$ tenemos una proyección P de $\mathbb{E}$ en $\mathbb{F}$ independientemente de que haya o no en $\mathbb{E}$ un producto euclidiano. Si existe también $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^{\perp}$ tenemos también la proyección *ortogonal* P_0. Esta tiene la propiedad de que para todo $x\in\mathbb{E}$ se cumple que $\|P_0(x)\|\leq\|x\|$. Pues bien, es *la única proyección con tal propiedad*, al menos en dimensión finita.

[20] Para $n\geq 3$ se usa **semiespacios**.

Problema 364 *Con las notaciones e hipótesis precedentes, si P cumple $\|P(x)\| \leq \|x\|$ para todo $x \in \mathbb{E}$ es porque $\mathbb{G} = \mathbb{F}^{\perp}$.*

9.8.4. El criterio de Sylvester

Vamos a dar un teorema, el **criterio de Sylvester**, que da condiciones necesarias y suficientes para determinar si una función bilineal simétrica σ en $\mathbb{E}$ es un producto euclidiano. Esto se puede hacer conociendo su matriz a en una base cualquiera $\mathcal{U}$. Dada una matriz $a \in \Bbbk^{n\times n}$, las **submatrices principales** son las de la forma

$$a_{(1)} = (a_{11}), \qquad a_{(2)} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ldots \quad a_{(k)} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix} \ldots .$$

Teorema 188 *Sea $\sigma : \mathbb{E} \times \mathbb{E} \to \mathbb{R}$ bilineal simétrica. Son propiedades equivalentes*

1. *σ es definida positiva; o sea, es un producto euclidiano.*
2. *En* cualquier *base $\mathcal{U}$, la matriz de σ tiene los determinantes Δ_n de las submatrices principales > 0.*
3. *En* alguna *base $\mathcal{U}$, la matriz de σ tiene los determinantes Δ_n de las submatrices principales > 0.*

Demostración. Si σ es un producto euclidiano definimos $\mathbb{F}_k = \lg(u_1, \ldots, u_k)$ y σ_k la restricción a $\mathbb{F}_k$. La matriz de σ_k en $\mathcal{U}_k = (u_1, \ldots, u_k)$ es la submatriz principal $a_{(k)}$, que es de hecho una matriz de Gram $G(u_1, \ldots, u_k)$ con determinante $\Gamma(u_1, \ldots, u_k) > 0$ (teorema 169). Queda visto que **1** $\implies$ **2** y es obvio que **2** $\implies$ **3**.

Aunque σ no sea un producto euclidiano, las matrices s y s' de σ en bases $\mathcal{U}$ y $\mathcal{U}'$ están relacionadas por $s' = c^{\top} sc$, siendo c una matriz de cambio de base y por tanto invertible. Obviamente $\det(s') = \det(s)\det(c)^2$ luego, si bien asignando a σ el determinante de s hacemos algo que depende de la base, sí es cierto que el *signo* de $\det(s)$ solo depende de σ. Esto será esencial en la demostración de **3** $\implies$ **1**, que probaremos por inducción sobre $n = \dim(\mathbb{E})$.

El caso $n = 1$ es trivial. Suponemos que se cumple **3** $\implies$ **1** si la dimensión del espacio es $< n$ y lo probamos para $\dim(\mathbb{E}) = n$. Con la base $\mathcal{U} = (u_1, \ldots, u_n)$ que cumple **3**, tomamos $\mathbb{F}_{n-1} = \lg(u_1, \ldots, u_{n-1})$ y $\sigma_{(n-1)}$ que es la restricción de σ a $\mathbb{F}_{(n-1)}$. Como las matrices $a_{(1)}, \ldots, a_{(n-1)}$ tienen determinante > 0, deducimos con la hipótesis inductiva que $\sigma_{(n-1)}$ es un producto euclidiano. Hay pues una base ortonormal $(v_1, \ldots, v_{n-1})$ de $\mathbb{F}_{(n-1)}$. Trabajamos con la base $\mathcal{B} = (v_1, \ldots, v_{n-1}, u_n)$ que no es ortonormal porque $\sigma(u_i, u_n) = h_i$, $i = 1, \ldots, , n$, no tiene por qué ser δ_{in}. De hecho,

$$b = \mathrm{mat}_{\mathcal{BB}}(\sigma) = \begin{pmatrix} 1 & 0 & \cdots & 0 & h_1 \\ 0 & 1 & \cdots & 0 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & h_{n-1} \\ h_1 & h_2 & \cdots & h_{n-1} & h_n \end{pmatrix}.$$

Vamos a reducir a b a su forma diagonal. A la última columna se le resta la primera multiplicada por h_1, luego la segunda por h_2, ... , hasta la $n-1$ por h_{n-1} llegando a b'. A continuación se hacen las operaciones homólogas por filas. El proceso es

$$b \to b' = \begin{pmatrix} 1 & 0 & \cdots & 0 & h_1 - h_1 \cdot 1 \\ 0 & 1 & \cdots & 0 & h_2 - h_2 \cdot 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & h_{n-1} - h_{n-1} \cdot 1 \\ h_1 & h_2 & \cdots & h_{n-1} & h_n - \sum_{i=1}^{n-1} h_i^2 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & h_n - \sum_{i=1}^{n-1} h_i^2 \end{pmatrix} = d$$

Se ha reducido b a forma diagonal d con $H = h_n - \sum_{i=1}^{n-1} h_i^2$ en la esquina inferior derecha. Claramente σ será un producto euclidiano si y solo si $H > 0$. Sucede que b, b' y d son matrices congruentes y los signos de sus determinantes son los mismos, luego $H = \det(d)$ y $\mathrm{sg}(H) = \mathrm{sg}\det(d) = \mathrm{sg}\det(b) > 0$. ♣

Problema 365 *En* $\mathbb{R}^4$ *se considera* σ *que en la base estándar tiene matriz*

$$a = \begin{pmatrix} 1 & 2 & 0 & 2 \\ 2 & 5 & 2 & 0 \\ 0 & 2 & p & 1 \\ 2 & 0 & 1 & p \end{pmatrix}.$$

Determinar para que valores de $p \in \mathbb{R}$ *es un producto euclidiano.* ♦

Solución. Los determinantes Δ_k son $1, 1, p-4$ y $p^2-24p-1$. Para que σ sea un producto escalar se necesita $p > 4$ y $p^2-24p-1 > 0$. Las raíces $r < s$ de $p^2-24p-1$ son $12 \pm \sqrt{145}$ y si $p \notin [r, s]$ es $\Delta_4 > 0$. No obstante, necesitamos también $\Delta_3 = p-4 > 0$ y $4 < s$, luego $p > s$ es la condición de que σ sea producto escalar. ♦

Problema 366 *En* $\mathbb{R}^3$ *se consideran formas* σ, τ *simétricas con matrices*

$$\begin{pmatrix} 1 & h & 0 \\ h & 1 & -h \\ 0 & -h & 1 \end{pmatrix} \quad y \quad \begin{pmatrix} 1 & 0 & p \\ 0 & p & 0 \\ p & 0 & 1 \end{pmatrix}$$

siendo $h, p \in \mathbb{R}$. *Determinar todos los valores de* h *y* p *para que sean productos euclidianos.*

Problema 367 *Probar, supuesto* $h > 0$, *los valores de* k *para que* σ *con matriz*

$$\begin{pmatrix} h+k & k & \cdots & k & k \\ k & h+k & \cdots & k & k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k & k & \cdots & h+k & k \\ k & k & \cdots & k & h+k \end{pmatrix} \in \mathbb{R}^{n\times n},$$

sea un producto euclidiano. Supuesto $k > 1$, $h = -1$, *¿puede ser* σ *un producto euclidiano? Nota: en el capítulo* Determinantes *se calculó que esta matriz tiene determinante* $h^{n-1}(h+kn)$.

Problema 368 *Diremos que* $\sigma : \mathbb{E} \times \mathbb{E} \to \mathbb{R}$ *bilineal simétrica es* ***definida negativa*** *si para todo* $x \neq 0$ *es* $\sigma(x, x) < 0$. *Probar un criterio análogo al de Sylvester para que* σ *sea definida negativa. Indicación:* σ *definida negativa equivale a que* $-\sigma$ *sea definida positiva.*

9.9. Conceptos afines euclidianos

Supongamos $\mathbb{E}$ un espacio euclidiano de dimensión n. Los principales conceptos que vamos a tratar tienen que ver con **distancia** y **perpendicularidad** (sinónimo: **ortogonalidad**). Con el producto euclidiano $\omega(x, y) = \langle x, y\rangle$ hay definida una **distancia** $d(x, y) = \|x - y\|$ (ver página 271). Verifica las condiciones **(a)** $d(x, y) \geq 0$ y solo 0 si $x = y$; **(b)** $d(x, y) = d(y, x)$; **(c)** $d(x, z) \leq d(x, y) + d(y, z)$, con $x, y, z \in \mathbb{E}$ arbitrarios. **(a)** y **(b)** son sencillas y **(c)** se probó tras el teorema 184. La distancia es el concepto fundamental al tratar **espacios métricos**[21] que aquí no trataremos pero estos espacios euclidianos, mayormente $(\mathbb{R}^n, \varepsilon)$, son ejemplos sencillos pero importantes de espacios métricos.

Se define la **distancia entre los subconjuntos** A y B no vacíos de $\mathbb{E}$ como

$$d(A, B) = \text{ínf}\{d(x, y) \mid x \in A,\ y \in B\}.$$

Una idea de uso continuo y demostración inmediata es que si tenemos la traslación $T(x) = x + v$ se tiene que $d(A, B) = d(T(A), T(B))$. Solo utilizaremos el concepto cuando A y B sean subespacios afines. Diremos que los subespacios $\mathbf{F} = s + \mathbb{F}$ y $\mathbf{G} = t + \mathbb{G}$ son **perpendiculares** u **ortogonales** si las direcciones $\mathbb{F}$ y $\mathbb{G}$ lo son; es decir si para $x \in \mathbb{F}$ e $y \in \mathbb{G}$ se cumple $\langle x, y\rangle = 0$.

Teorema 189 *Si* $\mathbf{F} = s + \mathbb{F}$ *y* $p \in \mathbb{E}$ *se tiene que* $d(p, \mathbf{F}) = \|P_{\mathbb{F}^\perp}(p - s)\|$. *Verbalmente:* $d(p, \mathbf{F})$ *es la longitud de la protección ortogonal de* $p - s$ *en el subespacio ortogonal a* $\mathbb{F}$, *la dirección de* $\mathbf{F}$. *Si* $P_{\mathbb{F}}(p - s) = y$ *entonces* $s + y$ *es el punto de* $\mathbf{F} = s + \mathbb{F}$ *a distancia mínima de* p.

[21] Es un par (X, d) con $X \neq \emptyset$ y $d : X \times X \to \mathbb{R}$ verificando **(a)**, **(b)** y **(c)**.

Demostración. Si $T(x) = x - s$ tenemos $d(p, \mathbf{F}) = d(T(p), T(\mathbf{F})) = d(p - s, \mathbb{F})$, y vimos en el teorema 183 que en $\mathbb{F}$ el punto a distancia mínima es $P_{\mathbb{F}}(p - s) = y$. Por consiguiente,

$$d(p, \mathbf{F})^2 = d(p - s, \mathbb{F})^2 = \|(p - s) - P_{\mathbb{F}}(p - s)\|^2 = \|P_{\mathbb{F}^\perp}(p - s)\|^2.$$

Como T preserva distancias, si $y \in \mathbb{F}$ está a distancia mínima de $p - s$, $s + y \in \mathbf{F}$ está a distancia mínima de $p - s + s = p$. ♣

Puede parecer que al ser $d(p, \mathbf{F}) = \|P_{\mathbb{F}^\perp}(p - s)\|$ y $s \in \mathbf{F}$ elegible de infinitas maneras en $\mathbf{F}$, hay el riesgo de una definición incorrecta, por dependencia de s. No hay tal peligro porque hay una definición alternativa (con el ínfimo) independiente de s.

Teorema 190 *Dados subespacios $\mathbf{F} = s + \mathbb{F}$ y $\mathbf{G} = t + \mathbb{G}$, se cumple que*

1. *Si son paralelos, digamos que $\mathbb{F} \subset \mathbb{G}$, al recorrer p el subespacio $\mathbf{F}$ la distancia $d(p, \mathbf{G})$ es constante. Este valor constante D es $d(\mathbf{F}, \mathbf{G})$.*
2. *Si $\mathbf{D}$ es una recta perpendicular a $\mathbf{F}$ y $\mathbf{G}$ (que pueden ser no paralelos) y los corta en s^* y t^*, entonces $d(\mathbf{F}, \mathbf{G}) = d(x^*, y^*)$.*

Demostración. Ampliemos una base ortonormal $(u_1, \ldots, u_m)$ de $\mathbb{G}$ hasta una base ortonormal $(u_1, \ldots, u_m, \ldots, u_n)$. Entonces,

$$P_{\mathbb{G}^\perp}(s - t) = \sum_{i=m+j}^{n} \langle s - t, u_j \rangle u_j, \quad d(s, \mathbf{G})^2 = \sum_{i=m+j}^{n} \langle s - t, u_j \rangle^2.$$

Si se sustituye s por $p = s + u$, $u \in \mathbb{F}$, otro punto de $\mathbf{F}$, tenemos, debido a que $\langle u, u_{m+j} \rangle = 0$ que

$$\langle p - t, u_j \rangle = \langle p - s + s - t, u_j \rangle = \langle s - t, u_j \rangle, \quad P_{\mathbb{G}^\perp}(p - t) = P_{\mathbb{G}^\perp}(s - t), \quad d(s, \mathbf{G})^2 = d(p, \mathbf{G})^2.$$

Probemos ahora que $D = d(p, \mathbf{G})$, independiente de $p \in \mathbf{F}$, es $d(\mathbf{F}, \mathbf{G})$. Si tenemos $p \in \mathbf{F}$ y $q \in \mathbf{G}$, se cumple $d(p, q) \geq d(p, \mathbf{G}) = D$, lo que lleva a $d(\mathbf{F}, \mathbf{G}) = \text{ínf}\{d(p, q) \mid p \in \mathbf{F},\ q \in \mathbf{G}\} \geq D$. Recíprocamente, $D = d(p, \mathbf{G})$ es expresable como $D = d(p, p')$ con $p' \in \mathbf{G}$. Por ser $d(\mathbf{F}, \mathbf{G})$ un ínfimo, $D = d(p, p') \geq d(\mathbf{F}, \mathbf{G})$.

Para **2** tomamos $x = s^* + u \in \mathbf{F}$ e $y = t^* + v \in \mathbf{G}$. Sea w un vector no nulo de la dirección de $\mathbf{D}$, y por tanto $s^* - t^* = \lambda w$. Entonces, como $\langle w, u \rangle = \langle w, v \rangle = 0$,

$$\|x - y\|^2 = \|(s^* - t^*) + (u - v)\|^2 = \|s^* - t^*\|^2 + \|u - v\|^2 \geq \|s^* - t^*\|^2 = d(x^*, y^*)^2.$$

Por la definición de $d(\mathbf{F}, \mathbf{G})$ como un ínfimo, obtenemos $d(\mathbf{F}, \mathbf{G}) \geq d(x^*, y^*)$. Por otra parte, x^* e y^* están en $\mathbf{F}$ y $\mathbf{G}$ y esto nos da, por ser $d(\mathbf{F}, \mathbf{G})$ un ínfimo, que $d(\mathbf{F}, \mathbf{G}) \leq d(x^*, y^*)$. ♣

Queremos dar fórmulas para calcular $d(p, \mathbf{F})$ con $\mathbf{F} = s + \mathbb{F}$. Todo depende del modo elegido para expresar $\mathbb{F}$. Lo que requiere menos discusión es el caso en donde $\mathbb{F}$ está en paramétricas en la forma $\mathbb{F} = \text{lg}(b_1, \ldots, b_m)$ pues $d(p, \mathbf{F}) = d(p - s, \mathbb{F})$ y para esto último tenemos la fórmula (9.11) con $x = p - s$ que nos da

$$d(p, \mathbf{F})^2 = d(p - s, \mathbb{F})^2 = \frac{\Gamma(p - s, b_1, \ldots, b_m)}{\Gamma(b_1, \ldots, b_m)}.$$

Si $(b_1, \ldots, b_m)$ fuese *además* ortogonal, los determinantes de Gram se simplificarían muchísimo. Puede optarse también por aplicar el procedimiento de Gram-Schmidt a $(b_1, \ldots, b_m)$ construyendo otra base $(v_1, \ldots, v_m)$ ortogonal y sería $\mathbb{F} = \text{lg}(v_1, \ldots, v_m)$. Es fácil con $(v_1, \ldots, v_m)$ calcular directamente $P_{\mathbb{F}}(p - s)$ y utilizar

$$\|p - s\|^2 = \|P_{\mathbb{F}}(p - s)\|^2 + \|P_{\mathbb{F}^\perp}(p - s)\|^2 = \|P_{\mathbb{F}}(p - s)\|^2 + d(p, \mathbf{F})^2.$$

Si $\mathbb{F}$ está en implícitas; o sea como solución de k ecuaciones *independientes* $f_1(x) = \ldots = f_k(x) = 0$ y las f_j formas lineales de $\mathbb{E}$, es conveniente escribir cada f_j como $f_j(x) = \langle a_j, x \rangle$; o sea f_j es el producto escalar por $a_j \in \mathbb{E}$. De este modo $\mathbb{F}$ viene dado por las ecuaciones $\langle a_1, x \rangle = \ldots = \langle a_k, x \rangle = 0$ y $\mathbb{F}^\perp = \text{lg}(a_1, \ldots, a_k)$ con base $(a_1, \ldots, a_k)$. Ahora, usando (9.11) pero en diferentes circunstancias,

$$d\left(p - s, \mathbb{F}^\perp\right)^2 = \frac{\Gamma(p - s, a_1, \ldots, a_k)}{\Gamma(a_1, \ldots, a_m)}.$$

Para tener $d(p,\mathbf{F})^2 = d(p-s,\mathbb{F})^2$ utilizamos $\|p-s\|^2 = d(p-s,\mathbb{F})^2 + d\left(p-s,\mathbb{F}^\perp\right)^2$.

Todo esto se simplifica mucho si $\mathbf{F}$ es una recta $\mathbf{D} = s+\lg(w)$ o un hiperplano de ecuación $\langle x-s,a\rangle = 0$ y más aún en el caso $(\mathbb{E},\omega) = \left(\mathbb{R}^3,\varepsilon\right)$ pues se evita usar complicados determinantes de Gram o se dispone del producto vectorial. Damos los detalles.

1. Sea $\mathbf{H}$ un hiperplano con ecuación $\langle x-s,a\rangle = 0$. Claramente $s \in \mathbf{H}$, $\mathbb{H} = \{x \mid \langle x,a\rangle = 0\}$ y $\mathbb{H}^\perp = \lg(a)$. Entonces,

$$d(p,\mathbf{H})^2 = \|P_{\mathbb{F}^\perp}(p-s)\|^2 = \left\|\frac{\langle p-s,a\rangle}{\|a\|^2}a\right\|^2 = \frac{\langle p-s,a\rangle^2}{\|a\|^2}, \quad d(p,\mathbf{H}) = \frac{|\langle p-s,a\rangle|}{\|a\|}.$$

   Lo que se hace es sustituir p en la ecuación de $\mathbf{H}$ de la forma $|\langle x-s,a\rangle| = 0$ y dividir por $\|a\|$.

2. Si $\mathbf{D} = s+\mathbb{D} = s+\lg(w)$ es una recta dada en paramétricas, tendremos

$$d(p,\mathbf{D})^2 = \|P_{\mathbb{D}^\perp}(p-s)\|^2 = \|p-s\|^2 - \|P_{\mathbb{D}}(p-s)\|^2 = \|p-s\|^2 - \left\|\frac{\langle p-s,w\rangle}{\|w\|^2}w\right\|^2$$
$$= \|p-s\|^2 - \frac{\langle p-s,w\rangle^2}{\|w\|^2} = \frac{\|p-s\|^2\|w\|^2 - \langle p-s,w\rangle^2}{\|w\|^2}.$$

3. Particularizamos a $(\mathbb{E},\omega) = \left(\mathbb{R}^3,\varepsilon\right)$. Al ser $\|p-s\|^2\|w\|^2 - \langle p-s,w\rangle^2 = \|(p-s)\times w\|^2$, con **2**,

$$d(p,\mathbf{D}) = \frac{\|(p-s)\times w\|}{\|w\|}.$$

   Si $\mathbf{D}$ viene en implícitas por dos ecuaciones $\langle x-s,u\rangle = \langle x-s,v\rangle = 0$, tenemos $\mathbb{D}^\perp = \lg(u,v)$ y $w = u\times v$ genera $\mathbb{D}$. Por tanto

$$d(p,\mathbf{D}) = \frac{\|(p-s)\times(u\times v)\|}{\|u\times v\|}.$$

   Si $\mathbf{H} = s+\lg(u,v)$ es un plano en paramétricas, puede usarse la fórmula general con $\mathbf{F} = \mathbf{H}$,

$$d(p,\mathbf{H})^2 = \frac{\Gamma(p-s,u,v)}{\Gamma(u,v)}.$$

   Mirando con atención también vemos en esta fórmula el producto vectorial $w = u\times v$. En efecto,

$$\Gamma(u,v) = \|u\|^2\|v\|^2 - \langle u,v\rangle^2 = \|u\times v\|^2, \quad \Gamma(p-s,u,v) = \det(p-s,u,v)^2 = \langle p-s,u\times v\rangle^2,$$

   como se tiene con fórmulas de la sección *Un primer contacto con el producto vectorial*. Por consiguiente,

$$d(p,\mathbf{H}) = \frac{\langle p-s,u\times v\rangle}{\|u\times v\|}.$$

Hemos querido ser exhaustivos en las fórmulas, pero, como cuestión personal, preferimos, aun en casos muy concretos, deducir la fórmula necesaria en abstracto, y sustituir al final. De todos modos hay una manera de recordar las fórmulas en $\mathbb{R}^3$ (y en dimensión 3 si se dispone de producto vectorial). Hay dos vectores *unitarios* $u_{\mathbb{D}}$ en la dirección $\mathbb{D}$ de $\mathbf{D}$ y $u_{\mathbb{H}}$ perpendicular a $\mathbb{H}$, la dirección del plano $\mathbf{H}$. Llamamos[22] a $u_{\mathbb{D}}$ y $u_{\mathbb{H}}$ los vectores unitarios principales asociados a $\mathbf{D}$ y $\mathbf{H}$. Para las fórmulas que vienen será irrelevante sustituir $u_{\mathbb{D}}$ por $-u_{\mathbb{D}}$ o $u_{\mathbb{H}}$ por $-u_{\mathbb{H}}$. El lector tiene fácil comprobar con **3** que

$$d(p,\mathbf{H}) = |\langle p-s,u_{\mathbb{H}}\rangle|, \quad d(p,\mathbf{D}) = \|(p-s)\times u_{\mathbb{D}}\|.$$

Verbalmente: la distancia de p a $\mathbf{H}$ es el *módulo del producto escalar* de $p-s$ por el vector unitario principal de $\mathbf{H}$, y la distancia de p a $\mathbf{D}$ es el *módulo del producto vectorial* de $p-s$ por el vector unitario principal de $\mathbf{D}$. Nos hemos apartado un poco de la terminología usual hablando de *módulo de un número* en vez de *valor absoluto* para que se vea el paralelismo. El tratamiento de las distancias punto-recta y punto-plano a un nivel elemental se basa en admitir la existencia de $u_{\mathbb{D}}$ y $u_{\mathbb{H}}$ y buscar su fórmula concreta con productos escalares, vectoriales o mixtos ("productos" $\langle a,b\times c\rangle = \det(a,b,c)$).

[22] No es terminología estándar, pero es útil.

Problema 369 *Para $p=(3,2,1)$ calcular la distancia de p a*

1. *El plano $\mathbf{H}_1$ con ecuación implícita $2(y-3)+2(z-3)=0$.*
2. *La recta $\mathbf{D}_1$ de ecuaciones paramétricas $x=3+\lambda$, $y=3+\lambda$, $z=3-\lambda$.*
3. *El plano $\mathbf{H}_2$ con ecuaciones paramétricas $x=3+\lambda+\mu$, $y=3+\mu$, $z=3+\lambda-\mu$*
4. *La recta $\mathbf{D}_2$ con ecuaciones implícitas $(x-3)+2(y-3)-4(z-3)=0$, $2(y-3)+2(z-3)=0$.* ♦

Solución. Se han elegido los $\mathbf{D}$ y $\mathbf{H}$ para que $s=(3,3,3)^\top$ esté en todos ellos. Como aparecerá bastante, digamos primero que $p-s=(0,-1,-2)$. La ecuación implícita de $\mathbf{H}_1$ es

$$\left\langle \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} x-3 \\ y-3 \\ z-3 \end{pmatrix} \right\rangle = 0.$$

El vector unitario principal a $\mathbf{H}_1$ es $(0,2,2)^\top$ *normalizado*; o sea $u_{\mathbb{H}_1}=\frac{1}{\sqrt{2}}(0,1,1)^\top$. Entonces,

$$d(p,\mathbf{H}_1)=\left\langle \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\rangle = \frac{3}{2}\sqrt{2}.$$

La ecuación paramétrica vectorial de $\mathbf{D}_1$ es

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

y su vector unitario principal es $(1,1,-1)^\top$ *normalizado*, que es $u_{\mathbb{D}_1}=\frac{1}{\sqrt{3}}(1,1,-1)^\top$. Entonces,

$$d(p,\mathbf{D}_1)=\left\| \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} \times \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\| = \frac{1}{\sqrt{3}}\left\| \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} \right\| = \frac{\sqrt{14}}{\sqrt{3}}.$$

La ecuación paramétrica vectorial de $\mathbf{H}_2$ es

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

El vector

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

es normal a $\mathbf{H}_2$ y normalizado a $\frac{1}{\sqrt{6}}(-1,2,1)^\top$ da $u_{\mathbb{H}_2}$, vector unitario principal de $\mathbf{H}_2$. Entonces,

$$d(p,\mathbf{H}_2)=\left| \left\langle \frac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} \right\rangle \right| = \frac{4}{\sqrt{6}}.$$

La recta $\mathbf{D}_2$ tiene ecuaciones implícitas

$$\left\langle \begin{pmatrix} x-3 \\ y-3 \\ z-3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -4 \end{pmatrix} \right\rangle = 0, \quad \left\langle \begin{pmatrix} x-3 \\ y-3 \\ z-3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \right\rangle = 0.$$

El vector

$$\begin{pmatrix} 1 \\ 2 \\ -4 \end{pmatrix} \times \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ -2 \\ 2 \end{pmatrix}$$

es de dirección de la recta y, normalizado a $\frac{1}{\sqrt{38}}(6,-1,1)$ es $u_{\mathbb{D}_2}$, vector unitario principal de $\mathbf{D}_2$. Entonces,

$$d(p,\mathbf{D}_2)=\left\|\begin{pmatrix}0\\-1\\-2\end{pmatrix}\times\frac{1}{\sqrt{38}}\begin{pmatrix}6\\-1\\1\end{pmatrix}\right\|=\frac{3\sqrt{21}}{\sqrt{38}}.\ \blacklozenge$$

Problema 370 *Póngase el lector otro problema parecido si quiere practicar.*

Nos queda lo que en apariencia puede ser la fórmula más difícil, que es la de la distancia entre dos rectas que se cruzan $\mathbf{F}=s+\lg(u)$ y $\mathbf{G}=t+\lg(v)$. Como usaremos el producto vectorial supondremos que estamos en $\left(\mathbb{R}^3,\varepsilon\right)$. Al cruzarse, (u,v) es independiente, y por ello $u\times v\neq 0$. Por **2** en el teorema 190 sabemos que si podemos encontrar puntos $x^*=s+\lambda u$ e $y^*=t+\mu v$ en $\mathbf{F}$ y $\mathbf{G}$ tales que la recta $\mathbf{D}$ que los une sea ortogonal a $\mathbf{F}$ y $\mathbf{G}$ se tendría que $d(x^*,y^*)=d(\mathbf{F},\mathbf{G})$. Como $\mathbf{D}$ tiene dirección $\lg(u\times v)$, es suficiente que probemos que hay solución (λ,μ,θ) de la ecuación

$$(s+\lambda u)-(t+\mu v)=\theta(u\times v)\ \text{ equivalente a }\ s-t=-\lambda u+\mu v+\theta(u\times v)$$

Esta solución existe debido a que $(u,v,u\times v)$ es base de $\mathbb{R}^3$. Podríamos incluso determinar x^* e y^* que son los puntos a mínima distancia. Sin embargo, si solo interesa $d(\mathbf{F},\mathbf{G})$ basta conocer $\|\theta(u\times v)\|$. Esto es fácil porque, como $u\times v$ es ortogonal a u y v,

$$\theta=\frac{\langle s-t,u\times v\rangle}{\|(u\times v)\|^2},\quad d(\mathbf{F},\mathbf{G})=\|\theta(u\times v)\|=\frac{|\langle s-t,u\times v\rangle|}{\|(u\times v)\|}.$$

Teorema 191 *Dadas rectas que se cruzan $\mathbf{F}=s+\lg(u)$ y $\mathbf{G}=t+\lg(v)$, la distancia entre ellas es*

$$d(\mathbf{F},\mathbf{G})^2=\frac{\langle s-t,u\times v\rangle^2}{\|(u\times v)\|^2}=\frac{\Gamma(s-t,u,v)}{\Gamma(u,v)},$$

y los puntos $x^*=s+\lambda u$ e $y^*=t+\mu v$ a distancia mínima se determinan calculando λ y μ en la ecuación $s-t=-\lambda u+\mu v+\theta(u\times v)$, siendo las soluciones (0jo al signo menos)

$$-\lambda=\frac{\begin{vmatrix}\langle s-t,u\rangle & \langle v,u\rangle\\ \langle s-t,v\rangle & \|v\|^2\end{vmatrix}}{\Gamma(u,v)},\qquad \mu=\frac{\begin{vmatrix}\|u\|^2 & \langle s-t,u\rangle\\ \langle u,v\rangle & \langle s-t,v\rangle\end{vmatrix}}{\Gamma(u,v)}.$$

La primera parte se ha probado y la segunda es un problema para el lector.

Problema 371 *Lo dicho, probar las fórmulas para λ y μ que dan x^* e y^* en el teorema precedente.*

El lector se puede poner un problema de puro cálculo a su gusto. Nosotros preferimos plantearle un problema teórico muy sencillo.

Problema 372 *En las fórmulas **1-4** de distancias de p a rectas o planos y la de la distancia entre rectas que se cruzan involucran puntos s y t de esas rectas o planos, que aparecen en las fórmulas numéricas. ¿Por qué sabemos que no dependen de s y t?*

Capítulo 10

Funciones en espacios euclidianos

Hasta ahora hemos estudiado tres tipos de objetos principales en un espacio vectorial $\mathbb{E}$, que son sus subespacios $\mathbb{F}$, sus endomorfismos L, y sus funciones bilineales simétricas σ. ¿Qué se puede añadir si en $\mathbb{E}$ tenemos un producto euclidiano $\omega(x,y) = \langle x,y\rangle$? Por lo que se refiere a subespacios tenemos el concepto de ortogonalidad o perpendicularidad, podemos a veces descomponer $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^{\perp}$, también estudiar el concepto de distancia y ángulo y buscar el punto más cercano en $\mathbb{F}$ a otro $x \in \mathbb{E}$, e incluso a expresar en implícitas $\mathbb{F}$ como el conjunto de vectores ortogonales a vectores $a_1, \ldots, a_k$. Y las bases ortogonales y ortonormales son fundamentales en todo esto. Nos preguntamos si los endomorfismos L y las funciones bilineales σ tendrán una expresión, si no más sencilla (muchos ceros y unos), al menos más informativa respecto a "lo que hacen" L o σ. La respuesta es sí porque hay teoremas de descripción, y con ella una clasificación, de ciertos endomorfismos y funciones bilineales, siendo la esencia de ello el cómo se expresan en bases *ortonormales* adecuadas. Es más fácil describir lo que pasa con las funciones σ porque lo que hay son bases *ortonormales* que diagonalizan σ, cosa que además permitirá una mejor descripción de las cónicas y cuádricas. No es cierto en general que si L es diagonalizable haya una base *ortonormal* en la que L tenga matriz diagonal. Sin embargo hay un tipo de endomorfismos que, sin ser diagonalizables, tienen una buena descripción con bases ortonormales. Son las **isometrías**, que se definen como las funciones lineales $L : \mathbb{E} \to \mathbb{E}$ tales que $\omega(L(x), L(y)) = \omega(x,y)$ para todo $x, y \in \mathbb{E}$. Puede verse que cumplen $d(L(x), L(y)) = d(x,y)$, luego preservan la distancia asociada a ω y esto las anuncia como muy interesantes. Hay teoremas estructurales de gran importancia que permiten clasificar las isometrías.

Problema 373 *Dar un ejemplo de $L : \mathbb{R}^2 \to \mathbb{R}^2$ que sea diagonalizable pero que sea imposible que esto suceda para una base ortogonal para ε, el producto estándar.*

La dimensión infinita se considera optativa, tanto en teoría como en problemas.

10.1. Adjunto de un endomorfismo

En toda la sección tenemos un espacio euclidiano $\mathbb{E}$ con producto escalar $\omega(x,y) = \langle x,y\rangle$ y un endomorfismo $L : \mathbb{E} \to \mathbb{E}$. En muchos casos será $\mathbb{E} = \mathbb{R}^n$, $a \in \mathbb{R}^{n\times n}$ una matriz, y $L(x) = ax$, si bien ω no tiene por qué ser necesariamente ε, el producto euclidiano estándar.

Cada L determina otro endomorfismo $L^* : \mathbb{E} \to \mathbb{E}$ con importantes relaciones con L. A pesar de la $*$ elegida para la notación, L^* no es un endomorfismo de $\mathbb{E}^*$ (el dual de $\mathbb{E}$) sino de $\mathbb{E}$. Diremos que un endomorfismo M es el **adjunto** de L si $\langle L(x), y\rangle = \langle x, M(y)\rangle$ para todo $x, y \in \mathbb{E}$.

Teorema 192 *Si $\mathbb{E}$ es de dimensión finita, existe y es único el adjunto M de L. Las matrices a y b de L y M en una base* arbitraria *$\mathcal{U}$ de $\mathbb{E}$ se relacionan por $a^{\top}\Omega = \Omega b$, siendo Ω la matriz de ω en la base $\mathcal{U}$. En el caso particular en que $\mathcal{U}$ sea ortonormal se tiene $b = a^{\top}$.*

Demostración. Si x e y tienen matrices de coordenadas ξ y η y suponemos que M existe, se tiene

$$\langle L(x), y\rangle = (a\xi)^{\top}\Omega\eta = \xi^{\top}a^{\top}\Omega\eta, \quad \langle x, M(y)\rangle = \xi^{\top}\Omega b\eta.$$

Tomando ξ y η en la base estándar de $\mathbb{R}^n$ se obtiene $\left(a^\top \Omega\right)_j^i = (\Omega b)_j^i$; o sea, $a^\top \Omega = \Omega b$. Como Ω es invertible, $b = \Omega^{-1} a^\top \Omega$, luego a determina b y M unívocamente.

Probada la unicidad, definimos M como el endomorfismo con matriz $b = \Omega^{-1} a^\top \Omega$. Dado que $a^\top \Omega = \Omega b$, probamos que $\langle L(x), y\rangle = \langle x, M(y)\rangle$ porque $\langle L(x), y\rangle = \xi^\top a^\top \Omega \eta$ y $\langle x, M(y)\rangle = \xi^\top \Omega b \eta$ para ξ y η arbitrarios. ♣

Se denotará por L^* al adjunto de L, que está unívocamente determinado por L (se prueba aun con dimensión infinita más abajo) teniendo presente que, aunque no figure en la notación, L^* depende no solo de L sino del producto ω elegido. Está definido como el *único* endomorfismo L^* que verifica

$$\langle L(x), y\rangle = \langle x, L^*(y)\rangle \quad \text{para todo } x, y \in \mathbb{E}. \tag{10.1}$$

Esta posibilidad de "saltar en $\langle \bullet, \bullet\rangle$ poniendo o quitando $*$" es de uso continuo. En bases no ortonormales, las matrices a y b de L y L^* pueden "parecerse muy poco". Por ejemplo, si en $\mathbb{R}^2$ damos ω y L con matrices en la base estándar (¡que no va a ser ortonormal!)

$$\Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \qquad a = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{10.2}$$

se tiene que L^* viene dado en la base estándar por

$$b = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}.$$

Hay veces en que un cálculo directo va mucho mejor como en el siguiente ejemplo donde $\mathbb{E} = \mathbb{R}^{n\times n}$. Vimos que $\langle a, b\rangle = \operatorname{tr}\left(a^\top b\right)$ es un producto euclidiano (y la base estándar es ortonormal, aunque no lo vamos a usar). Se fija una matriz ℓ y se define con ella $L : \mathbb{E} \to \mathbb{E}$ por $L(a) = \ell \cdot a$ (no olvidemos que a, b, ℓ son *matrices* $n \times n$). ¿Cómo es L^*? Trabajar con matrices en $\mathbb{E} = \mathbb{R}^{n\times n}$ es muy complicado pues serían $n^2 \times n^2$. Sin embargo es fácil comprobar (y bastante menos de imaginar) que la respuesta es $L^*(b) = \ell^\top \cdot b$ (se multiplica por $\ell^\top$ para L^* en vez de por ℓ). En efecto,

$$\langle L(a), b\rangle = \operatorname{tr}\left((\ell a)^\top b\right) = \operatorname{tr}\left(a^\top \ell^\top b\right), \quad \langle a, L^*(b)\rangle = \operatorname{tr}\left(a^\top \ell^\top b\right).$$

La definición de adjunto y su unicidad valen en dimensión infinita. No obstante, en dimensión infinita no está garantizada la existencia pero sí la unicidad. Esta se sigue de que si hubiera M_1 y M_2 cumpliendo la definición de adjunto se tendría $\langle x, (M_1 - M_2)(y)\rangle$ y un vector $(M_1 - M_2)(y)$ ortogonal a todo $\mathbb{E}$ es necesariamente cero. De acuerdo con esto, la función $L \to L^*$ de $\mathcal{L}(\mathbb{E}, \mathbb{E})$ en $\mathcal{L}(\mathbb{E}, \mathbb{E})$ es lineal e inyectiva y, en dimensión finita, un isomorfismo.

Son de uso muy frecuente las propiedades del siguiente teorema.

Teorema 193 *El adjunto de un endomorfismo cumple las siguientes propiedades*

1. *Si $L, M : \mathbb{E} \to \mathbb{E}$ son endomorfismos y $\lambda \in \mathbb{R}$, $(L+M)^* = L^* + M^*$ y $(\lambda L)^* = \lambda L^*$*
2. *$(L \circ M)^* = M^* \circ L^*$; o sea el adjunto de la composición es la composición de los adjuntos* en orden inverso. *El doble adjunto es el endomorfismo original; o sea, $(L^*)^* = L$*
3. *Si $P(X) = \alpha_0 + \alpha_1 X + \ldots + \alpha_n X^n$ es un polinomio y sustituimos X por L obteniendo un nuevo endomorfismo $P(L)$, se tiene que $P(L)^* = P(L^*)$.*
4. *L y L^* tienen el mismo polinomio característico; por tanto, tienen los mismos valores propios y con la misma multiplicidad (aunque no los mismos vectores propios[1]).*
5. *Para $\mathbb{F}$ subespacio de $\mathbb{E}$, si $\mathbb{F}$ estable por L, entonces $\mathbb{F}^\perp$ es estable por L^*; y si $\mathbb{F}$ es estable por L^*, entonces $\mathbb{F}^\perp$ es estable por L. Con símbolos, $L(\mathbb{F}) \subset \mathbb{F}$ implica $L^*\left(\mathbb{F}^\perp\right) \subset \mathbb{F}^\perp$ y $L^*(\mathbb{F}) \subset \mathbb{F}$ implica $L\left(\mathbb{F}^\perp\right) \subset \mathbb{F}^\perp$.*

[1] En el ejemplo de (10.2), L y L^* tienen en la base estándar las matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ y $\begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix}$, ambas con el mismo polinomio característico $(X-1)^2$ pero los autovectores son respectivamente los múltiplos de e_1 y $e_1 - e_2$.

Demostración. Los dos primeros apartados se pueden hacer observando que dos endomorfismos coinciden si coinciden sus matrices en una base $\mathcal{U}$. Si se elige $\mathcal{U}$ ortonormal, las propiedades a probar son inmediatas por las propiedades de la trasposición. Por ejemplo, sean a y b las matrices de L y M en $\mathcal{U}$. La matriz de $(L\circ M)^*$ es $(ab)^\top$ ya que ab es la matriz de $L\circ M$. Por otra parte la matriz de M^* y L^* son $b^\top$ y $a^\top$ y la de $M^*\circ L^*$ es $b^\top a^\top$. Dado que $(ab)^\top = b^\top a^\top$ se tiene que $(L\circ M)^* = M^*\circ L^*$. Es más aconsejable el hacer una demostración válida en dimensión infinita,[2] donde no hay bases ni matrices. Como el adjunto es único basta ver que para todo para (x,y) de $\mathbb{E}$ se verifica que

$$\langle L\circ M(x), y\rangle = \langle x, M^*\circ L^*(y)\rangle .$$

Pero sabiendo que L y M o L^* y M^* pueden "saltar de lado" poniendo o quitando estrellas, resulta

$$\langle L\circ M(x), y\rangle = \langle M(x), L^*(y)\rangle = \langle x, M^*(L^*(y))\rangle$$

que es lo que queríamos. Es inmediato **3** con **1** y **2**.

Probamos **4**. Sea $a = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$ en una base ortonormal. Los polinomios característicos de L y L^* se pueden calcular con la matriz que se desee y son $\det(a - XI)$ y $\det\left(a^\top - XI\right)$, el mismo polinomio.

Supongamos $L(\mathbb{F})\subset F$ y sea $x^\perp\in\mathbb{F}^\perp$. Para ver que $L^*\left(x^\perp\right)\in\mathbb{F}^\perp$ tomamos $y\in\mathbb{F}$ y comprobamos si $\left\langle L^*\left(x^\perp\right), y\right\rangle = 0$, cierto ya que $\left\langle L^*\left(x^\perp\right), y\right\rangle = \left\langle x^\perp, L(y)\right\rangle = 0$ porque $x\in\mathbb{F}^\perp$ y $L(y)\in\mathbb{F}$. Si sustituimos L por L^*, con lo recién probado, $L^*(\mathbb{F})\subset\mathbb{F}$ implica $(L^*)^*\left(\mathbb{F}^\perp\right)\subset\mathbb{F}^\perp$ y al ser $(L^*)^* = L$ resulta lo que queríamos: que $L^*(\mathbb{F})\subset\mathbb{F}$ implica $L\left(\mathbb{F}^\perp\right)\subset\mathbb{F}^\perp$. ♣

Problema 374 *Probar que, en dimensión finita, si L es invertible lo es también L^* y $(L^*)^{-1} = \left(L^{-1}\right)^*$*

Problema 375 *Probar que si $L = L^*$ y $M = M^*$ se tiene que $L\circ M = (L\circ M)^*$ si y solo si $L\circ M = M\circ L$.*

Problema 376 *Sea $\mathbb{E} = \mathbb{F}\oplus\mathbb{G}$ de modo que $\langle y, z\rangle = 0$ si $y\in\mathbb{F}$ y $z\in\mathbb{G}$. Probar que $\mathbb{G} = \mathbb{F}^\perp$. Usar esto para probar que si $P:\mathbb{E}\to\mathbb{E}$ es la proyección sobre $\mathbb{F}$ asociada $\mathbb{E} = \mathbb{F}\oplus\mathbb{G}$ la propiedad $\mathbb{G} = \mathbb{F}^\perp$ equivale a $P = P^*$. Así pues, $P = P^*$ si y solo si P es proyección asociada a una descomposición ortogonal.*

Problema 377 *Probar en dimensión finita que* $\mathrm{rg}(L) = \mathrm{rg}(L^*)$ *y que* $\mathrm{im}(L^*) = (\ker L)^\perp$.

10.2. Endomorfismos normales

Es una propiedad clave en los teoremas principales de este capítulo que $L^*\circ L = L\circ L^*$. Si un endomorfismo cumple esto se llama **normal**. Si a y b son las matrices de L y L^* en una base arbitraria $\mathcal{U}$ se ha de tener $ab = ba$, pero sabemos del teorema 192 que $\Omega^{-1}a^\top\Omega = b$ de modo que $ab = ba$ equivale a $a\Omega^{-1}a^\top\Omega = \Omega^{-1}a^\top\Omega a$, cosa difícil de manejar, si bien, cuando $\mathcal{U}$ es ortonormal, $\Omega = I$, y entonces la condición matricial es $aa^\top = a^\top a$. Una matriz $a\in\mathbb{R}^{n\times n}$ se llama **matriz normal** si $a^\top a = aa^\top$; o sea, conmuta con su traspuesta. Con lo anterior, esto equivale a que si con a definimos $L:\mathbb{R}^n\to\mathbb{R}^n$, $L(x) = ax$ y $\omega = \varepsilon$, entonces L es normal con la definición general si y solo si a es normal.

Aunque daremos teoremas generales para endomorfismos normales, nos interesa centrarnos en tres casos concretos.

1. El primero es $L = L^*$ y L se llama **autoadjunto** o **simétrico**.[3] La razón de "simétrico" es que si se define
$$\sigma:\mathbb{E}\times\mathbb{E}\to\mathbb{R},\quad \sigma(x,y) = \langle L(x), y\rangle$$
resulta
$$\sigma(x,y) = \langle L(x), y\rangle = \langle x, L^*(y)\rangle = \langle x, L(y)\rangle = \langle L(y), x\rangle = \sigma(y,x);$$
es decir, que la forma bilineal σ es *simétrica. Esto será muy importante porque teoremas sobre L pasarán a ser teoremas sobre σ*. Es *falso* que L y σ tengan la misma matriz para $\mathcal{U}$ o que la matriz de L en cualquier $\mathcal{U}$ sea simétrica, aunque esto es *cierto* si $\mathcal{U}$ es ortonormal, como veremos.

[2] Aunque aquí nos centramos en espacios de dimensión finita, muchas ideas de este capítulo se generaron para resolver problemas de Análisis en espacios de funciones.

[3] El nombre de "simétrico" si $L = L^*$ es peligroso y "autoadjunto" es preferible.

2. El segundo caso es $L^* \circ L = L \circ L^* = \mathrm{id}_{\mathbb{E}}$, lo que exige en particular que L sea biyectiva; o sea, un isomorfismo. Se dice entonces que L es una **isometría** o que L es **ortogonal** (preferimos el primer nombre). Se tiene que

$$\langle x, y\rangle = \langle x, L^*(L(y))\rangle = \langle L(x), L(y)\rangle \text{ para todo } x, y \in \mathbb{E}, \tag{10.3}$$

luego L preserva los productos de vectores. Si $x = y$ se sigue que $\|x\| = \|L(x)\|$ y como la distancia es $d(x,y) = \|x - y\|$, vemos que L preserva distancias y es el origen del nombre "isometría". Advertimos que (10.3) es equivalente *en dimensión finita* a $L^* \circ L = L \circ L^* = \mathrm{id}_{\mathbb{E}}$ porque (10.3) siempre implica $\|x\| = \|L(x)\|$ así que $L(x) = 0$ nos da $x = 0$ y L, sea como sea la dimensión, es inyectiva. Si embargo hay ejemplos en dimensión infinita donde se cumple (10.3) pero L no es biyectiva, de modo que $L^* \circ L = L \circ L^* = \mathrm{id}_{\mathbb{E}}$ es imposible.

3. El tercer caso, que tiene interés pero trataremos poco, es $L^* = -L$. Puede verse como en **1** que $\alpha(x,y) = \langle L(x), y\rangle$ es antisimétrica y por ello se dice que L es **antisimétrica.** Hay comentarios parecidos a los hechos para σ.

Problema 378 *Sea $\mathbb{E}$ el espacio de las sucesiones $(x_1, x_2, \ldots, x_n, \ldots)$ con solo un número finito de x_j no nulos. Se define $\langle x, y\rangle = \sum_{i=1}^{\infty} x_j y_j$ que en realidad tiene solo un número finito sumandos $\neq 0$. Tomemos $L(x) = (0, x_1, x_2, \ldots, x_n, \ldots)$; o sea, corremos un lugar a la derecha los términos de x. Probar que se cumple (10.3) pero L no es suprayectiva por tanto $L \circ L^* = \mathrm{id}_{\mathbb{E}}$ no puede darse.*

Si no se dice lo contrario se supone $\mathbb{E}$ de dimensión finita. Los teoremas más importantes sobre endomorfismos normales dicen que estos endomorfismos son diagonalizables o que están muy cerca de serlo, porque si no lo son hay sumandos invariantes de dimensión 2. Lo importante no es solo la simplicidad de la descomposición sino que *se puede conseguir con bases ortonormales.* Esto será fundamental para entender ciertos entes geométricos. En el caso de los endomorfismos autoadjuntos o simétricos el objeto serán las **cónicas** y **cuádricas**, ya vistas cuando no teníamos las ideas de longitud, perpendicularidad, etc., así que podemos esperar un estudio más completo. La clave estará en el estudio del caso $L = L^*$. Por otra parte, las isometrías $L^* \circ L = L \circ L^* = \mathrm{id}_{\mathbb{E}}$ serán las transformaciones que preservan la distancia y se intuirán como "movimientos rígidos"; por ejemplo, rotaciones y simetrías respecto a rectas y planos. Resumiendo mucho: los endomorfismos normales pueden verse como los que cumplen una condición algebraica que incluye los autoadjuntos y las isometrías, llevando su conocimiento a poder estudiar cónicas, cuádricas, rotaciones, simetrías, etc. porque se consigue referirlas de modo bastante simple a bases ortonormales.

Problema 379 *Sea $L : \mathbb{E} \in \mathbb{E}$ autoadjunta y $\sigma(x,y) = \langle L(x), y\rangle$. Tenemos para una base $\mathcal{U}$ las matrices a de L, s de σ y Ω de ω. Probar que*

1. *Sea como sea $\mathcal{U}$, la matriz s de σ es simétrica.*
2. *Se relacionan por $s = a^\top \Omega$ y si $\mathcal{U}$ es ortonormal, $s = a^\top = a$.*
3. *Si en $\mathbb{E} = \mathbb{R}^2$ tomamos las matrices*

$$\Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad s = \begin{pmatrix} 1 & 0 \\ 0 & h \end{pmatrix},$$

el producto euclidiano $\omega(x,y) = x^\top \Omega y$ y $\sigma : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $\sigma(x,y) = x^\top s y$. ¿Cuál es la matriz a de $L : \mathbb{R}^2 \to \mathbb{R}^2$ definido por $\sigma(x,y) = \langle L(x), y\rangle$ en la base estándar $\mathcal{E}$? ¿Tiene que ser a simétrica?

Las matrices ortogonales tienen una estrecha relación con las isometrías (por eso estas se llaman con frecuencia transformaciones ortogonales). No solo aparecen como las matrices de cambio de base cuando ambas son ortogonales, sino que son la matriz que adopta $L : \mathbb{E} \to \mathbb{E}$ cuando $\mathcal{U}$ es ortonormal.

Teorema 194 *Sea $L : \mathbb{E}_1 \to \mathbb{E}_2$ un isomorfismo entre los espacios euclidianos $(\mathbb{E}_1, \omega_1)$ y $(\mathbb{E}_2, \omega_2)$ de dimensión n. Son equivalentes las condiciones*

1. *La matriz a de L respecto de bases ortonormales $\mathcal{U}_{(1)}$ y $\mathcal{U}_{(2)}$ de $\mathbb{E}_1$ y $\mathbb{E}_2$ es una matriz ortogonal.*
2. *Para todo x, y en $\mathbb{E}_1$, $\omega_2(L(x), L(y)) = \omega_1(x,y)$.*

3. *Toda base ortonormal* $\mathcal{U}_{(1)}$ *de* $\mathbb{E}_{(1)}$ *cumple que* $L\left(\mathcal{U}_{(1)}\right) = \left(L\left(u_{(1)1}\right), \ldots, L\left(u_{(1)n}\right)\right)$ *es una base ortonormal de* $\mathbb{E}_{(2)}$.

4. *Hay al menos una base ortonormal* $\mathcal{U}_{(1)}$ *de* $\mathbb{E}_1$ *tal que* $\left(L\left(u_{(1)1}\right), \ldots, L\left(u_{(1)n}\right)\right)$ *es una base ortonormal de* $\mathbb{E}_{(2)}$.

Demostración. Damos demostración circular **1** $\Longrightarrow$ **2** $\Longrightarrow$ **3** $\Longrightarrow$ **4** $\Longrightarrow$ **1**. Si se tiene **1**, las matrices de los ω_i en las bases $\mathcal{U}_{(i)}$ son I. Entonces si $x, y \in \mathbb{E}_1$ tienen matrices de coordenadas ξ y η,

$$\omega_2\left(L\left(x\right), L\left(y\right)\right) = \left(a\xi\right)^\top I\left(a\eta\right) = \xi^\top a^\top Ia\eta \overset{*}{=} \xi^\top \eta = \omega_1\left(x, y\right),$$

usándose en $\overset{*}{=}$ que $a^\top a = I$. Visto que **1** $\Longrightarrow$ **2**, es inmediato que **2** $\Longrightarrow$ **3** $\Longrightarrow$ **4**. Supongamos **4** cierto. Sabemos que la columna j de a, que es a_j, es la matriz de las coordenadas de $L\left(u_{(1)j}\right)$ en la base $\mathcal{U}_{(2)}$., Por **4** y el que ω_2 tenga en $\mathcal{U}_{(2)}$ a I como $\mathrm{mat}_{\mathcal{U}_2\mathcal{U}_2}\left(\omega_2\right)$ se debe cumplir

$$\delta_{ij} = \omega_1\left(u_{(1)i}, u_{(1)j}\right) = \omega_2\left(L\left(u_{(1)i}\right), L\left(u_{(1)j}\right)\right) = \left(a_i\right)^\top Ia_j = \left(a^\top a\right)_{ij},$$

que es la condición de que a sea ortogonal. ♣

Problema 380 *Sea* $S\left(x\right) = x - 2\left\langle x, u\right\rangle u$ *con* $\left\|u\right\|^2 = 1$, *que es la simetría respecto a* $\mathbb{H}$, *el hiperplano perpendicular a* u. *Probar que es una isometría tanto por el procedimiento directo como calculando la matriz de* S *respecto a una base ortonormal* $\mathcal{U}$ *tal que* u *sea su primer vector.*

Como la composición de isometrías es una isometría, podríamos tomar vectores $u_1, \ldots, , u_k$ no nulos, considerar las correspondientes simetrías respecto a hiperplanos $\mathbb{H}_j : \left\langle x, u_j\right\rangle = 0$, que serían

$$S_j\left(x\right) = x - \frac{2\left\langle x, u_j\right\rangle}{\left\|u_j\right\|^2}u_j, \quad j = 1, \ldots, , k$$

y definir $L = S_k \circ \cdots \circ S_1$, obteniendo una isometría. ¿Son todas las isometrías factorizables de este modo? El **teorema de Cartan-Dieudonné**, que dice sí es cierto (en dimensión finita) y se probará más adelante. Informalmente: las simetrías respecto a hiperplanos son los ladrillos que, adecuadamente elegidos y montados, nos dan cualquier isometría.

El arquetipo de las matrices ortogonales son las matrices 2×2

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \quad \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}, \quad \alpha \in \mathbb{R}, \tag{10.4}$$

o bien en la forma

$$a = \begin{pmatrix} p & -Dq \\ q & Dp \end{pmatrix}, \quad \text{siendo } D = \det\left(a\right) = \pm 1,\ p^2 + q^2 = 1.$$

como se vio en el problema 333. [4] Si tenemos una base ortonormal $\mathcal{U}$ y definimos $\mathcal{V}$ por $v_j = \sum c_j^i u_i$, siendo c una matriz ortogonal, sabemos que $\mathcal{V}$ es también ortonormal. Recíprocamente, si el dato es ahora que $\mathcal{V}$ es ortonormal, obtenemos que c ha de ser ortogonal. Por ejemplo, si partimos de una base ortonormal $\left(u_1, u_2\right)$, la nueva base $\mathcal{V}$

$$\left(v_1, v_2\right) = \left(u_1, u_2\right)\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \text{ equivalente a } \mathcal{V} = \begin{cases} v_1 = \left(\cos\alpha\right)u_1 + \left(\sin\alpha\right)u_2 \\ v_2 = -\left(\sin\alpha\right)u_1 + \left(\cos\alpha\right)u_2 \end{cases}$$

es ortonormal.

Un comentario importante antes de seguir. Nos interesan sobre todo los endomorfismos autoadjuntos y las isometrías. El dar propiedades de endomorfismos normales en general sirve para no tener que repetir demostraciones bastante parecidas para $L = L^*$ o $L^{-1} = L$. Sin embargo esto es más que un puro ahorro expositivo. Hay una generalización del concepto de producto euclidiano a *espacios complejos*. Allí los conceptos (ortogonalidad, bases, proyecciones, normalidad, etc.) tienen un tratamiento paralelo al aquí expuesto. En un espacio complejo todo polinomio característico es producto de factores de primer grado, en principio insuficiente para que L sea diagonalizable. Sin embargo, se puede probar que *si* L *es normal, entonces es diagonalizable con base ortonormal.* Aquí, al ser $\Bbbk = \mathbb{R}$ solo conseguiremos diagonalizar si $L = L^*$, condición mucho más restrictiva que $L \circ L^* = L^* \circ L$.

[4] Un teorema de Análisis que dice que todo punto (p, q) del círculo con ecuación $X^2 + Y^2 = 1$ se puede escribir como $p = \cos\alpha$, $q = \sin\alpha$, con $\alpha \in \mathbb{R}$ unívocamente determinado salvo múltiplo entero de 2π. Esto último es como decir que si para otro α' se tuviera también $p = \cos\alpha'$, $q = \sin\alpha'$ se ha de cumplir que $\alpha - \alpha' = 2\pi n$ con $n \in \mathbb{Z}$. Basta entonces hacer esta sustitución y obtener que (10.4) *describe todas las matrices ortogonales para* $n = 2$.

Teorema 195 *Si L es normal verifica*

1. *Para todo $x \in \mathbb{E}$ es $\|L(x)\| = \|L^*(x)\|$.*
2. *Al sustituir X por L en el polinomio $P(X)$, el endomorfismo $P(L)$ es normal. En particular, $L-\lambda$ con $\lambda \in \mathbb{R}$ es normal.*
3. *L y L^* tienen los mismos vectores propios y con el mismo valor propio.*[5]
4. *Los vectores propios con distinto valor propio son ortogonales.*

Para el caso particular de las transformaciones ortogonales (o isometrías) se tiene además

1. *El determinante de L es ± 1 si L es una isometría.*
2. *La composición de isometrías y el inverso de una isometría es una isometría. La identidad es una isometría. Las isometrías de $\mathbb{E}$ forman con la composición de funciones como producto un grupo (no abeliano) llamado* ***grupo ortogonal*** *o* ***grupo de las isometrías (lineales).***

Demostración. Se tiene para **1** que, como por definición, $L^* \circ L = L \circ L^*$,

$$\|L(x)\|^2 = \langle L(x), L(x)\rangle = \langle x, L^* \circ L(x)\rangle = \langle x, L \circ L^*(x)\rangle = \langle L^*(x), L^*(x)\rangle = \|L^*(x)\|^2.$$

Equivale **2** a que $P(L)$ y $P(L)^*$ conmuten. Por **3** en el teorema 193, $P(L)^* = P(L^*)$ y

$$\begin{cases} P(L) \circ P(L)^* = \left(\sum_{i=0}^k \alpha_i L^i\right) \circ \left(\sum_{j=0}^k \alpha_j (L^*)^j\right) = \sum_{i.j=0}^k \alpha_i \alpha_j L^i \circ (L^*)^j \\ P(L)^* \circ P(L) = \left(\sum_{j=0}^k \alpha_j (L^*)^j\right) \circ \left(\sum_{i=0}^k \alpha_i L^i\right) = \sum_{i.j=0}^k \alpha_j \alpha_i (L^*)^j \circ L^i \end{cases}$$

La condición $L \circ L^* = L^* \circ L$; implica $L^i \circ (L^*)^j = (L^*)^j \circ L^i$ (inducción) y se cumple **2**.

Como $\mathrm{id}_{\mathbb{E}}^* = \mathrm{id}_{\mathbb{E}}$, se sigue de **2** que tanto $L - \lambda$ como $L^* - \lambda$ son normales. Entonces, por **1**, $\|L(x) - \lambda x\| = \|L^*(x) - \lambda x\|$. Resultan entonces las equivalencias **(a)** x es vector propio de L con valor propio λ, **(b)** $L(x) - \lambda x = 0$, **(c)** $\|L(x) - \lambda x\| = 0$, **(d)** $\|L^*(x) - \lambda x\| = 0$, **(e)** x es vector propio de L^* con valor propio λ. Queda así probado **3**.

Probemos **4.** Sean $L(x) = \lambda x$ y $L(y) = \mu y$. Por **3** también $L^*(y) = \mu y$. Entonces

$$0 = \langle L(x), y\rangle - \langle x, L^*(y)\rangle = (\lambda - \mu)\langle x, y\rangle$$

y $\lambda - \mu \neq 0$ implica $\langle x, y\rangle = 0$.

Si a es la matriz de L en una base ortonormal, como el determinante no depende de la base elegida, $\det(L) = \det(a)$. Pero $aa^\top = I$, luego $1 = \det(a)\det\left(a^\top\right) = \det(a)^2$ y $\det(a) = \pm 1$. El comprobar que las isometrías forman grupo es trivial y queda para el lector. ♣

Este teorema implica inmediatamente una serie de propiedades para las matrices normales (la que cumplen $aa^\top = a^\top a$), identificando a con L y $a^\top$ con L^*, endomorfismos de $\mathbb{R}^n$ y ε como producto euclidiano. Podemos pues decir que a y $a^\top$ tienen los mismos valores propios y vectores propios y que vectores propios de diferentes valores propios son ortogonales para ε. Esto tendrá gran trascendencia y llevará a los llamados **teoremas espectrales**.

Problema 381 *En un espacio euclidiano $(\mathbb{E}, \omega)$ damos una descomposición en suma directa $\mathbb{E} = \mathbb{F} \oplus \mathbb{G}$ y P y S la proyección y simetría asociadas a la descomposición, ambas fijando los vectores de $\mathbb{F}$. Probar que S es una isometría si y solo sí $\mathbb{G} = \mathbb{F}^\perp$. ¿Es S autoadjunta? ¿Es P una isometría? ¿Es P autoadjunta? ¿Depende una respuesta afirmativa a lo anterior de que sea $\mathbb{G} = \mathbb{F}^\perp$? Suponemos dimensión finita.*

Problema 382 *Calcular en $\mathbb{R}^3$ con el producto ε para $v = (1,2,3)^\top$ las matrices de la proyección P sobre $\mathbb{H}$ y de S la simetría respecto de $\mathbb{H}$ ambas para $\mathcal{E}$. Comprobar en ambos casos que son matrices simétricas u ortogonales como asegura la teoría.*

Problema 383 *Probar que si L normal cumple $L^k(x) = 0$ para $k \geq 2$ entonces $L(x) = 0$. Indicación: hay que maniobrar hasta conseguir $\langle L^r(x), L^r(x)\rangle = 0$ para cierto r.*

[5] Para L arbitrario, L y L^* tienen los mismos valores propios pero pueden tener distintos vectores propios.

10.3. Los teoremas espectrales

El conjunto de autovalores de un endomorfismo se llama el **espectro**. Los teoremas que dan una estructura en función del espectro se llaman **teoremas espectrales.** Los teoremas 197 y 200 más abajo son teoremas espectrales para endomorfismos y formas bilineales simétricas. Tratamos primero los endomorfismos y con muy poco trabajo conseguiremos los teoremas para formas simétricas.

10.3.1. Teoremas espectrales para endomorfismos

Vamos a ver que los endomorfismos autoadjuntos son diagonalizables pero con un valor añadido: *se puede conseguir que las bases que diagonalizan sean ortonormales.* En el capítulo *Autovalores y autovectores* sí adelantamos al menos que las matrices simétricas eran diagonalizables, pero no podíamos ir más lejos sin tener el concepto de producto euclidiano. Interesa estudiar si los endomorfismos normales son diagonalizables, y de entrada hay que saber si $C(X)$ es linealmente factorizable.

Teorema 196 *Sea $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo y $C(X)$ su polinomio característico. Entonces*

1. *Si L es autoadjunto (caso $L^* = L$), todas las raíces de $C(X)$ son reales y L es diagonalizable.*
2. *Si L es antisimétrico (caso $L^* = -L$), todas las raíces de $C(X)$ son complejos imaginarios (vemos $C(X)$ como polinomio complejo con coeficientes en $\mathbb{R}$) y L no es diagonalizable.*
3. *Si L es isometría (caso $L^* = L^{-1}$), las raíces reales de $C(X)$ son ± 1.*

Demostración. En **1** y **2** tomamos la matriz a de L en una base ortonormal, y se ha visto en *Valores propios de matrices simétricas y hermitianas* que los valores propios de a son reales en **1** e imaginarios puros en **2**. Si L es autoadjunto, a es simétrica y vimos que a es diagonalizable y L también. El caso **3** se puede tratar directamente porque $L(x) = \lambda x$ nos da

$$\|x\|^2 = \langle x, L^* \circ L(x)\rangle = \langle L(x), L(x)\rangle = \lambda^2 \left\|x^2\right\|.$$

Cancelando $\|x\|^2$ llegamos a $\lambda = \pm 1$ si λ es real. ♣

Como contraejemplo consideramos en $\left(\mathbb{R}^2, \varepsilon\right)$,

$$a = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad aa^\top - a^\top a = 0, \quad C_a(X) = X^2 - 2X + 2 = (X - (1-i))(X - (1+i))$$

y vemos que a es normal pero no diagonalizable pues $C(X)$ no es linealmente factorizable en $\mathbb{R}$.

Teorema 197 *El endomorfismo $L : \mathbb{E} \to \mathbb{E}$ es autoadjunto si y solo si se puede diagonalizar con base ortonormal.*

Demostración. Supongamos L autoadjunto. El teorema 196 dice en **1** que L es diagonalizable y lo expresaremos en la forma $\mathbb{E} = \bigoplus_{j=1}^{p} \mathbb{E}(\lambda_j)$. Se usa ahora **4** en el teorema 195 y los subespacios $\mathbb{E}(\lambda_j)$ son perpendiculares entre sí. Entonces, yuxtaponiendo bases ortonormales de los $\mathbb{E}(\lambda_j)$, se tiene una base ortonormal de $\mathbb{E}$.

Recíprocamente, si es diagonalizable en base ortonormal, digamos $\mathcal{U}$, la matriz a de L en esa base es diagonal y, en particular simétrica. Sabemos que en términos matriciales, los endomorfismos que en una base ortonormal tienen matriz simétrica son autoadjuntos. ♣

Resaltemos otra vez que lo que cuenta no es solo poder diagonalizar, sino poder hacerlo *con base ortonormal.* Es posible que L sea diagonalizable pero que no sea autoadjunta para ω, y que las bases que diagonalicen L no sean ortonormales para ω. Tomemos $\mathbb{E} = \mathbb{R}^2$ y L y ω que en la base estándar tienen matrices

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Obviamente, cualquier base $\mathcal{U}$ que diagonalice L debe tener $u_1 \in \mathbb{E}(1)$ y $u_2 \in \mathbb{E}(2)$, pero $\mathbb{E}(1)$ y $\mathbb{E}(2)$ no son ortogonales para ω, y $\mathcal{U}$ no puede ser ortogonal y menos aún ortonormal.

Problema 384 *Sea L un endomorfismo. Probar que son equivalentes* ***(a)*** *L es a la vez isometría y autoadjunto;* ***(b)*** *L es la simetría respecto a una descomposición ortogonal $\mathbb{F} \oplus \mathbb{F}^\perp$ de $\mathbb{E}$.*

Los problemas de puro cálculo en que se pide una base ortonormal que diagonalice L se hacen en dos pasos calculando primero una base de vectores propios como en el capítulo *Autovalores y autovectores,* y ortogonalizando las partes de esa base en cada subespacio propio de L. En realidad se sigue la demostración del teorema 197.

Problema 385 *Consideramos el endomorfismo $L : \mathbb{R}^3 \to \mathbb{R}^3$ que en la base estándar tiene matriz*

$$a = \begin{pmatrix} \alpha & 1 & -1 \\ 1 & \alpha & 1 \\ -1 & 1 & \alpha \end{pmatrix}, \quad \alpha \in \mathbb{R}.$$

Construir una base ortonormal $\mathcal{W}$ para el producto estándar ε de $\mathbb{R}^3$ que diagonalice L. ¿Cómo será $\operatorname{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$? ♦

Solución. El polinomio característico con el cambio $Y = \alpha - X$ es

$$\begin{vmatrix} Y & 1 & -1 \\ 1 & Y & 1 \\ -1 & 1 & Y \end{vmatrix} = Y^3 - 3Y - 2 = (Y-2)(Y+1)^2 = ((\alpha-2)-X)((\alpha+1)-X)^2$$

con raíces $\lambda = \alpha + 1$ (doble) y $\mu = \alpha - 2$. Calculando los espacios de soluciones de $(a - \lambda)x = 0$ y $(a - \mu)x = 0$ obtenemos los espacios propios

$$\mathbb{E}(\alpha+1) = \operatorname{lg}\left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right), \quad \mathbb{E}(\alpha-2) = \operatorname{lg}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Si llamamos u_1, u_2, u_3 a los tres vectores calculados, observamos que $\varepsilon(u_1, u_3) = \varepsilon(u_2, u_3) = 0$, cosa que ya sabíamos pues son vectores propios con distintos valores propios. Sin embargo $\varepsilon(u_1, u_2) = -1$. Ortogonalizamos (u_1, u_2) con

$$v_1 = u_1, \quad v_2 = u_2 - \frac{\varepsilon(u_2, v_1)}{\|v_1\|^2} v_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{(-1)}{2}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{pmatrix}.$$

Ahora (v_1, v_2, u_3) es base ortogonal y, dividiendo por normas, obtenemos una base ortonormal. ♦

Problema 386 *Tomamos una matriz a muy parecida a la del problema precedente; en concreto,*

$$b = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

El ordenador nos dice que hay tres valores propios distintos, incluso hasta nos da valores propios, que son $\frac{3}{2} \pm \frac{1}{2}\sqrt{5}$ y 3 con correspondientes vectores propios

$$(u_1, u_2, u_3) = \left(\begin{pmatrix} \frac{1}{2} - \frac{1}{2}\sqrt{5} \\ -\frac{1}{2} + \frac{1}{2}\sqrt{5} \\ 1 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} + \frac{1}{2}\sqrt{5} \\ -\frac{1}{2} - \frac{1}{2}\sqrt{5} \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right).$$

Se ve enseguida que (u_1, u_2, u_3) no es ortonormal. ¿Contradice esto el teorema espectral?

Problema 387 *En $\mathbb{R}^3$ tomamos $v = (1, 0, 1)^\top$ y dos productos euclidianos, el estándar ε y otro ω que en la base estándar $\mathcal{E}$ tiene matriz*

$$\operatorname{mat}_{\mathcal{E}\mathcal{E}}(\omega) = \Omega = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

Los planos $\mathbb{H}$ y $\mathbb{K}$ ortogonales a v según ε y según ω son diferentes. Calcular las matrices de las respectivas simetrías S y T en $\mathcal{E}$, constatando que S tiene matriz simétrica pero T no. ¿Por qué? ¿Por qué estas dos matrices deben tener los mismos autovalores?

Es razonable preguntarse si hay muchas matrices normales que no correspondan a endomorfismos autoadjuntos o isometrías. El caso más sencillo en $\left(\mathbb{R}^2,\varepsilon\right)$ lleva cierto trabajo. Un pequeño atajo es escribir $M=s+a$ con s simétrica y a antisimétrica.

Problema 388 *Probar que una matriz* 2×2 *escrita como*

$$M=s+a=\begin{pmatrix} p & q\\ q & r\end{pmatrix}+\begin{pmatrix} 0 & -t\\ t & 0\end{pmatrix}$$

es normal si y solo si $tq=t\left(p-r\right)=0$ *¿Cuáles de estas son normales no diagonalizables?*

Problema 389 *¿Hay matrices en* $\mathbb{R}^3$ *normales y diagonalizables que no representen en la base estándar una proyección o una simetría en* $\left(\mathbb{R}^3,\varepsilon\right)$*?*

En el problema 211 se vio que la matriz

$$a=\begin{pmatrix} \alpha & 1 & 1 & \cdots & 1\\ 1 & \alpha & 1 & \cdots & 1\\ 1 & 1 & \alpha & \cdots & 1\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & 1 & \cdots & \alpha\end{pmatrix}$$

era diagonalizable, con polinomio característico

$$C\left(X\right)=\left(\left(\alpha-1\right)-X\right)^{n-1}\left(\left(\alpha-1\right)+n-X\right)\text{ con raíces }\lambda_1=\alpha+\left(n-1\right)\text{ y }\lambda_2=\alpha-1.$$

Los espacios propios son los espacios de soluciones de los sistemas homogéneos definidos por

$$a-\lambda_1 I=\begin{pmatrix} 1-n & 1 & 1 & \cdots & 1\\ 1 & 1-n & 1 & \cdots & 1\\ 1 & 1 & 1-n & \cdots & 1\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & 1 & \cdots & 1-n\end{pmatrix},\quad a-\lambda_2 I=\begin{pmatrix} 1 & 1 & 1 & \cdots & 1\\ 1 & 1 & 1 & \cdots & 1\\ 1 & 1 & 1 & \cdots & 1\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & 1 & \cdots & 1\end{pmatrix}.$$

Problema 390 *Dar una base* ortogonal *que diagonalice* a*.*

Solución (inicio) $\ker\left(a-\lambda_1 I\right)=\lg\left(u\right)=\lg\left(1,1,\ldots,1,1\right)^{\top}$ y lo costoso es encontrar una base ortogonal $\left(v_2,\ldots,v_n\right)$ de $\ker\left(a-\lambda_2 I\right)$ de dimensión $n-1$. Aconsejamos calcularla a ojo directamente, mejor que intentar aplicar el método de Gram-Schmidt a una base conocida. ♠

Teorema 198 (de Schur) *Sea* L *un endomorfismo de* $\mathbb{E}$ *con polinomio característico linealmente factorizable. Entonces hay una base ortonormal de* $\mathbb{E}$ *donde* L *tiene matriz triangular. Si* L *fuera autoadjunto, en esa base, la matriz sería diagonal.*

Demostración. La demostración se hará por inducción sobre $n=\dim\left(\mathbb{E}\right)$ y es obvia si $n=1$. Supuesto cierto el teorema para espacios $\mathbb{F}$ con $\dim\left(\mathbb{F}\right)<n$ se pueden seguir dos vías:

1. En el capítulo *Autovalores y autovectores* se probó un teorema similar para $\mathbb{E}$ y $\Bbbk$ arbitrarios. La demostración del presente teorema es similar con matices. Si $C\left(X\right)=\pm\left(X-\lambda_1\right)^{m_1}\cdots\left(X-\lambda_p\right)^{m_p}$, ha de haber un λ, digamos que es λ_1, tal que $\operatorname{im}\left(L-\lambda_1\right)=\mathbb{F}\neq\mathbb{E}$, porque si no se tendría que $C\left(L\right)\left(\mathbb{E}\right)=\mathbb{E}$. El subespacio $\mathbb{F}$ es estable por L, luego la restricción de L a $M:\mathbb{F}\to\mathbb{F}$ tiene también polinomio característico $Q\left(X\right)$ linealmente factorizable porque $Q\left(X\right)$ divide a $C\left(X\right)$ (teorema 98). La hipótesis inductiva nos da una base $\left(u_1,\ldots,u_m\right)$ ortonormal de $\mathbb{F}$ donde M tiene matriz triangular. Ampliamos esa base a una base también ortonormal $\mathcal{U}=\left(u_1,\ldots,u_m,\ldots,u_n\right)$ de todo $\mathbb{E}$. Afirmamos que en esa base L tiene matriz triangular. En efecto, si $j\leq m$, $L\left(u_j\right)=M\left(u_j\right)=\sum_{i=1}^{m}a_j^i u_i$. Tomando $j=m+k$, se cumple que $\left(L-\lambda_1\right)\left(u_{m+k}\right)\in\mathbb{F}$, luego

$$\left(L-\lambda_1\right)\left(u_{m+k}\right)=\sum_{i=1}^{m}a_{m+k}^i u_i\ \text{ y }\ L\left(u_{m+k}\right)=\sum_{i=1}^{m}a_{m+k}^i u_i+\lambda_1 u_{m+k},$$

que prueba lo afirmado.

2. Tomamos un vector propio v de L que también lo será de L^* y con el mismo valor propio λ (véase **3** en el teorema 195). Sea $\mathbb{F}$ de dimensión $n-1$ el subespacio ortogonal a v. Afirmamos que $L(\mathbb{F}) \subset \mathbb{F}$. En efecto, para $x \in \mathbb{F}$, luego $\langle x, v\rangle = 0$, se tiene $\langle L(x), v\rangle = \langle x, L^*(v)\rangle = \langle x, \lambda v\rangle = 0$. Aplicar la hipótesis inductiva a la restricción M de L a $\mathbb{F}$ es posible porque, como en la primera demostración, su polinomio característico es linealmente factorizable. Sea pues $(u_1, \ldots, u_{n-1})$ base ortonormal de $\mathbb{F}$ tal que $M(u_j) = \sum_{k=1}^{j} a_j^k u_k$, que completamos a una base ortonormal $(u_1, \ldots, u_{n-1}, u_n)$, siendo $u_n = v/\|v\|$. Ya es inmediato que esta base satisface el teorema.

Si L es autoadjunto, **1** en el teorema 196 da que $C(X)$ es linealmente factorizable. La matriz a en la base ortonormal $\mathcal{V}$ del teorema de Schur es triangular y simétrica (por ser L autoadjunta y $\mathcal{U}$ ortonormal), luego ambas cosas implican que a es diagonal y que la base $\mathcal{V}$ es adecuada. ♣

Es curioso que la idea esencial de la primera demostración del teorema de Schur es muy sencilla y vale aunque no estemos en un espacio euclidiano como indica el siguiente teorema.

Teorema 199 *Sea L un endomorfismo de $\mathbb{E}$ y λ un valor propio de L. Entonces, cualquier subespacio $\mathbb{G}$ conteniendo a $\mathbb{F} = \operatorname{im}(L-\lambda)$ es estable por L.*

Si L tiene polinomio característico linealmente factorizable hay una sucesión de subespacios estables

$$\mathbb{F}_0 = 0 \subset \mathbb{F}_1 \subset \mathbb{F}_2 \subset \ldots \subset \mathbb{F}_{n-1} \subset \mathbb{F}_n = \mathbb{E} \;\; con \;\; \dim(\mathbb{F}_j) = j.$$

Hay una base $(u_1, \ldots, u_n)$ de $\mathbb{E}$ de modo que $(u_1, \ldots, u_j)$ es base $\mathbb{F}_j$ y L tiene en ella matriz triangular.

Demostración. Lo primero es casi inmediato. Si $x \in \mathbb{G}$, $L(x) - \lambda x = y \in \mathbb{F}$ y $L(x) = \lambda x + y \in \mathbb{G}$ porque es suma de dos vectores de $\mathbb{G}$. La segunda parte se prueba por inducción tomando un valor propio λ y $\mathbb{F} = \operatorname{im}(L-\lambda)$. Si $\dim \mathbb{F} = m$ será $\mathbb{F} = \mathbb{F}_m$. Cualquier elección de subespacios $\mathbb{F}_m \subset \mathbb{F}_{m+1} \subset \ldots \subset \mathbb{F}_{n-1} \subset \mathbb{F}_n = \mathbb{E}$ será, por la primera parte, sucesión de subespacios estables. Para completar esta sucesión a $\mathbb{F}_0 = 0 \subset \mathbb{F}_1 \subset \ldots \subset \mathbb{F}_{m-1} \subset \mathbb{F}_m$ se aplica la hipótesis inductiva a la restricción de L a $\mathbb{F} = \mathbb{F}_m$. La base $\mathcal{U}$ se construye pasa a paso siendo (u_1) base de $\mathbb{F}_1$, (u_1, u_2) base de $\mathbb{F}_2$ ampliada de (u_1) y llegando a $(u_1, \ldots, u_j)$ base $\mathbb{F}_j$ por sucesivas ampliaciones se amplía a $(u_1, \ldots, u_j, u_{j+1})$ añadiendo un vector $u_{j+1} \in \mathbb{F}_{j+1} - \mathbb{F}_j$. ♣

Si en $\mathbb{E}$ hay un producto euclidiano, podemos conseguir cuidando detalles que $(u_1, \ldots, u_n)$ sea base ortonormal y $\mathbb{F}_j = \lg(u_1, \ldots, u_j)$. En efecto, cor la hipótesis inductiva se tendría $(u_1, \ldots, u_m)$ familia ortonormal con $\mathbb{F}_j = \lg(u_1, \ldots, u_j)$ para $j \leq m = \dim(\mathbb{F}) = \dim(\operatorname{im}(L-\lambda))$. Para cualquier base ortonormal ampliada $(u_1, \ldots, u_m, \ldots, u_n)$, basta definir $\mathbb{F}_j = \lg(u_1, \ldots, u_j)$ y no habrá problema cuando sea $j > m$ porque los subespacios, al contener a $\operatorname{im}(L-\lambda)$, son estables. Esto es prácticamente la misma demostración **1** del teorema de Schur.

10.3.2. Teoremas espectrales para formas simétricas

A cada endomorfismo *autoadjunto* L del espacio $(\mathbb{E}, \omega)$ le podemos asignar una forma bineal *simétrica* σ en $\mathbb{E}$ por $\sigma(x,y) = \omega(L(x), y)$. En dimensión finita es muy fácil comprobar que la asignación $L \to \sigma$ es un isomorfismo entre el espacio de endomorfismos autoadjuntos y el de las formas bilineales simétricas. Si L tiene en la base $\mathcal{U}$ matriz a, tenemos para x, y de coordenadas ξ, η que

$$\sigma(x,y) = \omega(L(x), y) = (a\xi)^\top \Omega\eta = \xi^\top a^\top \Omega\eta, \quad \text{siendo } \Omega = \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\omega),$$

de donde se deduce que la matriz de σ en $\mathcal{U}$ es $s = a^\top\Omega$. Con símbolos más detallados (y pesados)

$$\operatorname{mat}_{\mathcal{U}\mathcal{U}}(\sigma) = \operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L)^\top \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\omega). \tag{10.5}$$

Si $\mathcal{U}$ es ortonormal, $\operatorname{mat}_{\mathcal{U}\mathcal{U}}(\omega) = I$ y $\operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \operatorname{mat}_{\mathcal{U}\mathcal{U}}(\omega)$. Por tanto, *referidas a bases ortonormales, L y la forma bilineal σ que le corresponde, tienen la misma matriz* (simétrica, naturalmente).

Teorema 200 *Se tiene $\sigma(x,y) = \omega(L(x), y)$ con L autoadjunta y σ simétrica, y $\mathcal{W}$ una base ortonormal. El que $\mathcal{W}$ diagonalice L equivale a que diagonalice σ. Como L admite bases ortonormales diagonalizadoras, hay bases de este tipo que diagonalizan σ.*

Demostración. Es inmediata porque si $L(w_i) = \lambda_i w_i$ para $i = 1, \ldots, n$,

$$\sigma(w_i, w_j) = \langle L(w_i), w_j\rangle = \langle \lambda_i w_i, w_j\rangle = \lambda_i \delta_{ij}.$$

Recíprocamente, si $\sigma(w_i, w_j) = \lambda_i \delta_{ij}$ obtenemos para $j = 1, \ldots, n$ que

$$\langle L(w_i) - \lambda_i w_i, w_j \rangle = \sigma(w_i, w_j) - \lambda_i \delta_{ij} = 0.$$

Como $\mathcal{W}$ es una base, $\langle L(w_i) - \lambda_i w_i, x \rangle = 0$ para todo $x \in \mathbb{E}$ y, como ω es no degenerada, $L(w_i) = \lambda_i w_i$ para $i = 1, \ldots, n$. ♣

Problema 391 *Nos dan tres matrices*

$$\Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad s = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad t = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

siendo Ω es la matriz de un producto euclidiano ω y s y t son matrices de dos formas simétricas σ y τ, todas respecto de $\mathcal{E}$. Calcular bases ortonormales para ω que diagonalicen σ y τ. ♦

Solución del primer caso. Sea L el endomorfismo asociado a σ, que tendrá en $\mathcal{E}$ matriz a verificando $a^\top \Omega = s$ (ver 10.5). Calculamos

$$a = \Omega^{-1} s = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Desde luego a no es simétrica, pero no debe sorprendernos porque $\mathcal{E}$ no es ortonormal para ω. Es fácil calcular (olvidándonos por entero de productos euclidianos; en el espíritu del capítulo *Autovalores y autovectores*) que $\mathcal{U} = \left((-2,1)^\top, (0,1)^\top\right)$ es base de vectores propios. Como corresponden a valores propios diferentes, han de ser ortogonales para ω y, aunque no es necesario, podemos constatar que

$$\begin{pmatrix} -2 \\ 1 \end{pmatrix}^\top \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0.$$

Esta base $\mathcal{U}$, una vez normalizada, cumple lo pedido. El caso de t es para el lector. ♦

El lector se habrá dado cuenta de que si le dan los datos de σ en $\mathbb{E} = \mathbb{R}^n$ en la base estándar con s simétrica, y si $\omega = \varepsilon$, el hallar una base ortonormal $\mathcal{W}$ que diagonalice σ es simplemente cuestión de hallar una base ortonormal de vectores propios para L, *que tiene en* $\mathcal{E}$ la misma matriz s. No hay pues nada nuevo que aprender a nivel de cálculo. Sin embargo, si con los datos σ y s se toma otro producto ω con matriz $\Omega \neq I$, hay que buscar vectores propios de $a = \Omega^{-1} s$. El cálculo se alarga un poco, pero la dificultad es conceptual. Si el lector quiere algo más de entrenamiento le proponemos en $\mathbb{E} = \mathbb{R}^3$ el caso

$$\Omega = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}, \quad \Omega^{-1} = \begin{pmatrix} 2 & 0 & -1 \\ 0 & \frac{1}{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad s = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Problema 392 *Hallar una base ortonormal para ω que diagonalice σ.*

La segunda parte del teorema 200, que da la posibilidad de diagonalizar σ *con una base ortonormal,* se llama el **teorema espectral para formas bilineales simétricas.** Quizás el lector esté un poco confundido porque parece que dada una matriz $s \in \mathbb{R}^{n \times n}$ simétrica se la puede manipular como en el capítulo *Formas bilineales y cuadráticas* o como en este, llegándose a una matriz diagonal. ¿Se logra lo mismo? En *Formas bilineales y cuadráticas* obtenemos bases diagonalizadoras pero sin referencia a productos euclidianos. Si queremos que sean ortonormales se precisa más trabajo. Por ejemplo,

$$\left(\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \rightarrow\Rightarrow \left(\begin{array}{cc} 1 & 0 \\ 0 & 0 \\ \hline 1 & -1 \\ 0 & 1 \end{array}\right)$$

dice que la matriz arriba a la izquierda tiene a la base $(e_1, -e_1 + e_2)$ como base diagonalizadora, pero esta base no es ortonormal para ε. Sin embargo $\left(\frac{1}{\sqrt{2}}(e_1 + e_2), \frac{1}{\sqrt{2}}(e_1 - e_2)\right)$ es base diagonalizadora ortonormal (para ε) formada por vectores propios.

Repasemos, *pero añadiremos algo importante*. Sea s simétrica y $C(X) = (\lambda_1 - X)^{m_1} \cdots (\lambda_p - X)^{m_p}$. Representaremos por Λ la sucesión de valores propios con su multiplicidad; o sea,

$$\Lambda = \left(\overbrace{\lambda_1, \ldots, \lambda_1}^{m_1}, \ldots, \overbrace{\lambda_k, \ldots, \lambda_k}^{m_k}, \ldots, \overbrace{\lambda_p, \ldots, \lambda_p}^{m_p} \right),$$

y por d a una matriz diagonal donde Λ es la sucesión de los términos que aparecen en ella. El producto euclidiano en $\mathbb{R}^n$ es el estándar y $L(x) = sx$ y $\sigma(x, y) = x^\top sy$ representan un endomorfismo autoadjunto y una forma bilineal simétrica tales que $\langle L(x), y \rangle = \sigma(x, y)$.

1. Si usamos las técnicas de formas simétricas *antes de conocer los productos euclidianos*, sabemos que hay una matriz u invertible tal que $u^\top su = \Delta$ es una matriz diagonal. La base formada por las columnas de u diagonaliza σ, pero no tiene por qué ser base ortogonal u ortonormal para ε. No obstante sí que es cierto que si σ fuese definida positiva, sería entonces *otro producto euclidiano,* y las columnas de u serían una base ortogonal para σ.

2. Con las técnicas de este capítulo, sabemos que hay una matriz w tal que $w^{-1}sw = d$. Las columnas representan una *base ortonormal $\mathcal{W}$ para ε de vectores propios de L*. El que sea $\mathcal{W}$ base ortonormal para ε equivale a que w sea *matriz ortogonal, luego* $w^{-1} = w^\top$ y puede escribirse también $w^\top sw = d$. La base $\mathcal{W}$ de las columnas de w es *también* base que diagonaliza σ.

3. El trabajo para llegar a d y w es en principio mayor que el de llegar a Δ y u por ser azaroso conocer las raíces de $C(X)$. (no lo notamos porque los problemas suelen estar preparados.) En todo caso w tiene la ventaja de ser ortonormal y $w^{-1} = w^\top$. Puede ser desde luego $d \neq \Delta$ pero el teorema de Sylvester (teorema 161) dice que en ambas son formas diagonalizadas de la misma s, luego la cantidad de términos positivos, negativos y nulos (¡que dan la signatura!) en d y Δ ha de ser igual. Es mucho más fácil saberlo por d ya que los d_i^i son los valores propios de L; o sea, las raíces de $C_s(X)$. *Este es el punto al que hicimos referencia más arriba*: si solo interesa la signatura y hay facilidad de conocer los *signos* (no los valores exactos) de las raíces de $C_s(X)$, podemos evitar el trabajo que calcularla pasando de s a Δ.

Teorema 201 *Si L se corresponde con σ por $\sigma(x, y) = \langle L(x), y \rangle$ y $C_L(X)$ tiene p raíces > 0, q raíces < 0 y $n - (p + q)$ raíces 0, la signatura es (p, q).*

Problema 393 *Comparar los procedimiento para conocer rango y signatura (pero no más) de*

$$s = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Hallar en todo caso una base ortonormal para ε que diagonalice s. ♦

Solución. Se tiene que $C(X) = X(X-1)(X-2)$. La diagonal de la forma diagonal es, salvo permutación de términos, $(0, 1, 2)$ y la signatura es $(2, 0)$ por lo que acabamos de decir. Si se diagonaliza como en el capítulo anterior,

$$\left(\frac{s}{I}\right) = \left(\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc} 1 & 0 & 1-1 \\ 0 & 1 & 0-0 \\ 1 & 0 & 1-1 \\ \hline 1 & 0 & 0-1 \\ 0 & 1 & 0-0 \\ 0 & 0 & 1-0 \end{array}\right) = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ \hline 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) \Rightarrow \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) = \left(\frac{\Delta}{b}\right)$$

y la matriz Δ nos dice que la signatura es $(2, 0)$. Las matrices bajo la raya horizontal no son necesarias para calcular la signatura, pero las hemos puesto para que se pueda ver que, si bien las columnas de b dan una base que diagonaliza $\sigma(x, y) = x^\top sy$, esta base no diagonaliza $L(x) = sx$. Esta es la diferencia de "ver s como forma bilineal" o "ver s como endomorfismo autoadjunto". El ordenador comprueba que

$$b^\top sb = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \Delta,$$

$$b^{-1}sb=\begin{pmatrix}1&0&-1\\0&1&0\\0&0&1\end{pmatrix}^{-1}\begin{pmatrix}1&0&1\\0&1&0\\1&0&1\end{pmatrix}\begin{pmatrix}1&0&-1\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}2&0&0\\0&1&0\\1&0&0\end{pmatrix}\quad\text{(no diagonal)}$$

Por supuesto, s "visto como endomorfismo autoadjunto" se puede diagonalizar con la base ortonormal de vectores propios $\mathcal{U}=\left(\left(-1,0,1\right)^{\top},\left(0,1,0\right)^{\top},\left(1,0,1\right)^{\top}\right)$. ♦

Vuelve a aparecer la matriz del problema 211

$$a=\begin{pmatrix}\alpha&1&1&\cdots&1\\1&\alpha&1&\cdots&1\\1&1&\alpha&\cdots&1\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&1&1&\cdots&\alpha\end{pmatrix}$$

que era diagonalizable, con polinomio característico

$$C\left(X\right)=\left(\left(\alpha-1\right)-X\right)^{n-1}\left(\left(\alpha-1\right)+n-X\right)\text{ y raíces }\lambda_1=\alpha+\left(n-1\right)\text{ y }\lambda_2=\alpha-1.$$

Problema 394 *En $\mathbb{R}^n$ con $n\geq 2$ nos dan la forma cuadrática*

$$Q\left(x\right)=\sum_{i=1}^{n}\alpha\left(x^i\right)^2+2\sum_{1\leq i<j\leq n}x^ix^j,\quad\alpha>0,$$

con forma bilineal asociada σ. Se pide la signatura de σ o Q y si σ es un producto escalar.

Problema 395 *Calcular la signatura de la matriz simétrica*

$$a=\begin{pmatrix}1&1&0&1\\1&1&1&0\\0&1&1&1\\1&0&1&1\end{pmatrix}$$

cuyo polinomio característico es $C\left(X\right)=X^4-4X^3+2X^2+4X-3$.

Ciertos teoremas sobre la situación de las raíces de un polinomio permiten obtener información sobre la signatura de s simétrica. Supongamos por ejemplo que $C\left(X\right)$ tuviese todos sus coeficientes >0. No existirían entonces raíces $\lambda>0$ y la signatura de s sería de la forma $\left(0,q\right)$. El teorema más conocido[6] es la **regla de Descartes**. Nos dan un polinomio $P\left(X\right)=a_nX^n+a_{n-1}X^{n-1}+\ldots+a_1X+a_0$ y contamos los cambios de signo (positivo a negativo o negativo a positivo, ignorando los $a_j=0$). Por ejemplo, si $P\left(X\right)=X^5-3X^3-X^2+6$, hay los cambios de signo $1\to-3$ y $-1\to6$. Sea este número $V=V\left(P\left(X\right)\right)$, que es 2 en el ejemplo. Entonces **(a)** El número de raíces >0, *contadas con su multiplicidad* es de la forma $V-2k$ con $k\in\mathbb{Z}$ pero indeterminado. **(b)** El número de raíces <0 es el número de raíces >0 de $Q\left(X\right)=P\left(-X\right)$ y se calcula como en **(a)**. El ordenador calcula para $P\left(X\right)=X^5-3X^3-X^2+6$ una raíz real <0 y cuatro complejas. La regla de Descartes da 2 o 0 raíces >0 y, como $P\left(-X\right)=-X^5+3X^3-X^2+6$ y $V\left(P\left(-X\right)\right)=3$ hay 3 o 1 raíces <0. Si exponemos en la forma $\left(p,q\right)$ en número de raíces positivas p y negativas q, la regla de Descartes ofrece $\left(2,3\right)$, $\left(2,1\right)$, $\left(0,3\right)$ y $\left(0,1\right)$ y de las cuatro posibilidades $\left(0,1\right)$ la correcta.

Problema 396 *Consideremos la matriz simétrica*

$$a=\begin{pmatrix}2&1&0&0\\1&1&1&0\\0&1&1&1\\0&0&1&0\end{pmatrix}$$

que tiene $C\left(X\right)=X^4-4X^3+2X^2+4X-1$. Probar con $C_a\left(X\right)$ que su signatura no puede ser $\left(2,2\right)$. ¿Puede ser $\left(4,0\right)$ o $\left(0,4\right)$? ¿Puede ser $\left(3,0\right)$?

Si se quiere calcular *exactamente* la signatura de a y no meramente excluir algunos casos, se ve que el procedimiento de operaciones columna/fila puede ser mucho más fiable. En este caso las raíces de $C_a\left(X\right)$ son $1\pm\sqrt{2+\sqrt{2}}$ y $1\pm\sqrt{2-\sqrt{2}}$ que no son fáciles de calcular, aunque si se obtienen o aproximan lo suficiente para saber que hay tres positivas y una negativa, la signatura de a es $\left(3,1\right)$.

[6] El **teorema de Budan-Fourier** mejora la regla de Descartes pues con una cuenta de cambios de signo da, con error de números pares, el número de raíces en un intervalo (y no meramente en $(0,\infty)$ y $(-\infty,0)$).

10.3.3. Un resumen de los teoremas espectrales

El adjetivo hace referencia al **espectro**, que es el conjunto de valores propios de un endomorfismo si bien lo importante de todos ellos es la posibilidad de diagonalizar con bases ortonormales. Hay tres elementos (ω, L, σ) siendo ω un producto euclidiano sobre $\mathbb{E}$, L un endomorfismo de $\mathbb{E}$ y σ una función bilineal simétrica sobre $\mathbb{E}$, ligados por $\sigma(x,y) = \omega(L(x), y)$, con lo cual L es autoadjunto.

1. El que L sea autoadjunta permite que L sea diagonalizable con base ortonormal para ω. Si cambia ω a ω', L deja de ser autoadjunta y, aunque sigue siendo diagonalizable, ya no lo será para una base ortonormal de ω'.

2. La base ortonormal que diagonaliza L, también diagonaliza σ. Esto permite conocer la signatura de σ contando con su multiplicidad los valores propios positivos y negativos de $C_L(X)$.

3. Si se toman como datos principales ω y σ; es decir, si se tienen dos formas bilineales simétricas con ω definida positiva, hay una base común que diagonaliza ambas formas. La demostración es sencilla: con ω y σ se construye L y la base ortogonal para ω que diagonaliza L, cumple las condiciones. Este teorema se llama el **teorema de los ejes principales**, por su interpretación con las cuádricas.

4. Estos teorema se pueden ver como teoremas que muestran que una matriz simétrica s es congruente o semejante a otra diagonal. Sin necesitar productos euclidianos ya sabemos que s es congruente con d diagonal; o sea, existe c invertible tal que $c^\top sc = d$. Sin embargo si s se identifica con $L : \mathbb{R}^n \to \mathbb{R}^n$, $L(x) = sx$ y $\omega = \varepsilon$, L es diagonalizable con $\mathcal{U} = (u_1, \ldots, u_n)$ ortonormal. Al yuxtapones las columnas de $(u_1, \ldots, u_n)$ tenemos una base ortogonal u y $u^{-1}su = \Delta$ diagonal. La ortogonalidad de u hace que sea $u^{-1} = u^\top$, luego también $u^\top su = \Delta$.

5. Hay una forma del teorema espectral para endomorfismos que se basa en el enfoque de los endomorfismos diagonalizables del teorema 130 del capítulo *Potencias y exponencial de endomorfismos. Si* $(\lambda_1, \ldots, \lambda_p)$ son los *distintos* valores propios, y P_j la proyección *ortogonal* sobre el subespacio propio $\mathbb{E}(\lambda_j)$ se tiene que estos subespacios son mutuamente perpendiculares y $L = \lambda_1 P_1 + \ldots + \lambda_p P_p$. Sin productos euclidianos hay un teorema similar pero que no hace referencia a ortogonalidad.

Las matrices simétricas son las más ricas en propiedades de todas las matrices.

10.4. Cuádricas en el espacio euclidiano

Sea $\mathbb{E}$ un espacio euclidiano con un producto $\omega = \langle \bullet, \bullet \rangle$. Recordamos que una **cuádrica** es un conjunto de puntos $\mathcal{C}$ de $\mathbb{E}$ con ecuación $\phi(x) = 0$, siendo ϕ de la forma $\phi(x) = \sigma(x,x) + 2f(x) + \alpha$ con σ bilineal simétrica, $f : \mathbb{E} \to \mathbb{R}$ lineal y $\alpha \in \mathbb{R}$. Nos interesa en adelante considerar $L : \mathbb{E} \to \mathbb{E}$ autoadjunto tal que $\sigma(x,y) = \langle L(x), y \rangle = \langle x, L(y) \rangle$ y f de la forma $f(x) = \langle F, x \rangle$ con $F \in \mathbb{E}$. Por consiguiente, $\phi(x) = \langle L(x), y \rangle + 2\langle F, x \rangle + \alpha$. El caso particular más importante supone $\mathbb{E} = \mathbb{R}^n$ y $L(x) = sx$ con s simétrica. Seguiremos en paralelo el tratamiento de las cuádricas del caso afín pero teniendo en cuenta que ahora hay estructura adicional en $\mathbb{E}$. Según nuestra conveniencia pondremos la ecuación en una u otra de las formas

$$\phi(x) = \sigma(x,x) + 2f(x) + \alpha = \langle L(x), x \rangle + 2\langle F, x \rangle + \alpha.$$

En general, se prefiere la forma con σ y f para no tener que repetir cálculos ya hechos, y la forma "asimétrica" con L y F porque es más clara a la hora de buscar una referencia adecuada para expresar la ecuación de $\mathcal{C}$. Será muy útil la fórmula

$$\phi(z+y) = \sigma(y,y) + 2[\sigma(z,y) + f(y)] + \phi(z) = \langle L(y), y \rangle + 2\langle L(z) + F, y \rangle + \phi(z) \tag{10.6}$$

Usaremos sobre todo la segunda forma. Si se quiere deducirla directamente como entrenamiento

$$\begin{aligned}
\phi(z+y) &= \langle L(z) + L(y), z+y \rangle + 2\langle F, z+y \rangle + \alpha \\
&= \langle L(z), z \rangle + \langle L(z), y \rangle + \langle L(y), z \rangle + \langle L(y), y \rangle + 2\langle F, z \rangle + 2\langle F, y \rangle + \alpha \\
&\overset{*}{=} \langle L(y), y \rangle + 2[\langle L(z), y \rangle + \langle F, y \rangle] + [\langle L(z), z \rangle + 2\langle F, z \rangle + \alpha] \\
&= \langle L(y), y \rangle + 2\langle L(z) + F, y \rangle + \phi(z).
\end{aligned}$$

En el paso $\overset{*}{=}$ se utiliza que, por ser L autoadjunta, $\langle L(z), y\rangle + \langle L(y), z\rangle = 2\langle L(z), y\rangle$.

El concepto de centro y su tratamiento no varía en absoluto respecto al caso afín puesto que no se usa en absoluto la existencia de un producto euclidiano. Reescribimos el teorema 162.

Teorema 202 *La simetría central $\Sigma_z(x) = 2z - x$ cumple $\phi(\Sigma_z(x)) = \phi(x) + 4(\langle L(z) + F, z - x\rangle)$. Si $\mathcal{C}$ no está contenida en un hiperplano afín se tiene que z es un centro de $\mathcal{C}$ si y solo si $L(z) + F = 0$.*

Los posibles centros de $\mathcal{C}$ son las soluciones de $L(x) + F = 0$. Evidentemente, si L es un isomorfismo o si $F = 0$, hay centros.

10.4.1. Ecuación normalizada si hay centros

Recordamos que una referencia afín es una sucesión $\mathcal{R} = (z, u_1, \ldots, u_n)$ en $\mathbb{E}$ con $(u_1, \ldots, u_n)$ base de $\mathbb{E}$. Para el producto ω nos interesan las **referencias euclidianas**, que son aquellas en donde $\mathcal{U} = (u_1, \ldots, u_n)$ es base ortonormal para ω. Las coordenadas $(x^1, \ldots, x^n)$ de x en $\mathcal{R}$ se calculan como con las afines por $x - z = \sum_{i=1}^n x^i u_i$, advirtiéndose que no hay que confundir las coordenadas en $\mathcal{U}$ con las coordenadas en $\mathcal{R}$. La ventaja de disponer de coordenadas en referencia euclidiana es que las distancias se calculan "como en $\mathbb{R}^n$ "; es decir,

$$d(x,y)^2 = \|x-y\|^2 = \left\| z + \sum_{i=1}^n x^i u_i - z - \sum_{i=1}^n y^i u_i \right\|^2 = \left\| \sum_{i=1}^n \left(x^i - y^i\right) u_i \right\|^2 = \sum_{i=1}^n \left(x^i - y^i\right)^2.$$

Nuestro objetivo es encontrar una referencia euclidiana $\mathcal{R}$ en donde $\mathcal{C}$ tenga una ecuación "sencilla e informativa". Tendremos la ventaja de que en aquellos conceptos referentes a $\mathcal{C}$ donde intervenga la longitud, podremos calcular en coordenadas. Volveremos sobre esto. Tal como hicimos para cuádricas afines tratamos primero el caso en el que hay centros. El papel que allí representó la base de Sylvester, lo desempeñará aqui una base de vectores propios de L autoadjunta (véase el teorema 197).

Teorema 203 *Si $\mathcal{C}$ tiene un centro z, en una referencia euclidiana $\mathcal{R} = (z, u_1, \ldots, u_n)$ con $\mathcal{U}$ base ortonormal de vectores propios de L, la cuádrica tiene ecuación*

$$0 = \phi(x) = \sum_{i=1}^n \lambda_i \left(x^i\right)^2 + \phi(z),$$

siendo los λ_i los valores propios de L.

Demostración. Se toma $\mathcal{R}$ como se ha dicho. Utilizando (10.6), al ser $L(z) + F = 0$,

$$\begin{aligned}\phi(x) = \phi\left(z + \sum_{i=1}^n x^i u_i\right) &= \left\langle L\left(\sum_{i=1}^n x^i u_i\right), \sum_{i=1}^n x^i u_i \right\rangle + \phi(z) \\ &= \left\langle \sum_{i=1}^n x^i \lambda_i u_i, \sum_{i=1}^n x^i u_i \right\rangle + \phi(z) = \sum_{i=1}^n \lambda_i \left(x^i\right)^2 + \phi(z).\end{aligned}$$

♣

Queda un punto algo delicado. Nos gustaría que $\mathcal{C}$ determinase unívocamente a los λ_i pero no es así. Si estamos en el caso $z \in \mathcal{C}$, será $\phi(z) = 0$ y la ecuación $0 = \sum_{i=1}^n \lambda_i \left(x^i\right)^2$ no altera sus soluciones al ser multiplicada por $\mu \neq 0$. Los coeficientes de los $\left(x^i\right)^2$ no dependen tan solo de $\mathcal{C}$, pero sí es cierto que dependen solo de $\mathcal{C}$ *salvo factor multiplicativo*. En la medida en que nos importen, por ejemplo, cocientes del tipo λ_i/λ_j, tales cocientes solo dependen de $\mathcal{C}$.

El caso $z \notin \mathcal{C}$ es mejor porque $\phi(z) \neq 0$ y se puede dividir por $-\phi(z)$ para que quede una ecuación

$$\sum_{i=1}^n \frac{\lambda_i}{-\phi(z)} \left(x^i\right)^2 = 1. \tag{10.7}$$

Ahora, los coeficientes $-\lambda_i/\phi(z)$ dependen tan solo de $\mathcal{C}$ y no de la ecuación elegida, suponiendo (razonablemente) que $\mathcal{C}$ no está contenida en un subespacio afín estricto de $\mathbb{E}$. En efecto, en tal caso,

otra ecuación $\phi'(x) = \langle L'(x), x\rangle + 2\langle F', x\rangle + \alpha' = 0$ verifica la proporcionalidad $L' = \mu L$, $F' = \mu F$ y $\alpha' = \mu\alpha$. Advertimos que los valores propios de $L' = \mu L$ son de la forma $\lambda_i' = \mu\lambda_i$. Por otra parte, para el centro z, que lo es con independencia de la ecuación elegida, $\phi'(z) = \mu\phi(z)$ y la ecuación de $\mathcal{C}$ escrita en la forma (10.7) tiene carácter intrínseco. *Con mayor precisión: los coeficientes de las* $\left(x^i\right)^2$, *salvo por su ordenación, solo dependen de* $\mathcal{C}$.

Supongamos por ejemplo que en el caso $n = 2$ tenemos que $\mathcal{C}$ tiene una ecuación $x^2 - 4y^2 = 0$ (ponemos x, y en vez de x^1, x^2). Representa $\mathcal{C}$ el par de rectas $\mathbb{D}_1 = x + 2y = 0$ y $\mathbb{D}_2 = x - 2y = 0$. Valdría también $6x^2 - 24y^2 = 0$, pero el cociente $\lambda_1/\lambda_2 = -1/2$ es intrínseco y refleja el ángulo entre $\mathbb{D}_1$ y $\mathbb{D}_2$. (Preferimos no extendernos más porque habría que desarrollar el concepto del ángulo entre dos rectas.) Si tenemos una elipse, la ecuación (10.7) se suele escribir en la forma

$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1, \quad \alpha \geq \beta > 0.$$

Los puntos de coordenadas $(\pm\alpha, 0)$ están a distancia máxima del centro $(0,0)$ y los puntos $(0, \pm\beta)$ están a distancia mínima, todos ellos puntos de la elipse. Son las longitudes de los ejes de la elipse.

Problema 397 *Estudiar en* $\left(\mathbb{R}^2, \varepsilon\right)$ *la cónica de ecuación*

$$\phi\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + 2\begin{pmatrix} 6 \\ -9 \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} + 3 = 0.$$

Dando una referencia en donde tenga una ecuación del tipo (10.7). ♦

Solución. A estas alturas ya es fácil saber que los vectores y valores propios de L son

$$(v_1, v_2) = \left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right), \qquad (\lambda_1, \lambda_2) = (-1, 3).$$

Hemos puesto $\mathcal{V} = (v_1, v_2)$ porque todavía no es base ortonormal, pero $\mathcal{U} = (u_1, u_2)$ con $u_i = \left(1/\sqrt{2}\right) v_i$ ya es base ortonormal. El posible centro $z = (p, q)^\top$ es la solución de

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} 6 \\ -9 \end{pmatrix}, \text{ que es } \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -6 \\ 9 \end{pmatrix} = \begin{pmatrix} 8 \\ -7 \end{pmatrix}.$$

Se calcula que $\phi(z) = 114 \neq 0$ luego hay un solo centro z que, además, no está en $\mathcal{C}$. Como la signatura de σ es $(1,1)$, tenemos una hipérbola que en las coordenadas (u, v) asociadas a

$$\mathcal{R} = (z, u_1, u_2) = \left(\begin{pmatrix} -8 \\ 7 \end{pmatrix}, \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\right)$$

tiene ecuación $-u^2 + 3v^2 + 114 = 0$. En la forma normalizada la ecuación es

$$\frac{-u^2}{-114} + \frac{3v^2}{-114} + \frac{114}{-114}; \text{ es decir } \frac{u^2}{114} - \frac{v^2}{-38} = 1. \blacklozenge$$

Problema 398 *Hacer lo mismo con*

$$\phi\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + 2\begin{pmatrix} 3 \\ -6 \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} - h = 0,$$

cuando h permita que $\mathcal{C}$ sea una elipse. Y si no, ¿qué pasa?

10.4.2. Ecuación normalizada si no hay centros

Observamos que si $L : \mathbb{E} \to \mathbb{E}$ es un endomorfismo *simétrico*, entonces $(\operatorname{im}(L))^\perp = \ker(L)$. En efecto, si $x \in (\operatorname{im}(L))^\perp$ se tiene $\langle x, L(y)\rangle = 0$ para todo y, pero entonces $0 = \langle L(x), y\rangle$, luego $L(x)$ es ortogonal a todo $\mathbb{E}$ y por tanto, $L(x) = 0$. Se tienen entonces las *descomposiciones ortogonales*

$$\mathbb{E} = (\operatorname{im}(L))^\perp \oplus \operatorname{im}(L) = \ker(L) \oplus (\ker(L))^\perp = \ker(L) \oplus \operatorname{im}(L), \tag{10.8}$$

siendo esta última la que más nos va a interesar. Suponiendo en adelante que no hay centros, descomponemos $F = F_0 + F_1$ con $F_0 \in \ker(L)$ y $F_1 \in \operatorname{im}(L)$, debiendo ser $F_0 \neq 0$, pues si no, $F = F_1 \in \operatorname{im} L$ y $L(x) + F = 0$ tiene soluciones. Tomamos z tal que $L(z) + F_1 = 0$. Una referencia con origen este z tendría una "buena" ecuación, pero afinaremos eligiendo de entre las soluciones de $L(x) + F = 0$ una $\bar{z}$ que *además* cumpla $\phi(\bar{z}) = 0$. Escribimos $\bar{z} = z + \lambda F_0$ con $\lambda \in \mathbb{R}$ a determinar y (10.6) con $y = \lambda F_0$ da

$$\phi(\bar{z}) = \phi(z + \lambda F_0) = \langle L(\lambda F_0), \lambda F_0 \rangle + 2\langle L(z) + F, \lambda F_0 \rangle + \phi(z) = 2\lambda \|F_0\|^2 + \phi(z).$$

Por tanto, para $\lambda = -\phi(z)/2\|F_0\|^2$ se tiene que $\bar{z} = z + \lambda F_0$ cumple $L(\bar{z}) + F_1 = 0$ y $\phi(\bar{z}) = 0$. Entonces se simplifica mucho la expresión de $\phi(\bar{z} + y)$ porque

$$\phi(\bar{z} + y) = \langle L(y), y \rangle + 2\langle L(\bar{z}) + F, y \rangle + \phi(\bar{z}) = \langle L(y), y \rangle + 2\langle F_0, y \rangle. \tag{10.9}$$

Antes de construir la referencia $\mathcal{R}$ en donde $\mathcal{C}$ tendrá la ecuación normalizada, hacemos una última observación. Si no hay centros, L no puede ser invertible, pues si lo fuera habría centros (la solución única de $L(x) + F = 0$). Ordenamos los valores propios $(\lambda_1, \ldots, \lambda_n)$ de L de forma que los últimos $\lambda_{r+1}, \ldots, \lambda_n$ sean los nulos y $r < n$. Construimos ahora una base $(u_1, \ldots, u_n)$ de vectores propios donde u_{r+1} sea $-F_0/\|F_0\|$.

Teorema 204 *En la referencia $\mathcal{R} = (\bar{z}, u_1, \ldots, u_n)$ que acabamos de describir, la ecuación de $\mathcal{C}$ es*

$$\sum_{i=1}^{r} \lambda_i \left(x^i\right)^2 - 2\|F_0\| x^{r+1} = 0,$$

siendo $(\lambda_1, \ldots, \lambda_r)$ los valores propios no nulos de L.

Demostración. Para $x = \bar{z} + y = \bar{z} + \sum_{i=1}^{n} x^i u_i$ tenemos con (10.9)

$$\begin{aligned}\phi(\bar{z} + y) = \langle L(y), y \rangle + 2\langle F_0, y \rangle &= \left\langle L\left(\sum_{i=1}^{n} x^i u_i\right), \sum_{i=1}^{n} x^i u_i \right\rangle - 2\|F_0\| \left\langle u_{r+1}, \sum_{i=1}^{n} x^i u_i \right\rangle \\ &= \left\langle \sum_{i=1}^{n} \lambda_i x^i u_i, \sum_{i=1}^{n} x^i u_i \right\rangle - 2\|F_0\| x^{r+1} = \sum_{i=1}^{r} \lambda_i \left(x^i\right)^2 - 2\|F_0\| x^{r+1},\end{aligned}$$

lo que nos da la ecuación igualando a cero. ♣

Interesa dividir por $2\|F_0\|$ y entonces la ecuación es

$$\sum_{i=1}^{r} \frac{\lambda_i}{\|F_0\|} \left(x^i\right)^2 = x^{r+1}.$$

Si ignoramos las coordenadas $x^{r+2}, \ldots, x^n$, vemos que la cuádrica es el grafo de

$$f : \mathbb{R}^r \to \mathbb{R}, \qquad f\left(x^1, \ldots, x^r\right) = \sum_{i=1}^{r} \frac{\lambda_i}{\|F_0\|} \left(x^i\right)^2.$$

Teorema 205 *Dada $\mathcal{C}$ con ecuación $\phi(x) = \langle L(x), x \rangle + 2\langle F, x \rangle + \alpha = 0$ hay ecuaciones estándar de la siguiente forma*

1. *Si hay un centro $z \in \mathcal{C}$, en una referencia $\mathcal{R} = (z, u_1, \ldots, u_m)$ donde $\mathcal{U} = (u_1, \ldots, u_m)$ es base ortonormal de vectores propios de L, se tiene una ecuación en las coordenadas de $\mathcal{R}$*

$$\sum_{i=1}^{n} \lambda_i \left(x^i\right)^2 = 0.$$

2. *Si hay centros z, pero ninguno en $\mathcal{C}$, en una referencia $\mathcal{R} = (z, u_1, \ldots, u_m)$ donde $\mathcal{U} = (u_1, \ldots, u_m)$ es base ortonormal de vectores propios de L, se tiene una ecuación en las coordenadas de $\mathcal{R}$,*

$$\sum_{i=1}^{n} \lambda_i \left(x^i\right)^2 + \phi(z) = 0.$$

3. *Si no hay centros, se descompone* $F = F_0 + F_1$ *con* $F_0 \in \ker(L)$ *y* $F_1 \in \operatorname{im}(L)$, *debiendo ser* $F_0 \neq 0$. *Hay entonces una referencia* $\mathcal{R} = (\bar{z}, u_1, \ldots, u_m)$ *donde* $\mathcal{U} = (u_1, \ldots, u_m)$ *es base ortonormal de vectores propios y de* L *tal que* ***(a)*** $u_{r+1}, \ldots, u_m$ *son los elementos de* $\mathcal{U}$ *con valor propio* 0; *y* ***(b)*** *Sea* $u_{r+1} = -F_0/\|F_0\|$. *La cuádrica* $\mathcal{C}$ *tiene ecuación en las coordenadas de* $\mathcal{R}$

$$\sum_{i=1}^{r} \lambda_i \left(x^i\right)^2 - 2\|F_0\| x^{r+1} = 0.$$

Las ecuaciones en ***2*** *y* ***3*** *se pueden poner en la forma*

$$\sum_{i=1}^{n} \frac{-\lambda_i}{\phi(z)} \left(x^i\right)^2 = 1, \qquad \sum_{i=1}^{r} \frac{\lambda_i}{\|F_0\|} \left(x^i\right)^2 = x^{r+1}$$

y los coeficientes de los x_i^2 *solo dependen de* $\mathcal{C}$.

Si no nos piden la referencia, se puede abreviar el cálculo recordando que las raíces de $C_L(X)$ son los valores propios.

Problema 399 *Encontrar las ecuaciones y referencia del teorema 205 para la cuádrica*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}^{\top} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 2\begin{pmatrix} 5 \\ -1 \\ -1 \end{pmatrix}^{\top} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + 4 = 0. \blacklozenge$$

Solución. Con el cambio $Y = 1 - X$ obtenemos $C(X) = Y^3 - 3Y + 2$ con raíces 1 (doble) y -2, que dan las raíces 0 (doble) y 3 de $C(X)$. Supongamos que no piden la referencia. El cálculo es bastante rápido ya que la sucesión $(\lambda_1, \lambda_2, \lambda_3)$ de valores propios es $(3, 0, 0)$. Por otra parte, la ecuación $L(x) + F = 0$ no tiene solución, luego en las coordenadas (u, v, w) de la referencia $\mathcal{R}$, la ecuación es $u^2 - 2\|F_0\| v = 0$. Queda conocer F_0. Descomponemos F en $F_0 + F_1 \in \ker(L) \oplus \operatorname{im}(L)$. Es obvio que $h = (1, 1, 1)^{\top}$ genera $\operatorname{im}(L)$ y por ello,

$$F_1 = \frac{\langle F, h\rangle}{\|h\|^2} h = \frac{3}{3} h = h, \qquad F_0 = F - F_1 = \begin{pmatrix} 5 \\ -1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \\ -2 \end{pmatrix}.$$

La ecuación buscada es $u^2 - 4\sqrt{6}v = 0$.

Si piden la referencia, debemos trabajar más. Calculamos primero una base de vectores propios

$$\mathcal{V} = (v_1, v_2, v_3) = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\right)$$

con v_1 el vector propio de $\lambda_1 = 3$. La base $\mathcal{V}$ no es ortogonal, pero podemos ortogonalizarla, bien con Gram-Schmidt o bien usando que al ser $\|v_2\|^2 = \|v_3\|^2 = 2$, los vectores $w_2 = v_2 + v_3$ y $w_2 = v_2 - v_3$. son ortogonales. Optamos por ello porque ilustra un procedimiento menos frecuente. La base

$$\mathcal{W} = (w_1, w_2, w_3) = \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right)$$

es ortonormal y, una vez normalizada, nos da

$$\mathcal{U} = (u_1, u_2, u_3) = \left(\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{6}}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right).$$

Esta es la base de la referencia $\mathcal{R}$. Solo queda calcular el origen $\bar{z}$. Primero hay que calcular una solución z de $L(x) + F_1 = 0$ y tomamos $z = (1/3)\, h = (1/3)\,(1, 1, 1)^{\top}$. Tendremos que $\bar{z} = z + \lambda F_0$ siendo

$$\lambda = \frac{-\phi(z)}{2\|F_0\|^2} = \frac{-7}{192}. \blacklozenge$$

El lector se convencerá pronto que los casos en que no hay centros son los que dan más trabajo.

Problema 400 *Clasificar las cónicas $\mathcal{C}_i$ de ecuación dada por ϕ_i,*

$$\phi_1\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} x \\ y \end{pmatrix}^T\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+2x-2y+1,\quad \phi_2\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} x \\ y \end{pmatrix}^T\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+2x+4y.$$

Problema 401 *Determinar el tipo de la cuádrica en $\mathbb{R}^3$ dada para cada $h\in\mathbb{R}$ por*

$$\phi(x)=\begin{pmatrix} x \\ y \\ z \end{pmatrix}^T\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}+2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}^T\begin{pmatrix} x \\ y \\ z \end{pmatrix}+h$$

Problema 402 *Clasificar la cónica dada por $\mathcal{C}: x^2+2hxy+y^2+2\left(1-h^2\right)x=0$, $0<h<1$ y, cuando sea una elipse, dar la longitud de su eje mayor. Nota: para abreviar los cálculos,*

$$\begin{pmatrix} 1 & h \\ h & 1 \end{pmatrix}^{-1}=\begin{pmatrix} -\frac{1}{h^2-1} & \frac{h}{h^2-1} \\ \frac{h}{h^2-1} & -\frac{1}{h^2-1} \end{pmatrix}.$$

Problema 403 *Clasificar la cuádrica de $\mathbb{R}^3$ dada por*

$$\phi\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} x \\ y \\ z \end{pmatrix}^T\begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}-3=0.$$

10.5. Isometrías o transformaciones ortogonales

En toda esta sección suponemos que L es un endomorfismo del espacio euclidiano $(\mathbb{E},\omega)$ de dimensión finita n que cumple $L^{-1}=L^*$, que se suele poner como $\langle L(x),L(y)\rangle=\langle x,y\rangle$. Se llaman estas funciones **isometrías** o **transformaciones ortogonales**. Por la propia definición queda claro que L es isometría respecto a cierto ω y puede perder el carácter si cambia ω. He aquí propiedades generales de las isometrías, algunas ya probadas.

1. Si en una base $\mathcal{U}$ arbitraria L tiene matriz a y ω matriz Ω, la condición equivalente a ser isometría es a sea invertible y $a^\top\Omega a=\Omega$. Si $\mathcal{U}$ es es ortonormal, $a^\top a=I$ y L tiene en base ortonormal matriz ortonormal, lo que implica $\det(L)=\pm1$. (Puede verse el teorema 194 si se quiere más detalle.)
2. Las isometrías preservan productos escalares, normas, ángulos y distancias.
3. Dada la descomposición $\mathbb{E}=\mathbb{F}\oplus\mathbb{F}^\perp$, las simetrías asociadas a ella son isometrías.
4. El conjunto $\mathcal{I}_\omega(\mathbb{E})$ de las isometrías de $\mathbb{E}$ para ω forman un grupo (no conmutativo) donde el producto es la composición de isometrías, el inverso es la isometría inversa, y la unidad es $\mathrm{id}_{\mathbb{E}}$.

Problema 404 *Probar lo que se dice en **4** sobre el grupo de las isometrías.*

Problema 405 *Probar lo que se dice en **2** sobre las isometrías. Añadimos que si una función $L:\mathbb{E}\to\mathbb{E}$, cumple $L(0)=0$ y $d(L(x),L(y))=d(x,y)$, entonces, aunque en principio no sea lineal, estas condiciones implican que sí lo es, y L es una isometría.*

Se tiene pues una definición equivalente, muy intuitiva pero que no usaremos apenas: una isometría lineal es una función de $\mathbb{E}$ en $\mathbb{E}$ que fija el origen y preserva distancias.

Solución parcial. De $d(L(x),L(y))=d(x,y)$ para $y=0$ se sigue que $\|L(x)\|=\|x\|$. Usamos que

$$\|x-y\|^2=\|x\|^2+\|y\|^2-2\langle x,y\rangle,\quad 2\langle x,y\rangle=\|x\|^2+\|y\|^2-\|x-y\|^2$$

y vamos a comparar

$$2\langle L(x),L(y)\rangle=\|L(x)\|^2+\|L(y)\|^2-\|L(x)-L(y)\|^2\quad\text{y}\quad 2\langle x,y\rangle=\|x\|^2+\|y\|^2-\|x-y\|^2.$$

Como $\|L(x)\|=\|x\|$ y $\|L(x)-L(y)\|^2=\|x-y\|^2$ deducimos que $\langle L(x),L(y)\rangle=\langle x,y\rangle$. Para llegar a $L(x+y)=L(x)+L(y)$ basta que sea $\|L(x+y)-L(x)-L(y)\|^2=0$. Esto es así porque... y aquí interviene el lector. ♦

Hay diversos tipos de isometrías. La primera clasificación es según el valor de su determinante, que recordamos vale ± 1 (teorema 195). Las isometrías con $\det(L) = 1$ tienen diversos nombres sin que haya unanimidad. Lo más frecuente es llamarlas **rotaciones** pero también pueden ser **isometrías propias, directas** (o incluso, lo más claro pero lo menos extendido) **isometrías positivas**. La terminología es menos unánime en el caso $\det(L) = -1$. Los nombres pueden ser **reflexiones** o **isometrías impropias, inversas o negativas**. Usaremos todas ellas. No obstante queremos evitar referirnos *en general* a las impropias como "simetrías". Hay dos razones. La primera es que si bien ciertas simetrías (por ejemplo, las que lo son respecto a un hiperplano) son isometrías impropias, también otras tienen determinante 1. La segunda razón es que hay isometrías con determinante -1 que no son simetrías de ningún tipo.

Nuestro propósito es clasificar las isometrías viendo, como en otras partes del curso, que se pueden elegir bases ortonormales donde su matriz es particularmente sencilla e informativa y que estas matrices permiten clasificarlas. Se verá también que se puede conseguir esta matriz sin necesidad de conocer explícitamente la base que permite escribirla.

10.6. Isometrías del plano euclidiano

El caso $n = \dim(\mathbb{E}) = 2$ es muy informativo. Fijamos $n = 2$ salvo aviso contrario, si bien $(\mathbb{E}, \omega)$ es arbitrario y la intuición la da el caso $\left(\mathbb{R}^2, \varepsilon\right)$. Empezamos, porque es más sencillo, con las isometrías inversas; o sea, $\det(L) = -1$. De hecho ya las conocemos: son las simetrías respecto a rectas.

Teorema 206 *Sea L una isometría inversa del plano euclidiano $(\mathbb{E}, \omega)$. Entonces L tiene dos valores propios ± 1 siendo sus espacios propios $\mathbb{E}(1)$ y $\mathbb{E}(-1)$ dos rectas ortogonales y L la simetría respecto a $\mathbb{E}(1)$. Por tanto si u y v son vectores propios con valor propio ± 1 respectivo, tenemos que*

$$L(x) = \frac{2\langle x, u\rangle}{\|u\|^2}u - x = x - \frac{2\langle x, v\rangle}{\|v\|^2}v.$$

Demostración. En cualquier base ortonormal,la matriz a de L es ortogonal con determinante 1. Entonces a, por el problema 333, tiene la forma

$$a = \begin{pmatrix} p & q \\ q & -p \end{pmatrix}, \quad \text{con } p^2 + q^2 = 1. \tag{10.10}$$

Es inmediato que $C_a(X) = X^2 - 1$, luego los valores propios son ± 1. Si comparamos L con S dado cualquiera de las dos fórmulas del enunciado, vemos que L y S coinciden sobre la base (u, v), luego $L = S$ y L es una simetría como se dijo. ♣

Teorema 207 *Sea L una isometría directa (rotación) del plano euclidiano $(\mathbb{E}, \omega)$. Entonces* $\det(L) = 1$ *y en* cualquier *base ortonormal $\mathcal{W} = (w_1, w_2)$, la matriz de L es de la forma*

$$\begin{pmatrix} p & -q \\ q & p \end{pmatrix}, \quad \textit{con } p^2 + q^2 = 1. \tag{10.11}$$

Si $L \neq \mathrm{id}_{\mathbb{E}}$, L solo fija el origen.

Demostración. La matriz de L es ortogonal por ser isometría y con determinante 1 por ser directa. En cualquier base ortonormal de ω, la matriz a de L es, de acuerdo con el problema 333, como se dice en el enunciado. Entonces $C_a(X) = X^2 + 2p + 1$ y no tiene raíces reales salvo si $a = I$, porque si no, $(2p)^2 - 4 < 0$ (recordar que $p^2 + q^2 = 1$.). Si no hay valores propios solo puede fijar el origen. ♣

Es importante observar que si L es una isometría, su matriz a en una base *arbitraria* debe cumplir ciertas condiciones. La clave es que la traza y determinante de cualquier endomorfismo son los mismos sea como sea la matriz y base que se utilice para calcularlos. Si L es impropia, la ecuación (10.10) nos dice que en la base construida, $\mathrm{tr}(a) = 0$ y $\det(a) = -1$ De modo análogo para L propia, (10.11) nos da $\mathrm{tr}(a) = 2p$ y $\det(a) = -1$, luego, además, $q = \pm\sqrt{1 - p^2}$ con incertidumbre respecto al signo de q. (Esta indefinición tiene más trascendencia de lo que parece.)

Problema 406 *Consideramos matrices*

$$a = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}, \quad b = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Se pregunta: (a) ¿Es a matriz de una simetría para ε como producto euclidiano? En caso afirmativo dígase respecto a que recta. (b) ¿Es b matriz de una rotación si ω tiene a Ω como matriz en $\mathcal{E}$? (c) Puede ser c matriz de una rotación o simetría respecto a ε o respecto a ω?

Problema 407 *Sean R_1 y R_2 rotaciones y S_1 y S_2 simetrías de $(\mathbb{E}, \omega)$. Probar que $R_1 \circ R_2$ y $S_1 \circ S_2$ son rotaciones y $R_1 \circ S_2$ y $S_1 \circ R_2$ son simetrías. La inversa de una rotación es una rotación y la de una simetría es una simetría (de hecho, ella misma).*

Teorema 208 *Sean u y v dos vectores unitarios distintos. Hay exactamente dos isometrías R y S que llevan u a v, de las cuales R es una rotación y S es una simetría.*

Demostración. Los vectores $u+v$ y $u-v$ son ortogonales porque

$$\langle u+v, u-v\rangle = \|u\|^2 - \|v\|^2 + \langle u, v\rangle - \langle v, u\rangle = 0.$$

Empezamos con las isometrías. Es muy intuitivo que si $u+v \neq 0$ y $\mathbb{F} = \lg(u+v)$, la simetría S respecto de $\mathbb{F}$ cumplirá. Su fórmula, al ser $u-v$ ortogonal a $\mathbb{F}$, es

$$S(x) = x - \frac{2\langle u-v, x\rangle}{\|u-v\|^2}(u-v).$$

Efectivamente, $S(u) = v$ como muestra el cálculo

$$S(x) = u - \frac{2\langle u-v, u\rangle}{\|u-v\|^2}(u-v) = u - \frac{2\|u\|^2 - 2\langle u, v\rangle}{\|u\|^2 + \|v\|^2 - 2\langle u, v\rangle}(u-v) = v,$$

ya que al sustituir $\|u\| = \|v\| = 1$ la fracción es 1 y $S(u) = u - (u-v) = v$. Incluso si $u+v=0$, la simetría S de más arriba cumple $S(u) = v$. Cualquier simetría T que cumpla $T(u) = v$, cumplirá también $T(v) = u$ porque $T^2 = \mathrm{id}_{\mathbb{E}}$. Con esto, $T(u-v) = -(u-v)$ y la recta ortogonal a $u-v$ tiene que ser la que da la simetría T, por lo que $T = S$ y la simetría es única.

Para la rotación tomamos una base ortonormal $\mathcal{W} = (w_1, w_2)$ tal que $w_1 = u$ y escribimos $R(u) = v = pw_1 + qw_2$. Sea R la rotación que en $\mathcal{W}$ tiene matriz

$$a = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}.$$

Es una rotación pues a es ortogonal y con determinante 1 (se usa $p^2 + q^2 = \|v\|^2 = 1$). El cálculo

$$\begin{pmatrix} p & -q \\ q & p \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}$$

nos da que $R(u) = v$. Cualquier otra rotación T que lleve u a v tendrá en $\mathcal{W}$ matriz b con $a_1 = b_1 = (p, q)^\top$. Las condiciones $\varepsilon(b_1, b_2) = 0$ y $\varepsilon(b_2, b_2) = 1$ nos dan enseguida $a = b$ y $R = T$. ♣

10.7. Ángulos y orientaciones en el plano

Quizás se extrañe el lector que a estas alturas no hayamos definido el ángulo de una rotación y probado afirmaciones como que el ángulo de la composición de rotaciones es la suma de los ángulos. El concepto de ángulo es muy útil sobre todo si todas las rotaciones tienen el mismo centro y nosotros, que tratamos solo rotaciones lineales, tenemos 0 como centro. Es una simplificación importante el asignar a las rotaciones sus ángulos. Sin embargo esto que a nivel intuitivo es fácil, es insospechadamente difícil de formalizar. Se ve que para que sea cierto que "el ángulo de la composición de rotaciones es la suma de los ángulos" se necesita que el ángulo sea un número módulo 2π y no un número real. (A fin de cuentas dos rotaciones de las de la vida diaria que se diferencian en un número exacto de vueltas mueven del mismo modo cualquier punto.) Pero aparte hay otra dificultad y es que asignar a la rotación R su ángulo no puede hacerse si no se distinguen previamente las bases orientadas a derechas y las orientadas a izquierdas sin que haya razón para preferir unas u otras. Todo esto es un tanto impreciso, pero precisarlo de modo riguroso no es nada fácil y hay que seguir un camino alambicado. El lector puede ver el teorema 209 un poco más abajo para darse cuenta al menos donde empiezan las dificultades. Las subsecciones

que siguen *Orientación de bases en el plano* y *Ángulo de una rotación* son material optativo y pueden saltarse en una primera lectura.

Si revisamos el teorema 207 vemos que dada la rotación R y una base ortonormal $\mathcal{W}$, la matriz de R en $\mathcal{W}$ es a en (10.11). En principio el par (p,q) que determina a dependerá de R pero también de la base $\mathcal{W}$ elegida. Es muy sorprendente que R determina "casi" unívocamente (p,q) y por tanto a, aunque variemos $\mathcal{W}$ *ortonormal.* El orden de cuantificadores en el teorema siguiente es esencial.

Teorema 209 *Dada la rotación* $R \neq \mathrm{id}$, existen p *y* q *tales que* para toda *base ortonormal* $\mathcal{W}$, *la matriz de* R *en* $\mathcal{W}$ *es de una de las dos formas*

$$a = \begin{pmatrix} p & -q \\ q & p \end{pmatrix} \text{ o bien } b = \begin{pmatrix} p & q \\ -q & p \end{pmatrix} \quad \text{con } q > 0.$$

De hecho, p *es el valor* constante *de* $\omega(u, R(u))$ *cuando* u *recorre el conjunto de vectores unitarios de* ω *y* $|q| = \sqrt{1-p^2}$. *Así pues,* R *determina el par* (p,q) *salvo el signo de* q.

Demostración. Tomemos *cualquier* $u = \alpha w_1 + \beta w_2$ unitario (luego $\alpha^2 + \beta^2 = 1$). Sea, según el caso, a o b la matriz de R. Al ser $\mathcal{W}$ ortonormal, $R(u)$ tendrá en coordenadas una de las matrices

$$\begin{pmatrix} p & -q \\ q & p \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} p\alpha - q\beta \\ q\alpha + p\beta \end{pmatrix} \text{ o bien } \begin{pmatrix} p & q \\ -q & p \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} p\alpha + q\beta \\ -q\alpha + p\beta \end{pmatrix},$$

y en ambos casos $\omega(u, R(u))$ es p. En efecto, $\alpha^2 + \beta^2 = 1$ y

$$\alpha(p\alpha - q\beta) + \beta(q\alpha + p\beta) = p\left(\alpha^2 + \beta^2\right) = p, \qquad \alpha(p\alpha + q\beta) + \beta(-q\alpha + p\beta) = p\left(\alpha^2 + \beta^2\right) = p.$$

De $p^2 + q^2 = 1$, ya sabido por el teorema 207, obtenemos $q = \pm\sqrt{1-p^2}$ sin que podamos determinar el signo. Al expresarse p de un modo en cuya definición no intervienen bases (el valor constante de $\omega(u, R(u))$), p es el mismo para todas las bases *ortonormales* y otro tanto se puede decir de $|q|$. ♣

Es tentador pensar que con algo más de esfuerzo conseguiremos deshacer la ambigüedad de signo y que R determinará (p,q) unívocamente, pero no es así. Veámoslo. Sea $\mathcal{W}$ la base ortonormal (w_1, w_2) y $\mathcal{W}'$ otra con $w_{1'} = w_1$ y $w_{2'} = -w_2$. Supongamos por ejemplo que R tiene la matriz a del teorema 207. Como la matriz de cambio de base es diagonal con $(1,-1)$ en ella, la matriz a' de R en $\mathcal{W}'$ será

$$a' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{-1}\begin{pmatrix} p & -q \\ q & p \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} p & q \\ -q & p \end{pmatrix} \neq a.$$

Muestra el ejemplo que si en ciertas bases R tiene matriz con $q > 0$, en otras tendrá $q < 0$.

Este tropiezo al no poder definir q a partir de R frustra la idea natural para definir el ángulo de R. ¿Cuál sería? Un teorema de Análisis dice que para cada punto del círculo $\mathbb{S}^1 \subset \mathbb{R}^2$ de ecuación $X^2 + Y^2 = 1$, y (p,q) lo es, existe un número θ tal que $X = \cos\theta$ y $Y = \operatorname{sen}\theta$. Este número θ no es único pero si θ' cumple las mismas ecuaciones, $\theta' - \theta = 2\pi n$ con $n \in \mathbb{Z}$. Abreviamos diciendo que θ y θ' **son iguales módulo** 2π. Dada R, querríamos definir el ángulo, escribiendo en una base ortonormal *cualquiera*

$$\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(R) = a = \begin{pmatrix} p & -q \\ q & p \end{pmatrix} = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}$$

y decir que θ es el ángulo de R. *Esta definición es incorrecta porque depende de la base.* Acabamos de ver un poco más arriba que si en $\mathcal{W}_1$ se tiene que θ_1 verifica que $p = \cos\theta_1$ y $q = \operatorname{sen}\theta_1$, hay bases $\mathcal{W}_2$ en las que el ángulo θ_2 tendría que cumplir $p = \cos\theta_2$ y $-q = \operatorname{sen}\theta_2$ por el cambio de signo de q. Esto exige que $\theta_1 = -\theta_2$ módulo 2π, pero no es lo que nos convendría, que es $\theta_1 = \theta_2$ módulo 2π. No todo está perdido porque sí que es cierto que puede ser $\mathcal{W}_1 \neq \mathcal{W}_2$ y, a la vez, que tenga L la misma matriz en las dos bases. Surge una idea en el horizonte y es que podríamos quedarnos solo con parte de las bases ortonormales y definir el ángulo como más arriba, *pero solo con ellas.* Bien pero ¿con que criterio se separan unas bases de otras? La respuesta está próxima pero el lector deberá decidir si quiere seguir o saltar las dos siguientes secciones.

10.7.1. Orientación de bases del plano

Vamos a estudiar si hay una fórmula para obtener q directamente en función de R, aunque ya vamos prevenidos de que hay dependencia de la base. Si fijada R asignamos a cada $u \in \mathbb{E}$ unitario y base ortonormal $\mathcal{W}$ el número $f(u,\mathcal{W}) = \omega(u, R(u))$, el teorema 207 dice que vale lo mismo sean como sean u unitario y $\mathcal{W}$ ortonormal, con solo dependencia de R. La idea es definir una función análoga $g(u,\mathcal{W})$ que dé el valor q y estudiar cómo depende de la base ortonormal $\mathcal{W}$. Primeramente definimos

$$\Delta_{\mathcal{W}} : \mathbb{E}\times\mathbb{E} \longrightarrow \mathbb{R}, \qquad \Delta_{\mathcal{W}}(x,y) = \det \begin{pmatrix} x^1 & y^1 \\ x^2 & y^2 \end{pmatrix} \tag{10.12}$$

con las coordenadas de x e y en $\mathcal{W}$. Si tenemos otra base $\mathcal{W}'$ se ve fácilmente que

$$\begin{pmatrix} x^{1'} & y^{1'} \\ x^{2'} & y^{2'} \end{pmatrix} = \begin{pmatrix} c_1^{1'} & c_2^{1'} \\ c_1^{2'} & c_2^{2'} \end{pmatrix} \begin{pmatrix} x^1 & y^1 \\ x^2 & y^2 \end{pmatrix}, \qquad \Delta_{\mathcal{W}'}(x,y) = (\det c)\,\Delta_{\mathcal{W}}(x,y). \tag{10.13}$$

siendo c la matriz c de cambio de bases. No hay que precisar mucho; lo que importa es que: **(a)** debe ser $\det(c) = \pm 1$ pues c es matriz de cambio entre bases ortonormales (teorema 171); y **(b)** la matriz c es la matriz de la función $H : \mathbb{E} \to \mathbb{E}$ que lleva $w_{j'}$ en w_j, $j = 1, 2$. Por tanto $\det c = 1$ equivale a que H sea una rotación y $\det c = -1$ equivale a que H sea una simetría. Deducimos que

Teorema 210 *Las funciones $\Delta_{\mathcal{W}}$ y $\Delta_{\mathcal{W}'}$ definidas por (10.12) son iguales si y solo si la función H que lleva una base en la otra es una rotación. Si H es una simetría, entonces $\Delta_{\mathcal{W}} = -\Delta_{\mathcal{W}'}$.*

Motivados por este teorema vamos a dividir el conjunto $\mathbf{O}$ de bases ortonormales de $\mathbb{E}$ en dos clases de acuerdo con una relación de equivalencia llamada **tener la misma orientación.** Lleva cierto tiempo explicarlo. Sabemos por el teorema 194 que si $H : \mathbb{E} \to \mathbb{E}$ está definida por llevar la base $\mathcal{W}$ en $\mathcal{W}'$, más precisamente, $H(w_i) = w_{i'}$, $i = 1, 2$, H es una isometría. Si c es la matriz que expresa las $w_{i'}$ en función de la w_i, esta matriz es precisamente $\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(H)$, así que

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(H) = \mathrm{mat}_{\mathcal{U}'}^{\mathcal{U}}(\mathrm{id}_{\mathbb{E}}) \quad \text{si } H(w_i) = w_{i'},\ i = 1, 2.$$

Se sigue entonces que la isometría H que superpone o transporta $\mathcal{W}$ a $\mathcal{W}'$ es una rotación o simetría si y solo si la matriz de cambio de base tiene respectivamente determinante positivo o negativo. Diremos entonces que dos bases ortonormales $\mathcal{W}$ y $\mathcal{W}'$ del plano euclidiano $(\mathbb{E}, \omega)$ **tienen la misma orientación** si la isometría H que transporta $\mathcal{W}$ a $\mathcal{W}'$ es una rotación y **tienen distinta orientación** (u **orientación opuesta**) si esta isometría es una simetría. Está claro que puede decirse como equivalente que la matriz de cambio de base tiene, respectivamente, signo positivo o negativo. Es muy fácil probar con el problema 407 que la relación, denotada por $\mathcal{W} \sim \mathcal{W}'$, es de equivalencia. Elegir una orientación (quedarse con la mitad de las bases clasificadas por $\sim$) se llama **orientar el plano** $(\mathbb{E}, \omega)$. Si $\mathbf{O}^+$ es la clase elegida, las bases $\mathcal{W} \in \mathbf{O}^+$ se llaman **positivas** o **bien orientadas** y las de la otra clase $\mathbf{O}^-$ son las **negativas** o **mal orientadas**. Normalmente se toma una base ortonormal $\mathcal{W}$ y se define **la orientación $\mathbf{O}_{\mathcal{W}}$ inducida por** $\mathcal{W}$ como la formada por todas las bases ortonormales $\mathcal{U}$ con la misma orientación que $\mathcal{W}$. Si hay una base "distinguida" $\mathcal{W}$, es conveniente elegir $\mathbf{O}_{\mathcal{W}}$ como orientación de $\mathbb{E}$. El ejemplo usual de esto es $\mathbb{R}^2$ que se orienta por defecto con $\mathbf{O}_{\mathcal{E}}$, la que llamaremos **orientación estándar**. Es muy fácil comprobar si $\mathcal{W} = (w_1, w_2)$ entonces $\mathcal{W}$ es positiva con la orientación $\mathbf{O}_{\mathcal{E}}$ si y solo si la matriz $w = (w_1, w_2)$ tiene determinante > 0. En general no hay criterio para que sea preferente cierta orientación en $\mathbb{E}$.

Hemos usado ω para dar una idea asequible de cómo distinguir la orientación de las bases según se superpongan con una rotación o con una simetría. Sin embargo ω es superfluo a la hora de definir orientaciones, si bien advertimos que lo que se gana en generalidad se pierde en intuición. El lector que haya estudiado la exposición amplia del producto vectorial le sonará algo familiar. En *El producto vectorial. Orientaciones* del capítulo precedente definimos en dimensión arbitraria y sin producto euclidiano en $\mathbb{E}$ algo muy semejante a lo hecho en el párrafo anterior. Lo recordamos.

Sea $\mathbb{E}$ un espacio real de dimensión n sin contar con ningún tipo de producto escalar. Dadas bases $\mathcal{U}$ y $\mathcal{V}$ hay matrices de cambio de $\mathcal{U}$ a $\mathcal{V}$ y $\mathcal{V}$ a $\mathcal{U}$, digamos c y d inversas una de otra. Definimos en $\mathbf{B}(\mathbb{E})$, el conjunto de bases de $\mathbb{E}$, una relación $\mathcal{U} \sim \mathcal{V}$ si las matrices de cambio tienen determinante > 0. (Al ser $c = d^{-1}$ solo hay que comprobar $\det(c) > 0$ o $\det(d) > 0$.) Esto es una relación de equivalencia con dos clases que llamaremos **orientaciones de** $\mathbb{E}$. Elegida una de ella como preferente, se denotará por $\mathbf{O}^+$, y a la otra por $\mathbf{O}^-$ siendo llamadas las bases de $\mathbf{O}^+$ (respectivamente de $\mathbf{O}^-$) **bases positivas**

(respectivamente **negativas**). Si fijamos una base $\mathcal{U}$ la **la orientación $\mathbf{O}_\mathcal{U}$ inducida por $\mathcal{U}$** es la formada por todas las bases ortonormales $\mathcal{V}$ tales que $\mathcal{U} \sim \mathcal{V}$. Como se ve, se generalizan las definiciones del plano euclidiano. La ventaja de esta ampliación de la definición es que podemos decir que $\mathcal{U}$ es una base positiva o negativa aunque no sea ortonormal. La ventaja de estudiar esto en el plano euclidiano es que permite visuslizar, al menos en $\left(\mathbb{R}^2, \varepsilon\right)$ el que dos bases tengan la misma orientación, que pasa justamente cuando por rotación puede llevarse una sobre la otra, y si no la tienen, se superponen con una simetría. Hay que señalar que aunque la posibilidad de superposición es la forma de intuir el concepto es el signo del determinante en la matriz de cambio de base lo que permite el cálculo. Lo más importante para lo que sigue es la relectura del teorema 215 en el sentido siguiente: *Si las bases ortonormales $\mathcal{W}$ y $\mathcal{W}'$ tienen la misma orientación, las funciones $\Delta_\mathcal{W}$ y $\Delta_{\mathcal{W}'}$ definidas por* (10.12) *son la misma.*

Con este largo preámbulo ya podemos definir la función $g_\mathcal{W}$ de la que hablamos poco antes de la ecuación (10.12). Es

$$g\left(u, \mathcal{W}\right) = \Delta_\mathcal{W}\left(u, R\left(u\right)\right).$$

A diferencia de $f\left(u, \mathcal{W}\right) = \omega\left(u, R\left(u\right)\right)$, g sí varía al cambiar la base, pero muy poco, apenas en un signo. Calculamos $g\left(u, \mathcal{W}\right)$ para $u = \alpha w_1 + \beta w_2$ con $\alpha^2 + \beta^2 = 1$ (o sea, unitario),

$$\mathrm{mat}^\mathcal{W}\left(R\left(u\right)\right) = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} p\alpha - q\beta \\ q\alpha + p\beta \end{pmatrix}, \quad g\left(u, \mathcal{W}\right) = \begin{vmatrix} \alpha & p\alpha - q\beta \\ \beta & q\alpha + p\beta \end{vmatrix} = q\left(\alpha^2 + \beta^2\right) = q.$$

Si sustituimos $\mathcal{W}$ por $\mathcal{W}'$ con la misma orientación, sabemos por el teorema 210 que $\Delta_\mathcal{W} = \Delta_{\mathcal{W}'}$, luego $g\left(u, \mathcal{W}\right) = g\left(u, \mathcal{W}'\right)$. Si por el contrario $\mathcal{W}'$ tiene la orientación opuesta, $\Delta_\mathcal{W} = -\Delta_{\mathcal{W}'}$ y $g\left(u, \mathcal{W}\right) = -g\left(u, \mathcal{W}'\right)$. Hemos probado un teorema fundamental.

Teorema 211 *Supongamos que tenemos el plano orientado por $\mathbf{O}^+$ y para un u unitario concreto conocemos $R\left(u\right)$, su imagen por la rotación R. La matriz de R en* toda *base ortonormal positiva $\mathcal{W}$ es*

$$a = \mathrm{mat}_\mathcal{W}^\mathcal{W}\left(R\right) = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}, \quad \textit{siendo } p = \omega\left(u, R\left(u\right)\right),\ q = \Delta_\mathcal{W}\left(u, R\left(u\right)\right).$$

Dependen p y $|q|$ tan solo de R y (p, q) tan solo de R y $\mathbf{O}^+$, independientemente de u. Si se trabaja con una base ortonormal negativa hay que sustituir q por $-q$.

Problema 408 *Dar las matrices en la base estándar de las rotaciones R_1, R_2 y R_3 que llevan respectivamente $(3, 4)$ a $(5, 0)$, $(3, 4)$ a $(0, 5)$ y $(3, 4)$ a $(-5, 0)$. Se supone el producto estándar.*

Si nos dan una rotación R en $(\mathbb{E}, \omega)$ con orientación $\mathbf{O}^+$ podemos calcular la matriz a, que es la misma para todas las bases ortonormales $\mathcal{W}$ de $\mathbf{O}^+$, con las fórmulas del teorema 211, pero estando obligados a construir una base ortonormal positiva $\mathcal{W}$ y $\Delta_\mathcal{W}$. Suele suceder que lo que conocemos es la matriz de R en una base $\mathcal{U}$ que sí es positiva, pero no ortonormal. Pues bien, puede sortearse la obligación de calcular explícitamente $\mathcal{W}$ y $\Delta_\mathcal{W}$ para tener el par (p, q). Ya dijimos hace tiempo que $2p = \mathrm{tr}\left(R\right)$ y $|q| = \sqrt{1 - p^2}$, luego solo queda por conocer $\mathrm{sg}\left(q\right)$ a partir de $\mathcal{U}$. Si $\mathcal{W}$ es una base ortonormal positiva como $\mathcal{U}$, las funciones $\Delta_\mathcal{U}$ y $\Delta_\mathcal{W}$ son en principio distintas, siendo de hecho $\Delta_\mathcal{W} = \det\left(c\right)\Delta_\mathcal{U}$ y c matriz de cambio de base. No obstante, dado que $\mathcal{U}$ y $\mathcal{W}$ son positivas, $\det\left(c\right) > 0$, luego los signos, que no los valores, de $\Delta_\mathcal{U}\left(x, y\right)$ y $\Delta_\mathcal{W}\left(x, y\right)$ son iguales, que es lo fundamental. Tomamos $u \neq 0$ y $v = u/\left\|u\right\|$ que será $v = w_1$ primer vector de una base ortonormal positiva $\mathcal{W}$ que no habrá que calcular. Entonces,

$$\mathrm{sg}\left(\Delta_\mathcal{U}\left(u, R\left(u\right)\right)\right) = \mathrm{sg}\left(\Delta_\mathcal{W}\left(u, R\left(u\right)\right)\right) = \mathrm{sg}\left(\left\|u\right\|^2 \Delta_\mathcal{W}\left(w, R\left(v\right)\right)\right) = \mathrm{sg}\left(\Delta_\mathcal{W}\left(v, R\left(v\right)\right)\right) = \mathrm{sg}\left(q\right).$$

Desde luego, $2p = \mathrm{tr}\left(R\right)$, $|q| = \sqrt{1 - p^2}$ y $\mathrm{sg}\left(q\right) = \mathrm{sg}\left(\Delta_\mathcal{U}\left(u, R\left(u\right)\right)\right)$ son fáciles de calcular. El ahorro de trabajo se ve en los ejemplos con productos euclidianos $\omega \neq \varepsilon$.

Problema 409 *En $\mathbb{R}^2$ nos dan $L : \mathbb{R}^2 \to \mathbb{R}^2$ y ω un producto euclidiano con matrices respectivas en $\mathcal{E}$*

$$a = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

1. *¿Por qué $a^\top \Omega a = \Omega$ (que se cumple) prueba que R es una rotación de $(\mathbb{E}, \omega)$?*
2. *¿Cómo es la matriz ℓ de L en una base ortonormal positiva $\mathcal{U}$ para la orientación estándar de $\mathbb{R}^2$?*

10.7.2. Ángulo de una rotación del plano orientado

De momento no hemos hablado del **ángulo de una rotación** pero hemos preparado todo con el trabajo de la sección precedente. Suponemos elegida en $(\mathbb{E},\omega)$ una orientación $\mathbf{O}^+$ y la definición se hará respecto a ella. Veremos que si se cambia a la orientación opuesta $\mathbf{O}^-$ el ángulo "invierte el signo". Si $(\mathbb{E},\omega)=\left(\mathbb{R}^2,\varepsilon\right)$ se elige por defecto la orientación de la base natural $\mathbf{O}_{\mathcal{E}}$.

Tratamos primero los **números módulo** 2π, que se representan por $\mathbb{R}/2\pi\mathbb{Z}$ habitualmente, pero que nosotros representaremos por $\mathbb{R}_{2\pi}$. (Elegimos esta notación atípica porque la definición y operaciones recuerda a las de $\mathbb{Z}_n$, las clases de restos módulo n vistas al principio del curso.) En $\mathbb{R}$ se establece una relación de equivalencia: $x\sim y$ si existe $n\in\mathbb{Z}$ tal que $x-y=2\pi n$ (x e y se diferencian en un múltiplo entero de 2π). Como en cualquier relación de equivalencia, la clase de $x\in\mathbb{R}$, que representaremos por $[x]$, es un subconjunto de $\mathbb{R}$; en concreto,

$$[x]=\{x+2\pi n\mid n\in\mathbb{Z}\}=\{\ldots,x-6\pi,x-4\pi,x-2\pi,x,x+2\pi,x+4\pi,x-6\pi,\ldots\}.$$

El conjunto cociente; o sea, el conjunto de todas las clases, será $\mathbb{R}_{2\pi}$, llamado **conjunto de números módulo** 2π. Por la propia definición, $[x]=[y]$ equivale a que sea $x-y=2\pi n$ con $n\in\mathbb{Z}$ y en particular $[x]=[x+8\pi]$, $[0]=[12\pi]$ o $[\pi]=[-\pi]$. En $\mathbb{R}_{2\pi}$ definimos la suma por $[x]+[y]=[x+y]$. Para ver que la definición es correcta ha de verificarse (es fácil) que si $[x]=[x']$ e $[y]=[y']$ entonces $[x+y]=[x'+y']$ para que la operación no dependa de los representantes elegidos sino solo de las clases. Con la suma $\mathbb{R}_{2\pi}$ es un grupo abeliano, siendo $[0]$ el elemento unidad y $[-x]$ el inverso de la clase $[x]$.

Ya podemos definir el **ángulo de la rotación** R, necesitando tan solo una base positiva $\mathcal{W}$ de la orientación elegida, pero sin que sea obligatorio tomar una concreta. Asignamos a R el par (p,q) que da la primera columna de su matriz en $\mathcal{W}$. Como $p^2+q^2=1$, el punto (p,q) está en el círculo $\mathcal{C}:X^2+Y^2=1$ y sabemos por Análisis que hay un $\theta\in\mathbb{R}$ no único, *pero sí único módulo* 2π, tal que $p=\cos\theta$, $q=\sin\theta$. Podemos definir ahora sin problemas el ángulo de R como la *clase* $[\theta]$ en $\mathbb{R}_{2\pi}$. El ángulo $[\theta]$ se obtiene resolviendo el par de ecuaciones

$$\langle u,R(u)\rangle=p=\cos\theta,\quad \Delta(u,R(u))=q=\sin\theta,$$

donde u, unitario, es de libre elección y Δ es la $\Delta_{\mathcal{W}}$ para una base ortonormal positiva. Una advertencia importante: resolver estas ecuaciones es casi siempre imposible si que quiere una solución exacta, y requiere una calculadora o tablas para una solución aproximada (con un número finito de decimales), interviniendo arcsin y arc cos, el arco seno y arco coseno. Los problemas prácticos necesitan para llegar a un resultado exacto ecuaciones como $p=q=\sqrt{2}/2$ que dan $[\theta]=[\pi/4]$. O cosas parecidas. *Lo mejor es trabajar con el par* $(\cos\theta,\sin\theta)$ *y usar las propiedades de estas funciones pero sin extraer el valor de* θ *mientras no sea absolutamente necesario.* En todo caso, si nos dicen que el ángulo de R es θ sabemos que en *cualquier* base ortonormal positiva de $(\mathbb{E},\omega)$ la matriz de R es

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

Como las funciones cos y sin son periódicas de periodo 2π la matriz será la misma para cualquier elemento de $[\theta]$. Como ejemplos tenemos

$$\begin{pmatrix}\cos\frac{\sqrt2}{2} & -\sin\frac{\sqrt2}{2}\\ \sin\frac{\sqrt2}{2} & \cos\frac{\sqrt2}{2}\end{pmatrix}\quad\text{y}\quad\begin{pmatrix}\cos\left(-\frac{\sqrt2}{2}\right) & -\sin\left(-\frac{\sqrt2}{2}\right)\\ \sin\left(-\frac{\sqrt2}{2}\right) & \cos\left(-\frac{\sqrt2}{2}\right)\end{pmatrix}=\begin{pmatrix}\cos\frac{\sqrt2}{2} & \sin\frac{\sqrt2}{2}\\ -\sin\frac{\sqrt2}{2} & \cos\frac{\sqrt2}{2}\end{pmatrix},$$

que son rotaciones diferentes de ángulos $[\pi/4]$ y $[-\pi/4]$. Tenemos también

$$\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},\quad\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\text{ y }\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$$

de ángulos respectivos $[\pi/2]$, $[-\pi/2]$ y $[\pi]$.

Problema 410 *En* $\mathbb{E}$ *elegimos la orientación que hace positiva la base ortonormal* $\mathcal{W}=(u,v)$. *Tenemos dos rotaciones* R_1 *y* R_2 *que llevan* u *respectivamente en* $z_1=\left(1/\sqrt{2}\right)(u+v)$ *y* $z_2=\left(1/\sqrt{2}\right)(u-v)$. *¿Qué ángulos tienen? ¿Y la que lleva* z_1 *a* z_2*?*

Problema 411 *Probar que si R_1 y R_2 son rotaciones con ángulos $[\theta_1]$ y $[\theta_2]$. entonces $R_1 \circ R_2$ tiene ángulo $[\theta_1 + \theta_2]$. Probar también que si R tiene ángulo $[\theta]$ el ángulo de la rotación inversa R^{-1} es $[-\theta]$. ¿Pueden coincidir en algún caso los ángulos de R y R^{-1}?*

Problema 412 *Sea $\mathcal{W}$ una base ortonormal positiva y vectores $u = (\cos\alpha)\, w_1 + (\sin\alpha)\, w_2$ y $v = (\cos\beta)\, w_1 + (\sin\beta)\, w_2$. Si R es la rotación que lleva u a $v = R(u)$ ¿cuál es su ángulo?* ♦

Solución. Si $[\theta]$ es el ángulo se debe tener

$$\begin{pmatrix} \cos\beta \\ \sin\beta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\alpha - \sin\theta\sin\alpha \\ \cos\theta\sin\alpha + \cos\alpha\sin\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta+\alpha) \\ \sin(\theta+\alpha) \end{pmatrix}$$

y esto nos da que para cierto $n \in \mathbb{Z}$ es $\beta = \theta + \alpha + 2\pi n$; o sea $[\beta - \alpha] = [\theta]$. Si nos hubieran pedido el ángulo para R' que lleva v a u, hubiésemos obtenido $[\alpha - \beta] = -[\beta - \alpha]$. ♦

Representemos dos vectores unitarios de $\mathbb{E}$ en la forma

$$s = (\cos\alpha)\, w_1 + (\sin\alpha)\, w_2, \quad t = (\cos\beta)\, w_1 + (\sin\beta)\, w_2 \tag{10.14}$$

siendo $\mathcal{W}$ una base ortonormal positiva. Sean $\mathbb{F}$ y $\mathbb{G}$ las rectas generadas por s y t y S y T las simetrías respecto a estas rectas. Se sabe por el problema 407 que $T \circ S$ es una rotación R. Se trata de calcular el ángulo de R en función de α y β. Enseguida usaremos que $\langle s, t\rangle = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\beta - \alpha)$. Sabemos que para *cualquier* base ortonormal positiva y vector unitario (vamos a tomar s) el ángulo $[\theta]$ de R es solución de

$$\cos\theta = \langle s, R(s)\rangle = \langle s, T(s)\rangle, \quad \sin\theta = \Delta_{\mathcal{W}}(s, R(s)) = \Delta_{\mathcal{W}}(s, T(s))$$

porque $S(s) = s$. Tenemos que $T(x) = 2\langle x, t\rangle\, t - x$ y $T(s) = 2\langle s, t\rangle\, t - s$. Entonces,

$$\cos\theta = \langle s, T(s)\rangle = \langle s, 2\langle s, t\rangle\, t - s\rangle = 2\langle s, t\rangle^2 - 1 = 2\cos^2(\beta - \alpha) - 1 = \cos(2(\beta - \alpha)).$$

$$\begin{aligned} \sin\theta = \Delta_{\mathcal{W}}(s, 2\langle s, t\rangle\, t - s) &= 2\langle s, t\rangle\, \Delta_{\mathcal{W}}(s, t) = 2\langle s, t\rangle \begin{vmatrix} \cos\alpha & \cos\beta \\ \sin\alpha & \sin\beta \end{vmatrix} \\ &= 2\cos(\beta - \alpha)\sin(\beta - \alpha) = \sin(2(\beta - a)) \end{aligned}$$

Vemos por tanto que θ y $2(\beta - \alpha)$ tienen igual seno y coseno, luego se diferencian en un múltiplo entero de 2π, que es como decir que $[\theta] = [2(\beta - \alpha)]$. Hemos probado entonces

Teorema 212 *Sean s y t como en (10.14) y S y T las simetrías respecto a las rectas que generan. La rotación $R = T \circ S$ tiene ángulo $[\theta] = [2(\beta - \alpha)]$.*

Este teorema se suele enunciar afirmando que al componer dos simetrías respecto a rectas $\mathbb{F}$ y $\mathbb{G}$ el ángulo de $R = T \circ S$ es el doble del ángulo que forman las rectas. Aparte de que hay que definir el ángulo de dos rectas (hay que estudiar qué generadores se toman) está el hecho de que si se ve un ángulo como la longitud de un arco del círculo de centro el origen y radio 1, cuesta llevar el control de la demostración si no se hace un dibujo. Sea como sea, el teorema se intuye enseguida como cierto. No hay nada contra hacer dibujos, salvo que por su propia concreción siempre tratan un caso particular. El lector puede intentar demostrar el teorema de modo informal, porque es instructivo. Hay que decir en favor del teorema, tal como lo presentamos, que no hay ambigüedad alguna porque si los generadores s y t de las rectas se escriben como en (10.14) el ángulo es $[\theta] = [2(\beta - \alpha)]$. Pero no hay cambio al elegir $s' = -s$ y $t' = t$ pues sería $[\alpha'] = [\alpha + \pi]$ y $[\beta'] = [\beta]$ y entonces

$$[\theta'] = [2(\beta' - \alpha')] = [2(\beta - \alpha - \pi)] = [2(\beta - \alpha) - 2\pi] = [2(\beta - \alpha)] = [\theta].$$

Las rectas no tienen asignado de modo unívoco un "ángulo de los generadores" pero sí hay de modo unívoco un ángulo de R.

10.8. Teorema de estructura de isometrías

Teorema 213 (estructura de isometrías) *Sea $L : \mathbb{E} \to \mathbb{E}$ una isometría de un espacio euclidiano. Existe una base ortonormal $\mathcal{W}$ en donde la matriz de L es de la forma cuasi-diagonal*

$$a = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = \begin{pmatrix} T_1 & & & & & & \\ & \ddots & & & & & \\ & & T_s & & & & \\ & & & R_1(p_1,q_1) & & & \\ & & & & \ddots & & \\ & & & & & R_k(p_t,q_t) & \end{pmatrix},$$

siendo las matrices $T \in \mathbb{R}^{1\times 1}$ con $+1$ o -1 como único coeficiente y las matrices $R \in \mathbb{R}^{2\times 2}$ de la forma

$$R_j(p_j,q_j) = \begin{pmatrix} p_j & -q_j \\ q_j & p_j \end{pmatrix}, \qquad p_j^2 + q_j^2 = 1, \qquad q_j > 0.$$

El número de matrices T con $+1$, el número de matrices T con -1, el número de matrices R_j, y los números p_j, q_j que aparecen en R_j, están unívocamente determinados por el polinomio característico $C(X)$ de L y por tanto, excepto en el orden en que las cajas 1×1 y 2×2 aparezcan en ella, la matriz de L con esta estructura es única.

La estructura de las cajas T_i y R_j se reconstruye a partir del polinomio característico, porque si se le factoriza como

$$C(X) = \pm (X-1)^A (X+1)^B \left[(X-p_1)^2 + (q_1)^2\right]^{C_1} \cdots \left[(X-p_k)^2 + (q_k)^2\right]^{C_k}, \qquad q_1,\ldots,q_k > 0.$$

hay A cajas 1×1 con 1, B cajas 1×1 con -1 y C_j cajas $R_j(p_j,q_j)$ para cada raíz $p_j + iq_j$ de $C(X)$ con $q_j > 0$.

Demostración. Si existe la base $\mathcal{W}$ donde $a = \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$ tiene la forma del enunciado, se calcula $C(X)$ por cajas muy fácilmente, y se ve que tiene la forma dicha. Como a viene determinada por el número de cajas $T_i = (1)$ que es el exponente de $(X-1)$, el de cajas $T_j = (-1)$ que es el exponente de $(X+1)$ y cada caja R_k contribuye con un factor $(X-p_k)^2 + (q_k)^2$ a $C(X)$, queda claro que L, que determina $C(X)$, determina también a de modo único (salvo permutaciones de cajas). *Esto permite conocer a sin tener que calcular $\mathcal{W}$*, a través de $C(X)$.

Probamos la existencia de $\mathcal{W}$. Veamos primero que existen subespacios $\mathbb{F}$ de dimensión 1 o 2 estables por L (recordamos que eso significa que $L(\mathbb{F}) \subset \mathbb{F}$ pero puede ser $L(x) \neq x$ para ciertos $x \in \mathbb{F}$). Si hay vectores propios w basta tomar $\mathbb{F} = \lg(w)$. Si no hay vectores propios, consideramos $S = L + L^*$, que es autoadjunta porque $S^* = L^* + L^{**} = L^* + L = S$. Necesariamente S tiene un vector propio w (teorema 196) con el que definimos $\mathbb{F} = \lg(w, L(w))$. Este $\mathbb{F}$ es bidimensional porque si no, w sería autovector de L. Afirmamos que $\mathbb{F}$ es estable. Para demostrarlo es suficiente que sea $L(L(w)) \in \mathbb{F}$. Esto es así ya que $S(w) = \lambda w$ da $\lambda w = L(w) + L^*(w) = L(w) + L^{-1}(w)$ y, aplicando L, $\lambda L(w) - w = L^2(w)$.

Necesitamos otro resultado: que si $\mathbb{F}$ es estable por la isometría L también lo es $\mathbb{F}^\perp$. Vimos en el teorema 193 que $L(\mathbb{F}) \subset \mathbb{F}$ implica $L^*\left(\mathbb{F}^\perp\right) \subset \mathbb{F}^\perp$. Dado que L es un isomorfismo, por razones de dimensión, los contenidos son igualdades, y $L(\mathbb{F}) = \mathbb{F}$ implica $L^*\left(\mathbb{F}^\perp\right) = \mathbb{F}^\perp$. Pero $L^* = L^{-1}$ y $L^*\left(\mathbb{F}^\perp\right) = L^{-1}\left(\mathbb{F}^\perp\right) = \mathbb{F}^\perp$ dan, aplicando L, que $\mathbb{F}^\perp = L\left(\mathbb{F}^\perp\right)$.

Con estos preliminares demostraremos la existencia en el teorema por inducción sobre $n = \dim \mathbb{E}$. Es cierto el teorema para $n = 1, 2$. En el caso $n = 1$ es obvio, y en el caso $n = 2$ resulta de los teoremas 206 o 207. Si $m \geq 3$ se toma una subespacio estable $\mathbb{F}$ de dimensión 1 o 2, que acabamos de ver que existe, y descomponemos $\mathbb{E} = \mathbb{F} \oplus \mathbb{F}^\perp$. Al ser $L(\mathbb{F}) = \mathbb{F}$ y $L\left(\mathbb{F}^\perp\right) = \mathbb{F}^\perp$, aplicamos la hipótesis de inducción a cada una de las restricciones y obtenemos la existencia de $\mathcal{W}$ uniendo dos bases de $\mathbb{F}$ y $\mathbb{F}^\perp$. ♣

Problema 413 *Probar con el teorema 213 que si $C(X)$ es el polinomio característico de una isometría L, todas sus raíces, sean reales o complejas, tienen módulo 1. Si $\dim(\mathbb{E})$ es impar hay al menos un vector w tal que $L(w) = w$ o $L(w) = -w$.*

Problema 414 *En $\mathbb{R}^3$ se considera el producto euclidiano ω y el endomorfismo L que en la base estándar $\mathcal{E}$ tienen respectivamente matrices*

$$\Omega = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad \ell = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

Demostrar que ω es un producto escalar y que con él, L es una isometría. Determinar una matriz a y una base $\mathcal{W}$ como en el teorema 213. ♦

Solución. Para ver que ω es un producto escalar se aplica el criterio de Sylvester (teorema 188), y lo es porque los determinantes de las submatrices principales son $2, 4$ y 4, los tres positivos. Para que L sea una isometría debe ser $\ell^\top \Omega \ell = \Omega$, cosa fácil de comprobar.

El polinomio característico de L es $-(1-X)^2(1+X)$ y el teorema de clasificación nos da

$$a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

y L es una simetría respecto a un plano. Véase que si no nos piden la base $\mathcal{W}$, hay poco trabajo si podemos factorizar $C(X)$, cosa fácil en los problemas preparados pero dificilísimo en general.[7]

Claramente L tiene a $\mathbb{F} = \lg(e_1, e_2)$ como subespacio propio, y por tanto estable. Es de hecho el subespacio propio del valor propio $\lambda = 1$. Necesitamos una base ortonormal de $\mathbb{F}$ y, como $\omega \neq \varepsilon$ podrís suceder que (e_1, e_2) no lo fuera. En realidad Ω nos da que es ortogonal con $\|e_1\|^2 = \|e_2\|^2 = 2$. Los vectores $w_1 = e_1/\sqrt{2}$ y $w_2 = e_2/\sqrt{2}$ formarán parte de $\mathcal{W}$. Como w_3 se necesita un vector unitario y ortogonal a w_1 y w_2 pero todo ello para ω (no para ε). Hay que resolver $e_1^\top \Omega x = e_2^\top \Omega x = 0$ que es

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^\top \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2x + z = 0, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}^\top \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2y + z = 0.$$

Esto da $\lg(-1, -1, 2)^\top$ como espacio de soluciones. Calculamos

$$\left\| \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\|^2 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}^T \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} = 4$$

y $w_3 = (-1, -1, 2)^\top / 2$. Se podía también haber calculado v_3 vector propio de $\lambda_2 = -1$ y haber definido $w_3 = v_3 / \|v_3\|$. ♦

Vamos a proceder a clasificar las isometrías lineales en dimensión 2 y 3. Hay que hacer unos comentarios muy generales sobre qué significa "clasificar". Una clasificación de los miembros de un conjunto X es una partición o división de X en conjuntos disjuntos A_i, que llamaremos **clases**, cuya unión sea todo X. Si algún $A_i \cap A_j \neq \emptyset$ o la unión no es todo X, no hay clasificación. Suponiendo que la clasificación sea correcta, lo que la hace *útil*, es disponer de funciones $f : X \to Y$ cumpliendo la propiedad de que si x y x' están en una misma clase, entonces $f(x) = f(x')$, lo que se suele usar para deducir de $f(x) \neq f(x')$ que x y x' están en clases distintas. Hay que tener cuidado pues quizás suceda que sea $f(x) = f(x')$ estando x y x' en clases distintas. A estas funciones f se las llama **invariantes**, y hacen honor a su nombre, pues, por definición, valen lo mismo dentro de cada clase. Una colección de invariantes $f_1, \ldots, f_p$ es una **colección completa de invariantes** si $f_1(x) = f_1(x'), \ldots, f_p(x) = f_p(x')$ implica que x y x' están en la misma clase. Entones, la coincidencia de los valores de todas las f es necesario pero también suficiente para que x y x' estén en la misma clase.

Damos un ejemplo intuitivo. Definimos como X el conjunto de círculos del plano y diremos que x y x' (que en este caso son círculos) están en la misma clase, si hay una traslación del plano que lleva uno en otro. La función $f : X \to \mathbb{R}$ que asigna a cada círculo su radio, es un invariante y ella sola es un invariante completo pues si dos círculos tienen el mismo radio, la traslación que lleva el primer centro en el segundo, lleva el primer círculo al segundo. Otro ejemplo más algebraico es tomar como X el conjunto de las funciones bilineales simétricas en $\mathbb{R}^n$. Dos formas σ y σ' estarán en la misma clase (por definición)

[7] Si $n = 3$ hay un poco más adelante agradables sorpresas.

si hay bases $\mathcal{U}$ y $\mathcal{U}'$ en donde σ y σ' tienen la misma matriz diagonal. Si f es la función que asigna a σ la signatura (p,q), se tiene que f es un invariante completo. En este caso $Y = \mathbb{N} \times \mathbb{N}$ y $f(\sigma) = (p,q)$.

Vamos a clasificar las isometrías lineales diciendo que L y L' estarán por definición en la misma clase si hay bases ortonormales $\mathcal{W}$ y $\mathcal{W}'$ en donde tienen L y L' la misma matriz del teorema 213. La idea es que si L y L' están en la misma clase son "el mismo tipo de transformación una vez elegidas coordenadas adecuadas para estudiarlos". Si se define $f(L) = \det(L)$ queda claro que f es un invariante pero no es completo. Por ejemplo, si $n = 2$, todas las rotaciones tienen determinante 1, pero p y q pueden variar al variar el polinomio característico de L.

Problema 415 *Demostrar que dos isometrías lineales están en la misma clase si y solo si tienen el mismo polinomio característico. ¿Es la traza un invariante? En $n = 2$ ¿puede ser una isometría cuya matriz tenga traza* 1 *la simetría respecto a una recta (hay que contar con la posibilidad de que el producto ω sea "raro")?*

Solución. El teorema 213 dice que con $C(X)$ se reconstruye la matriz, luego si los polinomios son iguales, también las matrices; aunque, eso sí, para dos bases diferentes.

La traza de un endomorfismo L (isometría o no) es la de su matriz en una base cualquiera. Si L y L' tienen la misma matriz del tipo del teorema 213; o sea, están en la misma clase, necesariamente tienen la misma traza pues la calcularemos con esa base y matriz. Así pues, $f(L) = \operatorname{tr}(L)$ es un invariante. Una simetría para $n = 2$, incluso en su versión más simple de que venga dada por una descomposición de $\mathbb{E}$ como suma directa de dos rectas, tiene una matriz diagonal con 1 y -1 en ella, luego su traza es cero. Si $\operatorname{tr}(L) \neq 0$, L no puede ser una simetría. ♠

10.8.1. Clasificación de las isometrías en dimensión dos

Casi todo el trabajo está ya hecho. Lo más rápido es examinar el polinomio característico $C(X)$. Si solo tiene raíces reales hay tres posibilidades nada más, $C(X) = (1-X)^2$, $C(X) = (1+X)^2$ y $C(X) = (1-X)(1+X)$ porque (problema 413 o teorema 196) los valores propios tienen módulo 1. Los dos primeros casos corresponden a $L = \pm\mathrm{id}$, que son rotaciones, y el tercero a una simetría respecto a la recta formada por el espacio propio de $\lambda = 1$. Si $C(X)$ no tiene raíces reales, se puede factorizar $C(X) = (X-p)^2 + q^2$ eligiendo $q > 0$ y siendo $p \pm iq$ las raíces complejas de $C(X)$. Aunque con $n = 2$ cuesta poco calcular $C(X)$, cuesta aún menos calcular $\operatorname{tr}(L)$. Será cero justamente si L es una simetría y $\operatorname{tr}(L) \neq 0$ si es una rotación. Obsérvese que como la traza no depende de la base y como en la base del teorema 213 es $\operatorname{tr}(L) = 2p$, tenemos la matriz a resolviendo $\operatorname{tr}(L) = 2p$ y $q = \sqrt{1-p^2}$. Esto es particularmente útil si no se trabaja con el producto estándar.

10.8.2. Clasificación de las isometrías en dimensión tres

Pueden clasificarse de varias maneras las isometrías $L : \mathbb{E} \to \mathbb{E}$.

1. Según $C_L(X)$, que es la del teorema 213 y sabemos que $C_L(X)$, que se conoce con sus raíces contando su multiplicidad, es un sistema completo de invariantes.
2. La segunda utiliza el polinomio característico, pero es menos fina. Atiende a las multiplicidades de las raíces ± 1 y si hay complejas no reales. Con esta clasificación, aparte de la identidad, tenemos tres simetrías (respecto a punto, eje o plano), rotaciones, y composiciones de una rotación distinta de $\mathrm{id}_{\mathbb{E}}$ y una simetría respecto a un plano ortogonal al eje de rotación.
3. La tercera se concentra en el subespacio $\mathbb{F} = \{x \mid L(x) = x\}$ (o sea, el subespacio de puntos fijos) y la clasificación es de acuerdo con $\dim(\mathbb{F})$. En este caso aparecen los casos anteriores pero se llega a ellos por la vía geométrica de examinar los puntos fijos de L en vez de por la algebraica de examinar $C_L(X)$.

Estudiamos la clasificación **2**. Hay seis posibles tipos de matrices que aparecen en el teorema 213,

$$a_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad a_{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad a_{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

$$a_{(4)}=\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{pmatrix},\qquad a_{(5)}=\begin{pmatrix}1&0&0\\0&p&-q\\0&q&p\end{pmatrix},\qquad a_{(6)}=\begin{pmatrix}-1&0&0\\0&p&-q\\0&q&p\end{pmatrix}.$$

y tenemos la tabla (para ahorrar espacio se quita el signo $-$ en $C(X)$)

$C(X)$ con $q>0$	**nombre o descripción**	**tipo de matriz**
$(X-1)^3$	identidad	$a_{(1)}$
$(X-1)^2(X+1)$	simetría respecto a un plano (especular)	$a_{(2)}$
$(X-1)(X+1)^2$	simetría respecto a un eje (axial)	$a_{(3)}$
$(X+1)^3$	simetría respecto a un punto (central)	$a_{(4)}$
$(X-1)\left((X-p)^2+q^2\right)$	rotación respecto a un eje sin ser simetría axial	$a_{(5)}$
$(X+1)\left((X-p)^2+q^2\right)$	composición de rotación y simetría (rotosimetría)	$a_{(6)}$

Las matrices $a_{(2)},a_{(3)}$ y $a_{(4)}$ son simetrías que fijan el plano $\lg(w_1,w_2)$, la recta $\lg(w_1)$ o 0. Es fácil convencerse visualmente de que las simetrías axiales son rotaciones "de 180 grados" (ponemos comillas porque no se ha definido el ángulo de una rotación en dimensión 3). En todo caso, $\det\left(a_{(3)}\right)=1$, luego es una rotación. Las matrices $a_{(5)}$ engloban todas las posibles rotaciones excepto las simetrías axiales y la identidad y por eso decimos que esta clasificación es menos fina que la del teorema 213. Claramente, salvo la identidad, toda rotación tiene una recta de puntos fijos $\lg(w_1)$ en la base del teorema 213, que es el subespacio propio de $\lambda=1$. A esta recta de puntos fijos, definida salvo si $L=\mathrm{id}$, se la llama el **eje de la rotación**. Finalmente, estudiamos que representa $a_{(6)}$ factorizándola como

$$\begin{pmatrix}-1&0&0\\0&p&-q\\0&q&p\end{pmatrix}=\begin{pmatrix}1&0&0\\0&p&-q\\0&q&p\end{pmatrix}\begin{pmatrix}-1&0&0\\0&1&0\\0&0&1\end{pmatrix}=\begin{pmatrix}-1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&p&-q\\0&q&p\end{pmatrix}. \tag{10.15}$$

Hemos factorizado L como composición de una rotación R y una simetría especular S de forma que conmutan pues $L=R\circ S=S\circ L$. Pero hay algo más: la rotación R no puede tener $q=0$ y fija exclusivamente su eje de rotación $\mathbb{D}=\lg(w_1)$ siendo por otra parte S la simetría especular respecto al plano $\mathbb{P}$ ortogonal a $\mathbb{D}$. Este tipo de isometría composición de una rotación que no es id y la simetría respecto al plano ortogonal a su eje se llama **rotosimetría**.[8] Si relajamos condiciones y admitimos $(p,q)=(-1,0)$, R es una simetría axial y L una simetría central.

Hacemos el resumen. Si se clasifica con el tipo de raíces $1,-1$ y compleja no real de $C(X)$, hay seis clases: la identidad; simetrías respecto a un punto, una recta, o un plano; rotaciones que no son la identidad; y composición una rotación $\neq \mathrm{id}_{\mathbb{E}}$ y una simetría especular respecto al plano ortogonal al eje de la rotación. Es atractivo simplificar en "simetrías", "rotaciones" y "rotosimetrías" ampliando un poco la definición de estas funciones, y es por lo que ponemos comillas. Se incluyen en "simetrías" todas las anteriores incluida las axiales y la central; se incluyen en "rotaciones" a la identidad y las axiales; y se incluye en "rotosimetrías" a la simetría central. El precio a pagar por la simplificación es que no hay una verdadera clasificación porque las simetrías axiales y la central están en dos clases.

Hacemos el resumen. Si se clasifica con el tipo de raíces $1,-1$ y compleja no real de $C(X)$, hay seis clases: la identidad; simetrías respecto a un punto, una recta, o un plano; rotaciones que no son la identidad; y composición una rotación $\neq \mathrm{id}_{\mathbb{E}}$ y una simetría especular respecto al plano ortogonal al eje de la rotación. Hay pues simetrías de tres tipos **central, axial y especular**, **rotaciones** (incluida la identidad, pero que se distingue de las otras al no tener eje) y **rotosimetrías.** La terminología puede causar cierta confusión. La simetría axial es de hecho una rotación (diríamos de ángulo π, si se hubiera definido el ángulo de una rotación con $\dim(\mathbb{E})=3$). Las rotosimetrías no son simetrías pero tienen en común con las simetrías especulares o centrales que su determinante es -1. Se puede definir el eje de la rotosimetría el espacio propio con autovalor -1. La rotosimetría lo tiene como subespacio estable pero no lo fija punto a punto.

Vamos ahora a la clasificación según $\dim(\mathbb{F})$.

1. $\dim(\mathbb{F})=3$. Todos los puntos quedan fijos y L es a **identidad**, un caso extremo de rotación.
2. $\dim(\mathbb{F})=2$. Queda fijo un plano $\mathbb{F}$. El polinomio característico tiene la raíz doble 1, luego es divisible por $(X-1)^2$. Con la tabla anterior, $C(X)=(X-1)^2(X+1)$, y L es una **simetría especular**.

[8] El nombre de **rotosimetrías** nos parece adecuado pero complicado. Lo cierto es que casi nadie lo usa.

3. $\dim(\mathbb{F}) = 1$. Queda fija la recta $\mathbb{F}$, que se llamará el **eje de la rotación**. Tendremos $C(X) = (X-1)Q(X)$ sin que $Q(X)$ pueda tener la raíz 1, ya que estaríamos en el caso **1** o **2**. Tiene que ser $Q(X) = (X-p)^2 + q^2$ con $q \geq 0$. El caso $q = 0$ implica $p = 1$ y $C(X) = (X-1)(X+1)^2$ y tenemos la **simetría axial respecto de** $\mathbb{F}$, que es un caso particular de rotación. En general, en la base del teorema 213, L tiene matriz

$$\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & p & -q \\ 0 & q & p \end{pmatrix}, \quad q \geq 0,$$

que es una **rotación** pues el determinante vale $p^2 + q^2 = 1$.

4. $\dim(\mathbb{F}) = 0$. No hay puntos fijos excepto 0. El polinomio $C(X)$ no puede tener la raíz 1 y la tabla de más arriba nos da $C(X) = (X+1)\left((X-p)^2 + q^2\right)$ con $q \geq 0$. El caso $q = 0$ da $C(X) = (X+1)^3$ y tenemos la **simetría central** $L(x) = -x$. Si $q > 0$, en la base del teorema 213 podemos factorizar $L = R \circ S = S \circ R$, siendo

$$\mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & p & -q \\ 0 & q & p \end{pmatrix}, \quad \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(R) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & p & -q \\ 0 & q & p \end{pmatrix}, \quad \mathrm{mat}_{\mathcal{W}}^{\mathcal{W}}(S) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Tenemos pues que L es composición de una rotación $R \neq \mathrm{id}_{\mathbb{E}}$ y una simetría especular S respecto al plano ortogonal al eje de R. Cuando el subespacio propio de L para $\lambda = -1$ es una recta, se le llama el **eje de la rotosimetría**.

Esta clasificación tiene cuatro clases y se podría reducir a tres si solo consideramos isometrías que no son la identidad. La primera clase son las rotaciones que fijan una recta (su eje), la segunda son las simetrías especulares que fijan un plano, y la tercera las rotosimetrías que solo fijan el origen (aunque si solo se dice esto, cuesta adivinar que son composición de rotación y simetría).

Los problemas de puro cálculo con vistas a clasificar una L concreta conocida a través de su matriz a en una base $\mathcal{U}$ (ortonormal o no) son bastante sencillos incluso si nos piden la clasificación más detallada que es la del teorema 213, porque es fácil calcular $C(X)$ *factorizado*. En efecto, aunque es de grado 3, ha de tener como raíz 1 o -1. Si α es esta raíz se divide $C(X)$ por $X - \alpha$ y $C(X) = (X-\alpha)Q(X)$ siendo ahora fácil factorizar $Q(X)$ porque es de segundo grado. Con las raíces de $C(X)$ se tiene enseguida la forma de la matriz como en el teorema 213. Hay que resaltar que como $C(X)$ es el mismo sea cual sea la base y ω, el trabajo se simplifica muchísimo. (Partimos en todo esto de que L es de verdad una isometría y que nos han planteado bien el problema.)

Aunque el cálculo de $C(X)$ factorizado es fácil, se puede simplificar todavía más la vía para llegar a la matriz del teorema 213 porque basta conocer $\det(a)$ y $\mathrm{tr}(a)$ para reconstruir esa matriz.

Teorema 214 *Supongamos para la base $\mathcal{W}$ del teorema 213, L tenga matriz*

$$a = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & -\gamma \\ 0 & \gamma & \beta \end{pmatrix} \quad \textit{siendo } \alpha = \pm 1,\ \ \beta^2 + \gamma^2 = 1 \ \textit{ y } \ \gamma \geq 0.$$

Entonces α, β y γ vienen dadas por

$$\alpha = \det(L), \quad \beta = \frac{\mathrm{tr}(L) - \det(L)}{2}, \quad \gamma = \sqrt{1-\beta^2} \tag{10.16}$$

y se puede reconstruir a a partir de datos de L que no dependen de la base ni del producto euclidiano ω.

Demostración. Traza y determinante de L no dependen de la base que se use para la matriz de L. Con $\mathcal{W}$ es inmediato que $\det(L) = \det(a) = \alpha$ y $\mathrm{tr}(L) = \mathrm{tr}(a) = \alpha + 2\beta$, de donde salen α y β. Solo nos queda γ, pero las condiciones $\beta^2 + \gamma^2 = 1$ y $\gamma \geq 0$ nos dan $\gamma = \sqrt{1-\beta^2}$. ♣

Problema 416 *Rehacer con esto el problema 414 para clasificar L (damos por hecho que L es isometría).*

Problema 417 *En $\mathbb{R}^3$ con el producto euclidiano estándar nos dan*

$$b=\begin{pmatrix}8&-4&1\\1&4&8\\-4&-7&4\end{pmatrix}\ \textit{que cumple}\ bb^\top=\begin{pmatrix}81&0&0\\0&81&0\\0&0&81\end{pmatrix}\ y\ \det(b)=9^3.$$

(Se puede admitir esto sin más para ahorrar cálculos.) Probar que si L tiene en la base estándar matriz $a=\frac{h}{9}b$ con $h=\pm1$, entonces L es una isometría. Clasificarla de las diversas formas explicadas. ♦

Solución. Si $a_h=\frac{h}{9}b_h$ tenemos una matriz ortogomal. Claramente,

$$\det(a)=\left(\frac{h}{9}\right)^3\det(b)=h=\pm1,\quad \operatorname{tr}(a)=\frac{h}{9}\operatorname{tr}(b)=\frac{16}{9}h.$$

Tenemos la tabla con las fórmulas (10.16)

$\det(L)$	$\operatorname{tr}(L)$	β	γ	matriz tipo	Por puntos fijos
1	$\frac{16}{9}$	$\frac{7}{18}$	$\frac{5\sqrt{11}}{18}$	$a_{(5)}$ rotación	rotación
-1	$-\frac{16}{9}$	$-\frac{7}{18}$	$\frac{5\sqrt{11}}{18}$	$a_{(6)}$ rotosimetría	rotosimetría

. ♦

Problema 418 *En $\mathbb{R}^3$ tenemos el producto euclidiano ω cuya matriz Ω en la base estándar es*

$$\Omega=\begin{pmatrix}2&0&1\\0&2&1\\1&1&2\end{pmatrix}.$$

Escribir las matrices de las tres simetrías S_i respecto a los planos ortogonales a e_1,e_2 y e_3 y calcular con ellas la matriz de $L=S_3\circ S_2\circ S_1$ todo con la base $\mathcal{E}$. ¿Qué tipo de isometría es L?

Problema 419 *En $(\mathbb{E},\omega)$ nos dan $\mathcal{U}=(u_1,u_2,u_3)$ base ortonormal y tenemos las tres simetrías S_i respecto a los planos ortogonales a los u_i. Probar "inteligentemente" que $L=S_3\circ S_2\circ S_1=-\operatorname{id}_{\mathbb{E}}$.*

En general, el clasificar una isometría L una vez que se conoce su matriz b en una base $\mathcal{U}$, *aunque no sea ortonormal*, es mucho más fácil de lo que parece, gracias al teorema 214 porque, una vez que nos aseguran que una matriz b es de una isometría L, los parámetros α,β, y γ se calculan con (10.16). No hace falta siquiera el conocer cómo es ω, aunque sí deben asegurarnos que b es matriz de una isometría. Si es por ejemplo $\det(b)=2$ o $\operatorname{tr}(b)=9$, esto no puede suceder. Tiene quizás más interés el dar ejemplos "raros" de $L:\mathbb{R}^3\to\mathbb{R}^3$ que sean isometrías para ω, construyendo $b=\operatorname{mat}_{\mathcal{E}}^{\mathcal{E}}(L)$ y $\Omega=\operatorname{mat}_{\mathcal{E}\mathcal{E}}(\omega)$.

Se toma h invertible y a ortogonal. Se define $\Omega=h^\top h$. La matriz b será $b=h^{-1}ah$. Afirmamos que si $L:\mathbb{R}^n\to\mathbb{R}^n$ es $L(x)=bx$ y ω es el producto euclidiano con matriz $\Omega=h^\top h$ en la base $\mathcal{E}$, se tiene que L es una isometría para ω. En general, en una base arbitraria la matriz p de L y la q de L^* se relacionan por $p^\top\Omega=\Omega q$. La condición $q=p^{-1}$ equivale a $p^\top\Omega=\Omega p^{-1}$ y $p^\top\Omega p=\Omega$. Aplicado a nuestro caso, L es una isometría si b cumple $b^\top\Omega b=\Omega$. ¿Se cumple? Sí, porque

$$b^\top\Omega b=\left(h^{-1}ah\right)^\top\left(h^\top h\right)\left(h^{-1}ah\right)=h^\top a^\top\left(h^{-1}\right)^\top h^\top hh^{-1}ah=h^\top a^\top ah=h^\top h=\Omega.$$

Tenemos también una base $\mathcal{V}=(v_1,\dots,v_n)$ ortonormal para ω donde L tiene a a como matriz. Basta que las v_j sean las columnas de $v=h^{-1}$. En efecto,

$$v^\top\Omega v=\left(h^{-1}\right)^\top h^\top hh^{-1}=I,\quad \operatorname{mat}_{\mathcal{V}}^{\mathcal{V}}(L)=\operatorname{mat}_{\mathcal{E}}^{\mathcal{V}}(\mathrm{id})\operatorname{mat}_{\mathcal{E}}^{\mathcal{E}}(L)\operatorname{mat}_{\mathcal{V}}^{\mathcal{E}}(\mathrm{id})=v^{-1}bv=a.$$

Preparamos un problema. Tomamos

$$h=\begin{pmatrix}1&0&1\\0&1&0\\0&0&1\end{pmatrix},\quad a=\frac{1}{3}\begin{pmatrix}1&2&2\\2&-2&1\\2&1&-2\end{pmatrix},\quad \Omega=\begin{pmatrix}1&0&1\\0&1&0\\1&0&2\end{pmatrix},\quad b=\begin{pmatrix}-\frac{1}{3}&\frac{1}{3}&1\\\frac{2}{3}&-\frac{2}{3}&1\\\frac{2}{3}&\frac{1}{3}&0\end{pmatrix}.$$

Problema 420 *En $\mathbb{R}^3$ nos dan L y ω con matrices b y Ω en $\mathcal{E}$ como acabamos de escribir. Probar que L es una isometría para ω y calcular la matriz y base $\mathcal{W}$ de L como en el teorema 213.* ♦

Solución. Si no sabemos cómo se ha preparado el problema, ha de verificarse que $b^\top \Omega b = \Omega$. Es un cálculo puro y lo damos por hecho. Para la matriz en la base $\mathcal{W}$ se calcula con las fórmula 10.16,

$$\alpha = \det(b) = 1, \quad \operatorname{tr}(b) = -1, \quad \beta = \frac{\operatorname{tr}(b) - \det(b)}{2} = -1, \quad \gamma = \sqrt{1-\beta^2} = 0.$$

Si conocemos cómo se ha preparado el problema, sabemos que $\det(b) = \det(a) = 1$, que es más fácil de calcular. La forma del teorema 213 es

$$\operatorname{mat}_{\mathcal{W}}^{\mathcal{W}}(L) = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & -\gamma \\ 0 & \gamma & \beta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

y L es una simetría axial. La base $\mathcal{W}$ será ortonormal y formada por vectores propios de L. Aquí es donde más hay que trabajar. Lo primero es calcular $C_b(X) = X^3 + X^2 - X - 1 = (X-1)(X+1)^2$, pero nos podemos ahorrar el trabajo porque será el mismo que tenga $\operatorname{mat}_{\mathcal{W}}^{\mathcal{W}}(L)$, que es obviamente $(X-1)(X+1)^2$. Ahora toca calcular una base de vectores propios. Obtenemos

$$\mathbb{E}(1) = \lg \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbb{E}(-1) = \lg \left(\begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \end{pmatrix} \right).$$

Sea $\mathcal{V} = (v_1, v_2, v_3)$ la base que forman estos tres vectores. Quizás llame la atención al lector que $\varepsilon(v_1, v_2) \neq 0$, siendo vectores propios de valores propios diferentes, *pero no hay contradicción pues nuesto producto es* $\omega \neq \varepsilon$. Puede probar el lector si quiere que $\omega(v_1, v_2) = \omega(v_1, v_3) = 0$ (el ordenador lo hace). Por otra parte

$$\omega(v_1, v_2) = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \end{pmatrix} = \frac{1}{4} \neq 0.$$

Habría que ortogonalizar $\mathcal{V} = (v_1, v_2, v_3)$ con el método de Gram-Schmidt (¡ojo!, utilizando ω) y la base que resulte, una vez normalizada, es $\mathcal{W}$. ♦

Problema 421 *Hacer un problema similar al anterior (como está preparado se puede hacer trampa) con una nueva* $\bar{h}$ *donde* $\bar{h}_3^1 = -1$ *y el resto como* h*. Tenemos manteniendo a como antes, pero* $\bar{\Omega} = \bar{h}^\top \bar{h}$,

$$\bar{b} = \begin{pmatrix} 1 & 1 & -1 \\ \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{4}{3} \end{pmatrix}, \quad \bar{\Omega} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 2 \end{pmatrix}$$

¿Seguirá siendo L *una simetría especular? ¿Cambiará* $\mathcal{W}$*?*

10.9. Factorización con simetrías respecto a hiperplanos

En esta sección, salvo aviso contrario las simetrías son respecto a hiperplanos y "vector fijo" significa "vector fijo no nulo". Usaremos los siguientes resultados en $(\mathbb{E}, \omega)$ de dimensión n

1. Dado $w \neq 0$, la simetría respecto al hiperplano $\mathbb{H}$ ortogonal a w es

$$S_w(x) = x - 2\frac{\langle x, w\rangle}{\|w\|^2} w.$$

   Obsérvese que si se sustituye w por kw con $k \neq 0$ se tiene que $S_w = S_{kw}$. Esto es de esperar puesto que la simetría depende de $\mathbb{H}$ y no de cómo se exprese $\mathbb{H}$. Abreviaremos diciendo que S_w es "la simetría del vector w."

2. Si tenemos u, v *distintos* con $\|u\| = \|v\|$, la simetría respecto al plano ortogonal a $u - v$ cumple $S(u) = v$. (Y como $S^2 = \mathrm{id}$, también $S(v) = u$.) Esto se ha hecho al demostrar el teorema 208, que aunque ahí supone $n = 2$, la demostración vale para n arbitraria.

3. Sean $u_1,\ldots,u_p$ vectores independientes y para cada u_j sea S_j la simetría de vector u_j. La composición $L=S_1\circ\cdots\circ S_p$ tiene un subespacio $\mathbb{F}$ de puntos fijos que contiene a $\lg(u_1,\ldots,u_p)^{\perp}$. En efecto, si $x\in\lg(u_1,\ldots,u_p)^{\perp}$ todo $\langle x,u_i\rangle=0$ y $S_i(x)=x$. Obtenemos también que $\dim\mathbb{F}\geq m-p$.

Teorema 215 (Cartan-Dieudonné) *Sea $L\neq \mathrm{id}_{\mathbb{E}}$ una isometría de $\mathbb{E}$ de dimensión n y $\mathbb{F}$ el subespacio de puntos fijos, de dimensión m. Entonces*

1. *Se puede factorizar $L=S_1\circ\cdots\circ S_{n-m}$ con $n-m$ simetrías y $n-m$ es el mínimo posible.*
2. *En esta factorización $L=S_1\circ\cdots\circ S_{n-m}$ se puede elegir libremente el primer o último factor S_1 o S_{n-p} con tal de que esa simetría fije también los puntos de $\mathbb{F}$. (Naturalmente, la elección de S_1 o S_{n-p} se hace inicialmente y condiciona la de las demás S_j.)* [9]

Demostración. Sea u_0 un vector tal que $L(u_0)\neq u_0$. Tomamos una simetría S_1 de vector $L(u_0)-u_0$ que intercambia u_0 con $L(u_0)$. El hiperplano $\mathbb{H}=\lg(L(u_0)-u_0)^{\perp}$ contiene a los puntos fijos de L porque si $L(x)=x$, al ser L isometría,

$$\langle L(u_0)-u_0,x\rangle=\langle L(u_0),x\rangle-\langle u_0,x\rangle=\langle L(u_0),L(x)\rangle-\langle u_0,x\rangle=0.$$

Es obvio que el subespacio $\mathbb{F}_1$ de puntos fijos de $L_1=S_1\circ L$ contiene a $\mathbb{F}=\mathbb{F}_0$ y a $u_0\notin\mathbb{F}_0$ por lo que $\dim(\mathbb{F}_1)\geq\dim(\mathbb{F}_0)+1$. Aplicamos el mismo razonamiento a L_1 con un vector u_1 tal que $L_1(u_1)\neq u_1$, una simetría S_2 intercambiando u_1 y $L_1(u_1)$, etcétera, obteniendo $L_2=S_2\circ L_1=S_2\circ S_1\circ L$ con subespacio de puntos fijos $\mathbb{F}_2\supset\mathbb{F}_1$ y $\dim(\mathbb{F}_2)\geq\dim(\mathbb{F}_1)+1$. Repitiendo el proceso inductivamente obtenemos una sucesión de isometrías $L_0=L,L_1,L_2,\ldots,L_k$, siendo $L_j=S_j\circ S_{j-1}\circ\cdots\circ S_1\circ L$, y cuyos subespacios de puntos fijos $\mathbb{F}_j$ cumplen

$$\mathbb{F}=\mathbb{F}_0\subset\mathbb{F}_1\subset\ldots\subset\mathbb{F}_k,\quad \dim(\mathbb{F}_j)\geq\dim(\mathbb{F}_{j-1})+1,\ j=1,\ldots,k.$$

Habremos tenido que parar cuando $\mathbb{F}_k=\mathbb{E}$, que supone $L_k=\mathrm{id}$ y $L=(S_k\circ S_{k-1}\circ\cdots\circ S_1)^{-1}=S_1\circ S_2\circ\cdots\circ S_k$. Sumando las desigualdades obtenemos $n=\dim(\mathbb{F}_k)\geq\dim(\mathbb{F}_0)+k=m+k$.

Hemos mostrado que L es factorizable con $k\leq n-m$ simetrías. ¿Que sucedería si fuese $k<n-m$? Los vectores $u_1,\ldots,u_k$ de las simetrías cumplirían $\dim\lg(u_1,\ldots,u_k)\leq k<n-m$ y en el preámbulo se dice que L fijaría los vectores de $(\lg(u_1,\ldots,u_q))^{\perp}$. Pero

$$\dim(\lg(u_1,\ldots,u_k))^{\perp}=n-\dim\lg(u_1,\ldots,u_k)\geq n-k>m=\dim(\mathbb{F})$$

y habría puntos fijos fuera de $\mathbb{F}$. Contradicción. Está probado **1**.

Supongamos $L=S_1\circ\cdots\circ S_p$ con un *número mínimo* p de simetrías y que queremos que aparezca S en vez de S_1, debiendo ser S una simetría que fija $\mathbb{F}$. Vamos a probar que $S\circ L=M$ es factorizable con $p-1$ simetrías *como número mínimo*. Sea $M=S_1'\circ\cdots\circ S_q'$ y q el número mínimo. Claramente M fija $\mathbb{F}$, luego su espacio de puntos fijos $\mathbb{G}$ contiene a $\mathbb{F}$. Tenemos ya probado en **1** que $n-p=\dim(\mathbb{F})\leq\dim(\mathbb{G})=n-q$, luego $q\leq p$.Si fuese $q=p$ tendríamos $L=S_1\circ\cdots\circ S_p=S\circ M=S\circ S_1'\circ\cdots\circ S_q'$, pero al ser $q=p$, se tiene la contradicción $\det(L)=(-1)^p=(-1)^{p+1}$. Si $q\leq p-2$ llegamos a otra contradicción porque entonces L se escribe con $q+1<p$ simetrías. Necesariamente, $q=p-1$ y $L=S\circ M=S\circ S_1'\circ\cdots\circ S_{p-1}'$ es factorizable con p simetrías y S en primer lugar.

Si se quiere que S figure en el lugar de S_p, se escribe $L^{-1}=S_p^{-1}\circ\ldots\circ S_1^{-1}=S_p\circ\ldots\circ S_1$, se aplica el párrafo anterior a L^{-1} escribiéndola como $L^{-1}=S\circ S_{p-1}'\circ\ldots\circ S_1'$ e invirtiendo otra vez queda factorizada L con S al final. ♣

La parte más importante del teorema 215, a saber, que L es factorizable con $n-m$ simetrías (sin entrar en si esta factorización es óptima en algún sentido), se puede deducir directamente del teorema 213 (de estructura de isometrías). El esquema, al que le faltan detalles, es el siguiente. Sea a la matriz del teorema con m cajas 1×1 con 1 y p cajas 1×1 con -1. Según sea $k=2q$ o $k=2q+1$ vemos a como

$$\begin{pmatrix} I_m & & & \\ & G_1 & & \\ & & \ddots & \\ & & & G_\ell \end{pmatrix}\ \text{o}\ \begin{pmatrix} I_m & & & & \\ & -1 & & & \\ & & G_1 & & \\ & & & \ddots & \\ & & & & G_\ell \end{pmatrix}.$$

[9] Hay que insistir, porque es un punto sutil, que no se pretende simplemente que S aparezca en la factorización de L, sino que aparezca en la que tiene el mínimo número de factores.

siendo (obsérvese que es $q_j \geq 0$ y no $q_j > 0$)

$$G_j = \begin{pmatrix} p_j & -q_j \\ q_j & p_j \end{pmatrix}, \qquad p_j^2 + q_j^2 = 1, \quad q_j \geq 0.$$

Es una comprobación pesada que a es el producto de las matrices

$$r_j = \begin{pmatrix} I_m & & & \\ & \ddots & & \\ & & G_j & \\ & & & \ddots \end{pmatrix} \text{ (solo hay un } G_j \text{ y } r_j \text{ es "casi" } I_n)$$

en el caso $n - m$ par, y a es el producto de las matrices

$$s_0 = \begin{pmatrix} I_m & & & & \\ & -1 & & & \\ & & I_2 & & \\ & & & \ddots & \\ & & & & I_2 \end{pmatrix}, \; r_j = \begin{pmatrix} I_m & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & G_j & \\ & & & & \ddots \end{pmatrix} \text{ (solo hay un } G_j \text{ y } r_j \text{ es "casi" } I_n).$$

en el caso $n - m$ impar. La matriz s_0 representa una simetría especular respecto al plano ortogonal a w_{m+1} y las r_j son rotaciones, que solo mueven los vectores de un plano. Cada r_j es factorizable como producto de dos simetrías. Si, digamos, $r_j = s'_j s''_j$ tenemos, según $n - m$ sea par o impar,

$$a = r_1 r_2 \cdots r_\ell = s'_1 s''_1 s'_2 s''_2 \cdots s'_\ell s''_\ell \text{ o bien } a = s_0 r_1 r_2 \cdots r_\ell = s_0 s'_1 s''_1 s'_2 s''_2 \cdots s'_\ell s''_\ell.$$

A partir de a es inmediato que el espacio de puntos fijos es $\mathbb{F} = \lg(w_1, \ldots, w_m)$ luego $n - m = 2\ell$ o $2\ell + 1$, y el número de simetrías que intervienen en la factorización construida es precisamente 2ℓ o $2\ell + 1$. Esto completa la demostración, pero no da información sobre mejores factorizaciones.

En los problemas que siguen se pide (entre otras cosas) factorizar una isometría como producto de simetrías. Dado que hemos dado con detalle la demostración del teorema de factorización con isometrías (teorema 215) lo más natural parece ser examinar la demostración y aplicarla al ejemplo numérico concreto que nos den. Sin embargo en dimensiones 2 y 3 se puede trabajar de modo más directo.

1. En dimensión 2, la isometría L, o es una simetría, en cuyo caso ya la tenemos factorizada, o es una rotación. Si elegimos una simetría cualquiera, $S \circ L$ es una simetría pues vimos en el teorema 197 que L es rotación o simetría y $\det(S \circ L) = \det S \det L = (-1)\,1 = -1$. Si *definimos* $S' = S \circ L$, al ser el cuadrado de cualquier simetría la identidad, llegamos a $S \circ S' = L$ y ya tenemos la factorización.

2. En dimensión 3 la isometría L puede ser

   a) Una simetría, y ya la tenemos factorizada. Detectaremos que L lo es porque $\det(L) = -1$ y 1 es valor propio.

   b) Una rotación. Detectaremos que L lo es porque $\det(L) = 1$. Entonces, excluyendo en adelante el caso trivial $L = \mathrm{id}_{\mathbb{E}}$, 1 es valor propio y su espacio propio $\mathbb{D}$ es una recta, el **eje de la rotación**. Sea S una simetría cualquiera respecto a un plano $\mathbb{P}$ que contenga a $\mathbb{D}$. Entonces $S \circ L = S'$ es una simetría que necesariamente fija $\mathbb{D}$. La justificación es que $\det(S \circ L) = -1$ y $S \circ L = S'$ no puede ser una rotosimetría porque fija $\mathbb{D}$ y las rotosimetrías solo fijan 0. Como en el apartado anterior se llega a $S \circ S' = L$ y ya tenemos la factorización.

   c) Una rotosimetría (isometría inversa que no es una simetría). Detectaremos que L lo es porque $\det(L) = -1$ y -1 es valor propio. Sea $\mathbb{D}$ el espacio propio de -1, luego si $z \in \mathbb{D}$, $L(z) = -z$.[10] Se define S como la simetría respecto al plano $\mathbb{P}$ ortogonal a $\mathbb{D}$. Claramente $M = S \circ L$ fija $\mathbb{D}$ y tiene determinante 1, luego es una rotación. Ya sabemos por **(b)** que existen simetrías S' y S'' (respecto a planos conteniendo a $\mathbb{D}$) tales que $M = S \circ L = S' \circ S''$ y llegamos a $L = S \circ S \circ L = S \circ S' \circ S''$.

[10] No está muy extendido pero podemos llamar a $\mathbb{D}$ el **eje de la rotosimetría.**

Obsérvese que se puede elegir con bastante libertad una de las simetrías. En los problemas, para no complicar los cálculos en exceso, se supone que $\mathbb{E}$ es $\mathbb{R}^2$ o $\mathbb{R}^3$ y $\omega = \varepsilon$.

Problema 422 *Dar en $\mathbb{R}^2$ la matriz de la simetría S respecto de la recta $\mathbb{F}$ generada por $(1,2)$. A continuación, para la rotación R de matriz*

$$R = \frac{1}{3}\begin{pmatrix} 1 & -2\sqrt{2} \\ 2\sqrt{2} & 1 \end{pmatrix}.$$

determinar simetrías S' y S'' que $R = S' \circ S = S \circ S''$. ♦

Solución. El vector $w = (-2,1)$ es ortogonal a $\mathbb{F}$ y tiene $\|w\|^2 = 5$, de modo que $S = S_w$ y su matriz son

$$S_w\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \frac{2}{5}\left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\rangle \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 2y - 3x \\ 4x + 3y \end{pmatrix}, \quad \mathrm{mat}_{\mathcal{E}\mathcal{E}}(S) = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}.$$

Ahora calculamos

$$R \circ S = \frac{1}{3}\begin{pmatrix} 1 & -2\sqrt{2} \\ 2\sqrt{2} & 1 \end{pmatrix}\frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix} = \frac{1}{15}\begin{pmatrix} -8\sqrt{2} - 3 & 4 - 6\sqrt{2} \\ 4 - 6\sqrt{2} & 8\sqrt{2} + 3 \end{pmatrix}$$

que es una simetría S', simplemente porque $\det(S') = \det(R)\det(S) = -1$. Multiplicando a la derecha en $R \circ S = S'$ por S queda $R = S' \circ S$. El otro caso queda para el lector. ♦

Problema 423 *Tomar en $\mathbb{R}^2$ la rotación de matriz*

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

Expresarla como producto de dos simetrías una de las cuales sea la simetría respecto al eje OX. ¿Se puede expresar R como producto de tres simetrías distintas? ¿Se puede expresar $S \circ R \circ S$ como producto de dos simetrías?

Problema 424 *En el problema 417 vimos que $L : \mathbb{R}^3 \to \mathbb{R}^3$ dada respectivamente por las matrices*

$$a_{(1)} = \frac{1}{9}\begin{pmatrix} 8 & -4 & 1 \\ 1 & 4 & 8 \\ -4 & -7 & 4 \end{pmatrix}, \qquad a_{(2)} = -\frac{1}{9}\begin{pmatrix} 8 & -4 & 1 \\ 1 & 4 & 8 \\ -4 & -7 & 4 \end{pmatrix}$$

representa una isometría. Se pide su factorización las como producto de simetrías.

Problema 425 *Nos dan en la base estándar de $\mathbb{R}^3$ un endomorfismo L de matriz*

$$a = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & -1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

Probar que es una isometría y dar su matriz según el teorema 213. Factorizar L como $R \circ S$ con R una rotación de eje $\mathbb{D}$ y S una simetría respecto al plano $\mathbb{P}$ ortogonal a $\mathbb{D}$.

Problema 426 *La función $L : \mathbb{R}^3 \to \mathbb{R}^3$ que en la base estándar tiene matriz*

$$a = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix},$$

tiene una matriz parecida, pero no igual, a las matrices del teorema 213. ¿Es una isometría? Si lo es, decir de qué tipo es, y factorizarla como producto de simetrías.

Problema 427 *Suponemos que $S : \mathbb{R}^3 \to \mathbb{R}^3$ es una simetría especular. Probar que $L(x) = -S(x)$ es también una isometría. ¿De qué tipo? ¿Puede factorizarse L como producto de dos simetrías especulares una de las cuales sea S?*

10.10. Rotaciones, rotosimetrías, y producto vectorial

En esta sección $\mathbb{E}$ representará un espacio euclidiano *orientado de dimensión* 3. Usaremos su volumen Δ y el producto vectorial $\times$, recordando que dependen de la elección del producto euclidiano $\omega(x,y) = \langle x,y\rangle$ y la orientación $\mathbf{O}$. Si el lector no conoce estos conceptos generales, puede suponer que trabaja en $\left(\mathbb{R}^3,\varepsilon\right)$ y que $\Delta(x,y,x) = \det(x,y,z)$. Tomamos una isometría L y *vamos a excluir salvo aviso explícito, el caso* $L^2 = \mathrm{id}$. *Quedan pues descartadas las simetrías (axiales o especulares) y la identidad.* Si $d = \det(L)$, sabemos que $d = \pm 1$ distingue las rotaciones de las rotosimetrías. Como $L^2 = \mathrm{id}$ está excluido solo hay un subespacio propio $\mathbb{D}$, que es una recta y es el eje de la rotación o rotosimetría. Para las rotaciones, $\mathbb{D}$ corresponde al valor propio 1 y para las rotosimetrías al valor propio -1.

Teorema 216 *Sea $\mathbb{D}$ el eje de L y $d = \det(L) = \pm 1$, que es su único valor propio real. Dado v unitario y ortogonal a $\mathbb{D}$ tenemos que*

1. *Si θ es el ángulo no orientado entre v y $L(v)$, entonces $\|v \times L(v)\| = \sin\theta$.*
2. *Para $u = (v \times L(v))/\|v \times L(v)\|$, y por tanto, $v \times L(v) = (\sin\theta)\,u$, se tiene la fórmula*
$$L(x) = (\cos\theta)\,x + (d - \cos\theta)\,\langle u,x\rangle\,u + \sin\theta\,(u \times x).$$
3. *En la base $\mathcal{B} = (u, v, u \times v)$, que es positiva, L tiene matriz,*
$$\mathrm{mat}_{\mathcal{B}}^{\mathcal{B}}(L) = \begin{pmatrix} d & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$

Demostración. Como descartamos $L^2 = \mathrm{id}_{\mathbb{E}}$, existe el eje $\mathbb{D}$ de L y no puede ser $L(v)$ proporcional a v, así que $\|v \times L(v)\| \neq 0$. Ahora calculamos, teniendo en cuenta que $\|v\| = \|L(v)\| = 1$,

$$\|v \times L(v)\|^2 = \|v\|^2\,\|L(v)\|^2 - \langle v, L(v)\rangle^2 = 1 - \cos^2\theta = \sin^2\theta.$$

El ángulo no orientado está en $[0,\pi]$ y ahí el seno es ≥ 0, luego resulta $\|v \times L(v)\| = \sin\theta$.

Sea $M(x) = (\cos\theta)\,x + (d - \cos\theta)\,\langle u,x\rangle\,u + \sin\theta\,(u \times x)$. Probaremos que $L = M$. La restricción a $\mathbb{D}^\perp$ es $M(x) = (\cos\theta)\,x + \sin\theta\,(u \times x)$. Esta restricción es una rotación (no consecuencia obvia de la definición) porque en la base ortonormal $(v, u \times v)$ de $\mathbb{D}^\perp$,

$$\begin{cases} M(v) = (\cos\theta)\,v + (\sin\theta)\,(u \times v) \\ M(u \times v) = (\cos\theta)\,(u \times v) + (\sin\theta)\,(u \times (u \times v)) = (\cos\theta)\,(u \times v) - (\sin\theta)\,v \end{cases}$$

y por tanto, la matriz de $M|_{\mathbb{D}^\perp}$, M restringida a $\mathbb{D}^\perp$, es

$$a = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Al ser $aa^\top = a^\top a = I$ se cumple lo afirmado: $M|_{\mathbb{D}^\perp}$ es una rotación. Como $L|_{\mathbb{D}^\perp}$ también lo es, para que sea $M|_{\mathbb{D}^\perp} = L|_{\mathbb{D}^\perp}$ es suficiente que se cumpla $L(v) = M(v)$. Esto es así porque

$$\begin{aligned} L(v) &= \langle L(v), v\rangle\,v + \langle L(v), u \times v\rangle\,(u \times v) = (\cos\theta)\,v + \Delta(u,v,L(v))\,(u \times v) \\ &= (\cos\theta)\,v + \langle u, v \times L(v)\rangle\,(u \times v) = (\cos\theta)\,v + (\sin\theta)\,\|u\|^2\,(u \times v) \\ &= (\cos\theta)\,v + (\sin\theta)\,(u \times v). \end{aligned}$$

Es inmediato que $L(u) = M(u) = du$, luego $L = M$ como queríamos probar.

En el transcurso de la demostración se ha mostrado también que

$$L(v) = M(v) = (\cos\theta)\,v + (\sin\theta)\,(u \times v), \quad L(u \times v) = M(u \times v) = (\cos\theta)\,(u \times v) - (\sin\theta)\,v,$$

y esto implica que la matriz de L en $\mathcal{B} = (u, v, u \times v)$ es la dicha. ♣

Podría pensarse que $u = (v \times L(v))/\|v \times L(v)\|$ y el ángulo $\theta \in [0,\pi]$ pudieran depender de la elección de $v \in \mathbb{D}^\perp$, pero no es así. Fijemos $v \in \mathbb{D}^\perp$ unitario, lo que da θ (quizás dependiente de v) pero permite usar la fórmula de L. Entonces, para cualquier $z \in \mathbb{D}^\perp$, incluso no unitario,

$$\langle z, L(z)\rangle = \langle z, (\cos\theta)\,z + \sin\theta\,(u \times z)\rangle = \|z\|^2\cos\theta,$$

$$\Delta(u,z,L(z)) = (u,z,(\cos\theta)\,z+\sin\theta\,(u\times z)) = \sin\theta\Delta(u,z,u\times z) = \sin\theta\,(u\times z,u\times z)$$
$$= \sin\theta\left(\|u\|^2\|z\|^2 - \langle z,u\rangle^2\right) = \sin\theta\,(\|z\|\cdot 1-0) = \|z\|^2\sin\theta.$$

Dicho con otras palabras: las funciones

$$f,g:\mathbb{D}^\perp-\{0\}\to\mathbb{R},\quad f(z)=\frac{\langle z,L(z)\rangle}{\|z\|^2},\quad g(z)=\frac{\Delta(u,z,L(z))}{\|z\|^2}$$

toman valor constante; precisamente $\cos\theta$ y $\sin\theta$. De este modo se ve que θ no depende de v. El vector u, en principio $u=u_v$, también es único. Es en efecto un vector unitario en $\mathbb{D}$ y solo hay dos posibilidades y de ellas solo una da que $\mathbb{D}^\perp-\{0\}\to\mathbb{R}$ es positiva. Por tanto u, a pesar de la aparente dependencia de v unitario en $\mathbb{D}^\perp$, es único.

Para una isometría L con $L^2\neq \mathrm{id}$ se define el **ángulo de la isometría** como $\theta\in(0,\pi)$ tal como se ha construido, y el **vector de la isometría** como $(\sin\theta)\,u = v\times L(v)$. Todo esto depende de la elección de orientación, que con ω determina $\times$. Si tomamos la orientación opuesta $\mathbf{O}^-$, se cambia $\times$ por $\times' = -\times$ y éstos ángulos y vectores cambian de signo. Llama la atención que si nien para $n=2$ el ángulo se definía como un número módulo 2π, para $n=3$ es un número de los usuales en $(0,\pi)$, que se puede definir no solo para rotaciones sino también para rotosimetrías, pero no para todas las rotaciones porque hemos excluido las simetrías axiales. Aun así, podemos asignar a una simetría axial L el ángulo π y todo encaja bien porque $\sin\pi=0$ y para u unitario en el eje (hay dos posibilidades, pero no afecta a la fórmula)

$$L(x) = (\cos\pi)\,x + (1-\cos\pi)\,\langle u,x\rangle\,u = -x+2\,\langle u,x\rangle\,u,$$

la antigua conocida fórmula de la simetría axial.

El cálculo de la matriz de **3** se puede hacer sin conocer la base $\mathcal{B}$ ya que

$$d=\det(L),\quad 2\cos\theta+d=\operatorname{tr}(L),\quad \sin\theta=\sqrt{1-\cos^2\theta}. \tag{10.17}$$

(se usa que $\sin\theta\geq 0$ porque $\theta\in(0,\pi)$).

Problema 428 *Aplicar la teoría anterior a $L:\mathbb{R}^3\to\mathbb{R}^3$ con el producto euclidiano y orientación estándar, que tiene matriz*

$$\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(L) = a = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix}.\ \blacklozenge$$

Solución. Con las fórmulas (10.17) todo, salvo la base, es casi inmediato porque

$$d=1,\quad \operatorname{tr}(L)=0,\quad \cos\theta=-\frac{1}{2},\quad \sin\theta=\frac{\sqrt{3}}{2},\quad \mathrm{mat}_{\mathcal{B}}^{\mathcal{B}}(L)=\begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix},$$

y L es una rotación de ángulo $\frac{2\pi}{3}$. Para la base $\mathcal{B}=(u,v,u\times v)$ calculamos primero que el subespacio propio del valor propio 1 está generado por $\left(\sqrt{2},0,1\right)$. El vector $\left(-1,0,\sqrt{2}\right)$ es ortogonal a él y

$$v=\begin{pmatrix} -\frac{1}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix},\quad L(v)=\begin{pmatrix} \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{2}\sqrt{3}}{6} \end{pmatrix},\quad v\times L(v)=\begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix},\quad \|v\times L(v)\|=\frac{\sqrt{3}}{2}.$$

Esto último ya lo sabíamos porque $\|v\times L(v)\|=\sin\theta$. Seguimos calculando

$$u=\frac{v\times L(v)}{\|v\times L(v)\|}=\begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{3}}{3} \end{pmatrix},\quad u\times v=\begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{3}}{3} \end{pmatrix}\times\begin{pmatrix} -\frac{1}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}=\begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix},$$

y la base $\mathcal{B}$ es

$$\mathcal{B}=(v,u\times v,u)=\left(\begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{3}}{3} \end{pmatrix},\begin{pmatrix} -\frac{1}{\sqrt{3}} \\ 0 \\ \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix},\begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}\right).$$

Por si flaquea la fe, el ordenador calcula a partir $\mathrm{mat}_{\mathcal{B}}^{\mathcal{E}}(\mathrm{id})$ con columnas los vectores de $\mathcal{B}$, que

$$\begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & -1 \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{\sqrt{3}} & 0 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & -1 \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{\sqrt{3}} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \blacklozenge$$

Si el lector estudia la sección sobre transformaciones antisimétricas en dimensión 3 con el producto vectorial[11] podrá introducir el **vector axial de una isometría** L. Se utiliza que $A = \frac{1}{2}(L - L^*)$ es claramente antisimétrica y entonces hay un único vector $h \in \mathbb{E}$ tal que $A(x) = h \times x$. En el caso $A = 0$, que equivale a $L^2 = \mathrm{id}_{\mathbb{E}}$, se tiene $h = 0$ siendo en los otros casos el vector axial no nulo. Probamos que este h así definido es el vector $(\sin\theta)\, u = v \times L(v)$ construido con el teorema 216. En efecto, con este teorema sabemos que en $\mathcal{B} = (u, v, u \times v)$,

$$\mathrm{mat}_{\mathcal{B}}^{B}(A) = \frac{1}{2}\begin{pmatrix} d & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} - \frac{1}{2}\begin{pmatrix} d & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}^{\top} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\sin\theta \\ 0 & \sin\theta & 0 \end{pmatrix}.$$

Con esta matriz, el vector $h = (\sin\theta)\, u$ verifica

$$A(u) = 0 = h \times u, \quad A(v) = h \times v, \quad A(u \times v) = (-\sin\theta)\, v = h \times (u \times v)$$

siendo lo último cierto porque

$$h \times (u \times v) = (\sin\theta)\, u \times (u \times v) = (\sin\theta)\left(\langle u, v\rangle\, u - \|u\|^2 v\right) = (-\sin\theta)\, v.$$

Se ha probado el teorema

Teorema 217 *El vector axial h, definido por $\frac{1}{2}(L - L^*)(x) = h \times x$ tiene en la base $\mathcal{B} = (u, v, u \times v)$ las definiciones alternativas $h = \sin\theta u = v \times L(v)$ y por consiguiente*

$$L(x) = (\cos\theta)\, x + (d - \cos\theta)\, \langle u, x\rangle\, u + h \times x.$$

Esto tiene la posible ventaja de acortar un poco el cálculo de la base $\mathcal{B} = (u, v, u \times v)$ porque, al menos en $\left(\mathbb{R}^3, \varepsilon\right)$ con la orientación estándar

$$\begin{pmatrix} 0 & -p & q \\ p & 0 & -r \\ -q & r & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -py + qz \\ px - rz \\ -qx + ry \end{pmatrix} = \begin{pmatrix} r \\ q \\ p \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

luego la matriz antisimétrica a de la izquierda (¡poniendo los coeficientes como se indican, signo incluido!) se corresponde con $h = (r, q, p)^{\top}$.

Problema 429 *Calcular directamente el ángulo y vector axial de la isometría del problema 428.* $\blacklozenge$

Solución. Calculamos

$$\frac{1}{2}\left(a - a^{\top}\right) = \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{pmatrix}^{\top} = \frac{1}{2}\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -\sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}$$

y con la fórmula previa al problema, $h = \frac{1}{2}\left(\sqrt{2}, 0, 1\right)^{\top}$ y $\|h\| = \frac{1}{2}\sqrt{3} = \sin\theta$. El ángulo es $\theta = \frac{2\pi}{3}$. Probablemente, para conocer el ángulo y eje es más corto usar las fórmulas (10.17) y calcular el eje como la recta de vectores propios. $\blacklozenge$

[11] Estamos dando por sabido algo aún no explicado. Desde luego, esto es incorrecto pero la alternativa era cortar cierto tiempo el estudio de las isometrías intercalando las trasformaciones antisimétricas.

Problema 430 *En* $\left(\mathbb{R}^3, \varepsilon\right)$ *nos dan* L *con matriz en la base estándar*

$$a = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \\ 2 & -2 & 1 \end{pmatrix},$$

que es una rotación. Calcular su eje, coseno y seno del ángulo, vector axial, y matriz y base como en el teorema 216.

Problema 431 *Sea* $a \in \mathbb{R}^{3\times 3}$ *una matriz ortogonal con determinante* 1. *Para* $k = \pm 1$, *la matrices* ka *representan dos isometrías de* $\mathbb{R}^3$ *con el producto estándar, una directa y otra inversa. ¿Cómo se relacionan los ángulos? ¿Y los vectores axiales?*

10.11. Endomorfismos antisimétricos

Sea $L : \mathbb{E} \to \mathbb{E}$ un endomorfismo **antisimétrico** del espacio euclidiano $(\mathbb{E}, \omega)$ de dimensión n. Por definición, $\langle L(x), y\rangle = -\langle x, L(y)\rangle$ y la antisimetría equivale a que en cualquier base ortonormal la matriz a de L sea antisimétrica, $a^\top = -a$. Como consecuencia, $\det(L) = (-1)^n \det(L)$ y si n es impar, $\det(L) = 0$. Como $\langle L(x), x\rangle = 0$, siempre $L(x)$ es ortogonal a x. Si además, x es vector propio con valor λ, obtenemos $\lambda \|x\|^2 = 0$ y $\lambda = 0$.

A L asociaremos $M = L^2$ que es simétrica porque $M^* = \left(L^2\right)^* = (L^*)^2 = (-L)^2 = L^2 = M$. Como $M = L^2$ es simétrica, sus valores propios son reales.

Teorema 218 *Sea* u *un vector propio unitario de* $M = L^2$ *con valor propio* $\lambda \neq 0$. *Entonces*

1. $\lambda < 0$, *porque es de hecho* $\lambda = -\|L(u)\|^2$,
2. *El subespacio* $\mathbb{P}_u = \lg(u, L(u))$ *es un plano (o sea, es bidimensional) y es estable por* L.
3. *La restricción de* L *a* $\mathbb{P}_u$ *tiene en la base ortonormal* $\mathcal{U} = (u, L(u)/\|L(u)\|)$ *matriz*

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}\left(L|_{\mathbb{P}_u}\right) = \begin{pmatrix} 0 & -\|L(u)\| \\ \|L(u)\| & 0 \end{pmatrix}.$$

Demostración. Se prueba **1** fácilmente porque

$$\left\langle L^2(u), u\right\rangle = \lambda\|u\|^2 = \lambda, \qquad \left\langle L^2(u), u\right\rangle = \langle L(u), -L(u)\rangle = -\|L(u)\|^2.$$

Si fuese $\dim(\mathbb{P}_u) = 1$ se tendría que $L(u) = \mu u$ y $M(u) = L^2(u) = \mu^2 u = \lambda u$ y $\mu^2 = \lambda < 0$. Imposible. Por otra parte, $L(L(u)) = \lambda u$ por hipótesis, y $\lg(u, L(u))$ es estable por L. Para **3** tenemos

$$L(u) = \|L(u)\|\frac{L(u)}{\|L(u)\|}, \quad L\left(\frac{L(u)}{\|L(u)\|}\right) = \frac{1}{\|L(u)\|}L^2(u) = \frac{\lambda}{\|L(u)\|}u = \frac{-\|L(u)\|^2}{\|L(u)\|}u = -\|L(u)\|u,$$

y resulta la matriz anunciada. ♣

Teorema 219 (estructural de endomorfismos antisimétricos) *Cada endomorfismo antisimétrico* L *de* $\mathbb{E}$ *admite una base ortonormal* $\mathcal{W} = (w_1, \ldots, w_n)$ *donde su matriz tiene forma cuasidiagonal*

$$a = \begin{pmatrix} C_1 & & & & & \\ & \ddots & & & & \\ & & C_m & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \text{ con } C_j = \begin{pmatrix} 0 & -\|L(w_j)\| \\ \|L(w_j)\| & 0 \end{pmatrix},$$

siendo todas las C_j *distintas de la matriz cero y* $2m = r$ *el rango de* L, *que es par.* $C(X)$ *determina la matriz unívocamente puesto que* $C(X) = \left(X^2 + \|L(w_1)\|^2\right)\cdots\left(X^2 + \|L(w_m)\|^2\right)$.

Demostración. Usaremos el teorema 218. Sabemos que $M = L^2$ es simétrica y tendrá todos sus autovalores reales. Si $M = 0$, se sigue que $0 = \langle L(L(v)), v\rangle = -\langle L(v), L(v)\rangle = -\|L(v)\|^2$, de modo que $L = 0$ y el teorema es evidente. En adelante entendemos que $M \neq 0$, tomamos u vector propio unitario de M, y definimos $\mathbb{P}_u = \lg(u, L(u))$, que es un subespacio estable. Por **5** en el teorema 193, $\mathbb{P}_u^\perp$ es estable por $L^* = -L$ y $\mathbb{P}_u^\perp$ es estable por L. La inducción es evidente. Para $n = 1$ debe ser $M = L = 0$ porque si no habría dos vectores independientes u y $L(u)$. Supongamos cierto el teorema para dimensiones $k < n$. Si L tiene un vector propio w, necesariamente $L(w) = 0$ y hay una suma directa y ortogonal de subespacios estables $\mathbb{E} = \lg(w) \oplus \lg(w)^\perp$. Aplicando la hipótesis inductiva a $\lg(w)^\perp$ obtenemos el teorema. Si L no tiene vectores propios, aplicamos la hipótesis inductiva a $\mathbb{E} = \mathbb{P}_u \oplus \mathbb{P}_u^\perp$ e igualmente probamos el teorema. ♣

Es fácil aplicar el teorema si $n = 2{,}3$. Excluyamos $L = 0$. En el caso $n = 2$, calculamos $C(X)$, necesariamente de la forma $X^2 + h^2$ con $h > 0$. La matriz y la base serán

$$a = \begin{pmatrix} 0 & -h \\ h & 0 \end{pmatrix}, \quad \mathcal{W} = (u, L(u)/\|L(u)\|), \ u \text{ vector propio unitario de } L^2.$$

Problema 432 *En $\mathbb{R}^2$ consideramos L y ω que en la base estándar tienen matrices*

$$b = \begin{pmatrix} -4 & -8 \\ 4 & 4 \end{pmatrix} \ y \ \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Probar que L es antisimétrica y determinar una base y la matriz de L como en el teorema 219. ♦

Solución. L es antisimétrica porque $b^\top\Omega = \Omega(-b)$; o sea, $b^\top\Omega + \Omega b = 0$. Como $C(X) = X^2 + 4^2$,

$$a = \begin{pmatrix} 0 & -4 \\ 4 & 0 \end{pmatrix}.$$

Para la base hay que calcular L^2 y un vector propio unitario de L^2. Los cálculos son

$$b^2 = \begin{pmatrix} -4 & -8 \\ 4 & 4 \end{pmatrix}^2 = \begin{pmatrix} -16 & 0 \\ 0 & -16 \end{pmatrix}, \quad u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \|u\|^2 = 1,$$

$$bu = \begin{pmatrix} -4 \\ 4 \end{pmatrix}, \quad \|bu\|^2 = \begin{pmatrix} -4 \\ 4 \end{pmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} -4 \\ 4 \end{pmatrix} = 16,$$

$$\mathcal{U} = \left(u, \frac{L(u)}{\|L(u)\|}\right) = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right).$$

Efectivamente, el ordenador comprueba que

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id})\,\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(L)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -4 & -8 \\ 4 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & -4 \\ 4 & 0 \end{pmatrix}. \ ♦$$

Le vamos a decir al lector como ponerse todos los problemas numéricos que desee *en cualquier dimensión*. Trabajamos en $\mathbb{R}^n$ y tomamos una matriz h invertible, definiendo ω con matriz Ω en $\mathcal{E}$ dada por $\Omega = h^\top h$. La matriz $u = h^{-1}$ tiene columnas que constituyen una base ortonormal debido a que $u_i^\top \Omega u_j$ es la componente (i,j) de $u^\top \Omega u = \left(h^{-1}\right)^\top h^\top h h^{-1} = I$. Tomamos una matriz a tal que $a^\top + a = 0$ (antisimétrica). y definimos $b = h^{-1}ah = uau^{-1}$. Para que b sea antisimétrica (como endomorfismo, no como matriz) se debe cumplir $b^\top\Omega = \Omega(-b)$; o sea, $b^\top\Omega + \Omega b = 0$. Esto es así porque

$$\begin{aligned} b^\top\Omega + \Omega b &= \left(h^{-1}ah\right)^\top h^\top h + h^\top h \left(h^{-1}ah\right) = h^\top a^\top \left(h^{-1}\right)^\top h^\top h + h^\top h h^{-1} a h \\ &= h^\top a^\top h + h^\top a h = h^\top \left(a^\top + a\right) h = 0. \end{aligned}$$

Preparando de este modo el problema sabemos que a es la matriz del teorema 219 y las columnas de u forman base ortonormal del tipo deseado. La justificación es como la del final del problema anterior

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \mathrm{mat}_{\mathcal{E}}^{\mathcal{U}}(\mathrm{id})\,\mathrm{mat}_{\mathcal{E}}^{\mathcal{E}}(L)\,\mathrm{mat}_{\mathcal{U}}^{\mathcal{E}}(\mathrm{id}) = u^{-1}bu = a.$$

Problema 433 *Ponerse un problema similar al anterior para $\mathbb{E} = \mathbb{R}^3$.*

En el caso $n = 3$ preferimos dejar los cálculos e ir a un estudio más teórico basado en las ventajas que ofrece el producto vectorial. Como se va a utilizar el producto vectorial, se puede suponer que $(\mathbb{E}, \omega) = \left(\mathbb{R}^3, \varepsilon\right)$ si el lector no ha estudiado la orientación en general. En ese caso la base ortonormal $\mathcal{U}$ será $\mathcal{E}$. Si no, en $\mathbb{E}$ tenemos una orientación $\mathcal{O}$ con la que definimos el producto vectorial $\times$ y el elemento de volumen Δ, que en el caso particular es $\Delta(x, y, z) = \det(x, y, z)$. Lo esencial de lo que vamos a hacer es probar que *si L es antisimétrica existe $h \in \mathbb{E}$ único tal que $L(x) = h \times x$ y que la función $h \to L$ de $\mathbb{E}$ al espacio de los endomorfismos antisimétricos es un isomorfismo*. Simplificando: Cualquier L antisimétrico es la premultiplicación vectorial por un h que L determina de modo isomórfico. Esto facilita el estudio de los endomorfismos antisimétricos si $n = 3$. Una primera demostración se basa en un puro cálculo. Probamos primero que $L(x) = h \times x$ es antisimétrico porque

$$\langle L(x), y\rangle + \langle x, L(y)\rangle = \langle h \times x, y\rangle + \langle x, h \times y\rangle = \Delta(h, x, y) + \Delta(h, y, x) = 0.$$

(En el caso particular $(\mathbb{E}, \omega) = \left(\mathbb{R}^3, \varepsilon\right)$ lo que se tiene es $\det(h, x, y) + \det(h, y, x) = 0$.) Si tomamos una base ortonormal $\mathcal{U}$ bien orientada, tenemos $e_i \times e_j = e_k$ si (i, j, k) es permutación circular de $(1, 2, 3)$. En el caso particular, $e_1 \times e_2 = e_3$, $e_1 \times e_3 = -e_2$ y $e_2 \times e_3 = e_1$. Es muy fácil ver que si $h = pu_1 - qu_2 + ru_3$ (¡ojo al signo!) se cumple que

$$\mathrm{mat}_{\mathcal{U}}^{\mathcal{U}}(L) = \begin{pmatrix} 0 & -r & -q \\ r & 0 & -p \\ q & p & 0 \end{pmatrix} = a \quad \text{si } h = pu_1 - qu_2 + ru_3 \text{ (¡ojo al signo!).}$$

Por ejemplo, la primera columna sale de

$$L(u_1) = (pu_1 - qu_2 + ru_3) \times u_1 = -qu_2 \times u_1 + ru_3 \times u_1 = qu_3 + ru_2.$$

Sucede que cualquier matriz antisimétrica a es como la que acabamos de escribir, luego, conocida a antisimetrica hay un h y solo uno tal que $L(x) = h \times x$ y la afirmación en cursiva queda probada. Hay otra demostración con interés pero menos directa.

Teorema 220 *Si $L \neq 0$ es antisimétrico y u unitario es ortogonal a $\ker(L)$, el vector $h = u \times L(u)$ cumple que $L(x) = h \times x$ y es el único h tal que $L(x) = h \times x$.*

Demostración. Como $u \neq 0$ no está en $\ker L$, tiene que ser $L(u) \neq 0$ y, como vimos, ortogonal a $\ker L$. Claramente, $(u, L(u) / \|L(u)\|)$ es base ortonormal de $(\ker L)^{\perp}$ y $u \times L(u)$ genera la recta $\ker L$. Si $M(x) = (u \times L(u)) \times x$ debemos probar que $L = M$. Esto es así porque, $\langle u, L(u)\rangle = 0$, luego

$$M(u \times L(u)) = (u \times L(u)) \times (u \times L(u)) = 0 = L(u \times L(u)),$$

$$M(u) = (u \times L(u)) \times u = \|u\|^2 L(u) - \langle u, L(u)\rangle u = L(u),$$

$$M(L(u)) = (u \times L(u)) \times L(u) = \langle u, L(u)\rangle L(u) - \|L(u)\|^2 u = -\|L(u)\|^2 u.$$

¿Es $-\|L(u)\|^2 u = L(L(u))$? Sí porque al ser $L^2(u)$ ortogonal a $L(u)$, ha de ser $L^2(u)$ proporcional a u, y

$$L^2(u) = \left\langle L^2(u), u\right\rangle u = \langle L(u), -L(u)\rangle u = -\left\|L^2(u)\right\| u.$$

Si dos vectores h_1 y h_2 cumplen $h_1 \times x = h_2 \times x$ para todo x, $(h_1 - h_2) \times x = 0$ para todo x y esto solo es posible si $h_1 - h_2 = 0$. Esto concluye la demostración. ♣

En la base $\mathcal{B} = (u \times L(u) / \|L(u)\|, u, L(u) / \|L(u)\|)$, que es ortogonal pero no ortonormal,

$$\mathrm{mat}_{\mathcal{B}}^{\mathcal{B}}(L) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\|L(u)\| \\ 0 & \|L(u)\| & 0 \end{pmatrix}.$$

10.12. Teorema estructural de isometrías afines

Sea $\Lambda(x) = t + L(x)$ una **isometría afín** de $\mathbb{E}$, definida como una transformación afín donde $L : \mathbb{E} \to \mathbb{E}$ es una *isometría lineal*. Pretendemos encontrar una referencia *euclidiana* $\mathcal{R} = (p, u_1, \ldots, u_n)$ en donde Λ tenga una **ecuación normalizada** (otro nombre puede ser **ecuación canónica** o **estándar**) que permita clasificar estas isometrías y estudiar su acción de mover puntos con sencillez. Recordamos

definiciones básicas que valen para cualquier función afín Λ (quizás no isometría) y referencia $\mathcal{R}$ (quizás no euclidiana). En $\mathcal{R}$ las **coordenadas afines de** $x \in \mathbb{E}$ son $\left(x^1, \ldots, x^n\right)$ determinadas por $x - p = \sum_{j=1}^{n} x^j u_j$ (son las "vectoriales" en $(u_1, \ldots, u_n)$ de $x - p$; no de x). La expresión de Λ en $\mathcal{R}$ (que se puede generalizar a $\Lambda : \mathbb{E} \to \mathbb{F}$ y dos referencias $\mathcal{R}$ y $\mathcal{S}$ en $\mathbb{E}$ y $\mathbb{F}$) viene dada en coordenadas expresando las coordenadas de $\Lambda(x) = y$ en $\mathcal{R}$ en función de las de x en $\mathcal{R}$. Necesitamos dos datos: **(a)** la matriz $a = \operatorname{mat}_{\mathcal{U}}^{\mathcal{U}}(L)$ y **(b)** las coordenadas α^i de $\Lambda(p)$ en $\mathcal{R}$, esto es $\Lambda(p) - p = \sum_{i=1}^{n} \alpha^i u_i$. Con esto se calcula $\Lambda(x) - p = y - p = \sum y^i u_i$ de otro modo

$$y - p = \Lambda(x) - p = \Lambda(x) - \Lambda(p) + \Lambda(p) - p = L(x - p) + \sum_{i=1}^{n} \alpha^i u_i = L(x - p) + \sum_{i=1}^{n} \alpha^i u_i$$
$$= L\left(\sum_{j=1}^{n} x^j u_j\right) + \sum_{i=1}^{n} \alpha^i u_i = \sum_{j=1}^{n} x^j \left(\sum_{i=1}^{n} a_j^i u_i\right) + \sum_{i=1}^{n} \alpha^i u_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_j^i x^j\right) u_i + \sum_{i=1}^{n} \alpha^i u_i.$$

Igualamos los coeficientes de las u_i y resultan las fórmulas, sea con ecuaciones o sea con matrices,

$$\begin{cases} y^1 = \Lambda(x)^1 = \sum_{j=1}^{n} a_j^1 x^j + \alpha^1 \\ \vdots\vdots \\ y^n = \Lambda(x)^n = \sum_{j=1}^{n} a_j^n x^j + \alpha^n \end{cases} \qquad \begin{pmatrix} \Lambda(x)^1 \\ \vdots \\ \Lambda(x)^m \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix} + \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^n \end{pmatrix}.$$

Insistimos en que las x^i y $\Lambda(x)^i$ son las coordenadas *en la referencia* $\mathcal{R}$ y no las coordenadas *en la base* $\mathcal{U}$. Los cálculos valen aunque $\mathcal{U}$ no sea base ortonormal, pero si lo es y Λ es una isometría afín, la matriz a de L debe ser ortogonal.

Un importante inciso algebraico. Sea $N : \mathbb{E} \to \mathbb{E}$ un endomorfismo del espacio euclidiano $\mathbb{E}$ de dimensión finita y N^* su adjunto. Tenemos $(\operatorname{im} N)^{\perp} = \ker(N^*)$ porque si $x \in (\operatorname{im} N)^{\perp}$ se tiene $\langle x, N(y)\rangle = 0$ para todo y. Pero entonces $0 = \langle x, N(y)\rangle = \langle N^*(x), y\rangle$ para todo y, lo que implica $N^*(x) = 0$. Visto que $(\operatorname{im} N)^{\perp} \subset \ker(N^*)$, la igualdad de dimensiones implica que $(\operatorname{im} N)^{\perp} = \ker(N^*)$. (Hay otras igualdades análogas; véase el problema 377.) Con esto, $\mathbb{E} = \operatorname{im} N \oplus (\operatorname{im} N)^{\perp}$ nos lleva a $\mathbb{E} = \operatorname{im} N \oplus \ker N^*$. Pero *si añadimos la hipótesis de que* N *es normal*, obtenemos $\ker(N) = \ker(N^*)$ ya que

$$\langle N(x), N(x)\rangle = \langle x, N^*(N(x))\rangle = \langle x, N(N^*(x))\rangle = \langle N^*(x), N^*(x)\rangle,$$

luego $N(x) = 0$ equivale a $N^*(x) = 0$ y $\ker N = \ker N^*$. Evidentemente, $\mathbb{E} = \operatorname{im} N \oplus \ker N^*$ implica $\mathbb{E} = \operatorname{im} N \oplus \ker N$. Aplicaremos esto a $N = \mathrm{id} - L$ siendo L una isometría, por lo que N es normal (fácil). Será fundamental la descomposición en suma directa *ortogonal*

$$\mathbb{E} = \operatorname{im}(\mathrm{id} - L) \oplus \ker(\mathrm{id} - L), \quad \text{y en particular } t = (p - L(p)) + z. \tag{10.18}$$

Conocido t, queda unívocamente determinado z, pero no así p. No obstante, quedará p unívocamente determinado si $\mathrm{id} - L$ es inyectiva.

Intuimos que si Λ tiene un punto fijo p, se comportará en una referencia $\mathcal{R}$ de origen p como su parte lineal L, que siendo una isometría, está bien estudiada. Es una intuición correcta. Puede suceder que Λ no tenga puntos fijos p y parece natural buscar una traslación T tal que $T^{-1} \circ \Lambda = \Gamma$ tenga un punto fijo p. Como Λ y Γ tienen la misma parte lineal L, se intuye también correctamente que lo que sepamos de Γ, algo modificado por T, nos dará la información sobre Λ. Hay que empezar por ver si existen traslaciones T tales que $T^{-1} \circ \Lambda = \Gamma$ fije puntos y cuántas existen. Probaremos que siempre existen y lo sorprendente es que solo hay una posible T.

Teorema 221 *Hay un único* z *tal que* $T_z^{-1} \circ \Lambda = T_{-z} \circ \Lambda = \Gamma$ *tenga al menos un punto fijo. Los pares* (p, z) *tales que* $T_{-z} \circ \Lambda = \Gamma$ *fija* p *son los que dan la descomposición ortogonal* $t = (p - L(p)) + z$. *Los puntos fijos de* Λ, *si existen, forman un subespacio afín, de ecuación* $(\mathrm{id} - L)(x) = t$.

Demostración. Supongamos que exista (p, z) tal que $T_{-z} \circ \Lambda = \Gamma$ fije p. Entonces $-z + \Lambda(p) = -z + t + L(p) = p$ y $t = (p - L(p)) + z$, luego z queda unívocamente fijo por t como su componente en $\ker(\mathrm{id} - L)$. Probada la unicidad de z y $T = T_z$ su existencia es inmediata descomponiendo $t = (p - L(p)) + z$ con (10.18) y repitiendo el cálculo anterior. Si p es punto fijo de Λ (no decimos de Γ) se tiene $t + L(p) = p$ y p cumple $(\mathrm{id} - L)(x) = t$, valiendo asimismo la recíproca. ♣

Dijimos que puede ser que Λ no fije punto alguno pero si lo hace hay otra factorización.

Teorema 222 *Si Λ fija q y L es su parte lineal, se puede factorizar $\Lambda = T_q \circ L \circ (T_q)^{-1} = T_q \circ L \circ T_{-q}$.*

Demostración. Restamos $\Lambda(x) = t + L(x)$ y $q = t + L(x)$, que da $\Lambda(x) - q = L(x-q)$ y la factorización. ♣

La factorización *unica* $\Lambda = T_z \circ \Gamma$ se llama la **factorización canónica** (de Λ). El vector z, unívocamente dado por el teorema 221, se llama el **vector de desplazamiento** (o **deslizamiento**) y su módulo $\|z\| = \mu \geq 0$ el **módulo de desplazamiento** (o **deslizamiento**).

Teorema 223 (de estructura) *Sea $\Lambda = T_z \circ \Gamma$ una isometría afín factorizada como en el teorema 221 con vector de desplazamiento z y módulo de desplazamiento μ. Existe una referencia ortonormal, con origen en cualquier punto fijo p de Γ, en donde Λ tiene la expresión en coordenadas*

$$\begin{pmatrix} \Lambda(x)^1 \\ \Lambda(x)^2 \\ \vdots \\ \Lambda(x)^n \end{pmatrix} = a \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} + \begin{pmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

siendo a una matriz $n \times n$ cuasi-diagonal

$$a = \begin{pmatrix} T_1 & & & & & & \\ & \ddots & & & & & \\ & & T_s & & & & \\ & & & R_1(p_1,q_1) & & & \\ & & & & \ddots & & \\ & & & & & R_k(p_k,q_k) \end{pmatrix}.$$

Las matrices T son 1×1 con 1 o -1 en su único coeficiente; las R son 2×2 de la forma

$$R_j = \begin{pmatrix} p_j & -q_j \\ q_j & p_j \end{pmatrix}, \qquad p_j^2 + q_j^2 = 1, \quad q_j > 0,$$

y los restantes coeficientes de a son cero. Si es $\mu > 0$ tiene que ser $a_1^1 = 1$.

La matriz a y el coeficiente $\mu \geq 0$ dependen solamente de Λ.

Demostración. Definimos $\mathcal{R} = (p, u_1, \ldots, u_n)$ con origen p punto fijo de Γ y $(u_1, \ldots, u_n)$ una base ortonormal como en el teorema 213 para que L tenga la matriz a que presentamos. Las fórmulas generales de Γ en coordenadas se simplifican porque $\Gamma(p) = p$ y

$$\begin{pmatrix} \Gamma(x)^1 \\ \vdots \\ \Gamma(x)^m \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix},$$

pues todas las α^i son cero. Para $z = 0$ será $\Lambda = \Gamma$ y, al ser $\mu = 0$, hemos acabado en este caso.

Si es $z \neq 0$, es $L(z) = z$ y L tiene el valor propio 1. Tomamos una base ortonormal $(u_1, \ldots, u_m)$ como la del teorema de estructura de isometrías con $u_1 = z/\|z\| = z/\mu$ para L.[12] La matriz a de L tendrá $a_1^1 = 1$ porque $L(u_1) = u_1$. La expresión de T_z en coordenadas de la referencia $\mathcal{U} = (p, u_1, \ldots, u_m)$ se calcula por

$$T_z(y) - p = z + y - p = \mu u_1 + \sum_{i=1}^{n} y^i u_i = \left(\mu + y^1\right) u_1 + \sum_{i=2}^{n} y^i u_i,$$

lo que da como expresión en coordenadas para $\Lambda = T_z \circ \Gamma$ que si $\Gamma(x) = y$,

$$\begin{pmatrix} \Lambda(x)^1 \\ \vdots \\ \Lambda(x)^m \end{pmatrix} = T_z \begin{pmatrix} y^1 \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} y^1 + \mu \\ \vdots \\ y^n \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix} + \begin{pmatrix} \mu \\ \vdots \\ 0 \end{pmatrix},$$

[12] Si se es estricto, hay que revisar la demostración del teorema 213, observando que si L tenía el valor propio 1 con vector propio unitario u, se podía tomar este vector como primer vector de la base que se construye.

como queríamos demostrar.

La unicidad es trivial porque Λ determina unívocamente L y z. Entonces, por el teorema 213, L determina unívocamente a, y μ es $\|z\|$. ♣

Se pueden añadir más detalles al teorema examinando la demostración. Puesto que la herramienta esencial es el teorema de estructura de isometrías lineales aplicado a L (que es la parte lineal tanto de Λ como de Γ) podemos decir que la matriz a es construible a partir del polinomio característico $C(X)$ de L. El escalar μ es el módulo de desplazamiento, luego el par $(C(X), \mu)$ permite reconstruir la expresión en coordenadas de Λ tal como se da en el teorema, pero no $\mathcal{R}$. La más detallada clasificación se obtiene estableciendo que $\Lambda_1 = \Lambda_2$ si y solo si en referencias euclidianas (posiblemente diferentes) $\mathcal{R}_1$ y $\mathcal{R}_2$, tienen la misma expresión normalizada del teorema 223. Otra clasificación, interesante si $n = 2, 3$ es ver como es $\mathbf{F}$, espacio de puntos fijos de Λ, que es un subespacio afín o $\emptyset$, y estudiar el comportamiento de T_z en relación con $\mathbf{F}$. Esto da una idea de la acción geométrica de las diversas isometrías.

Problema 434 *Sean Λ_1 y Λ_2 las isometrías afines de $\left(\mathbb{R}^2, \varepsilon\right)$ dadas por*

$$\Lambda_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Lambda_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Hacer el trabajo de los teoremas 221 y 223 para estas isometrías. ♦

Solución. Empezamos con Λ_1. Claramente $\det(L_1) = -1$, luego L_1 es una simetría lineal. La imagen de $\mathrm{id} - L$ esta generada por $(1,-1)^\top$ y el núcleo por $(1,1)^\top$. Determinamos z en la descomposición $t = y + z$ por

$$z = \frac{1}{\left\|(1,1)^\top\right\|^2} \left\langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \text{ luego } \mu = \|z\| = \frac{3\sqrt{2}}{2}.$$

Seguimos calculando p como solución de

$$\left(I - a_{(1)}\right)(x) = y = t - z = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

que es $p = \left(0, \frac{1}{2}\right)$. Una base ortogonal de vectores propios de L_1 es $\mathcal{U} = (u_1, u_2)$ siendo $u_1 = \frac{1}{\sqrt{2}}(1,1)^\top$ y $u_2 = \frac{1}{\sqrt{2}}(-1,1)^\top$ con valores propios $+1$ y -1. La referencia es

$$\mathcal{R} = \left(\begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right),$$

y la ecuación normalizada en las coordenadas *afines euclidianas* (u, v) de la nueva referencia es

$$\Lambda \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \frac{3}{2} \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}.$$

Cuando estudiemos con más detalle las isometrías (enseguida) veremos que hay composición de una simetría y una traslación a lo largo del eje de simetría. En efecto, primero $(u, v)^\top \to (u, -v)^\top$ y luego se traslada a lo largo del eje U a $\left(u + \frac{3\sqrt{2}}{2}, -v\right)^\top$. Si no nos preocupara la referencia y solo quisiéramos la ecuación, calcularíamos

$$C_{a_{(1)}}(X) = X^2 - 1 = (X - 1)(X + 1),$$

que nos daría la matriz. Falta $\mu = \frac{3\sqrt{2}}{2}$ y ahí si hace falta un cálculo como al principio.

Seguimos con Λ_2. Ahora, $\det(L_2) = 1$ y sabemos que L_2 es una rotación, obviamente distinta de $\pm\,\mathrm{id}$, luego no tendrá valores propios e $\mathrm{id} - L_2$ será invertible. Habrá un solo punto fijo p (el centro de la rotación) que se obtiene resolviendo

$$(I - a_2)\, x = t, \quad \begin{pmatrix} 1 - \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} - 1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

que nos da $p = \frac{1}{2}\left(-1-2\sqrt{2}, \sqrt{2}+3\right)^\top$. Como base ortonormal asociada casi no hay que trabajar, porque vale $(e_1, e_2) = \mathcal{E}$, y $\mathcal{R} = (p, e_1, e_2)$ es la referencia pedida. La ecuación normalizada en las coordenadas *afines* (u, v) de la nueva referencia es

$$\Lambda \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix},$$

ya que $\ker(\mathrm{id} - L_2) = 0$, luego $\mu = 0$. ♦

Problema 435 *Hacer el problema similar para*

$$\Lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ h \end{pmatrix}, \quad h \in \mathbb{R}.$$

Hemos hecho cálculos sin preocuparnos demasiado de los atajos porque queremos ilustrar los teoremas, pero si nos conformamos con información más limitada, fundamentalmente la que da la factorización $\Lambda = T_z \circ \Gamma$, los cálculos pueden ser más breves. El vector de desplazamiento z, que juega un papel esencial entra en escena por una razón algebraica (nos facilita la descomposición $t = (p - L(p)) + z$ en los cálculos) pero nos gustaría una definición más geométrica. Ahí va.

Teorema 224 *Para la isometría afín $\Lambda(x) = t + L(x)$ existe un vector z y solo uno que cumple el par de condiciones* ***(a)*** *$T_z \circ \Lambda = \Lambda \circ T_z$ y* ***(b)*** *$\Gamma = T_{-z} \circ \Lambda$ tiene al menos un punto fijo p. Los vectores p, z citados son los que aparecen en la descomposición de suma directa*

$$t = y + z = (L(p) - p) + z, \qquad y \in \mathrm{im}(\mathrm{id} - L), \quad z \in \ker(\mathrm{id} - L).$$

Demostración. La condición de p fijo da $p = (T_{-z} \circ \Lambda)(p) = -z + t + L(p)$ y $t = p - L(p) + z$ con $p - L(p)$ evidentemente en $\mathrm{im}(\mathrm{id} - L)$. Por otra parte,

$$(T_z \circ \Lambda)(x) = z + t + L(x), \qquad (\Lambda \circ T_z)(x) = t + L(x + z) = t + L(x) + L(z)$$

implica $z - L(z) = 0$. Con todo esto vemos que $t = p - L(p) + z$ es la descomposición en suma directa $\mathbb{E} = \mathrm{im}(\mathrm{id} - L) \oplus \ker(\mathrm{id} - L)$ que ya conocemos. ♣

Obtenemos un dato adicional de este teorema, y es que T_z conmuta con la parte fija Γ, cosa interesante en ciertas demostraciones teóricas. Otra cosa obvia es que los puntos fijos p (si existen) de Λ son las soluciones de $(\mathrm{id} - L)(x) = t$ y es por tanto este conjunto de puntos fijos $\mathbf{F}$ es un subespacio afín con dirección $\mathbb{F} = \ker(\mathrm{id} - L)$. Observemos que en particular el vector de desplazamiento, si $\mathbf{F} \neq \emptyset$, está en la dirección $\mathbb{F}$ de $\mathbf{F}$. Esto tendrá su importancia.

10.13. Descripción de las isometrías

La vía natural es usar la factorización $\Lambda = T_z \circ \Gamma$ que Λ determina unívocamente. El caso $z = 0$ es equivalente a que Λ tenga puntos fijos porque, automáticamente, existen p tales que $(\mathrm{id} - L)(p) = t$ y tales puntos son precisamente los que fija Λ. En todo caso, Γ tiene puntos fijos p y cualquiera de ellos permite escribir $\Gamma(x) = p + L(x - p)$. En una referencia con origen en p, los cambios de variable $\xi = x - p$ y $\eta = \Gamma(x) - p$ nos dicen que Γ se identifica con $\eta = L(\xi)$. Esto se puede precisar mejor, pero en esencia afirma que Γ se identifica con L y que si $\mathbb{F}$ es el subespacio *vectorial* de puntos fijos por L, $\mathbf{F} = p + \mathbb{F}$ es el subespacio *afín* de puntos fijos por Γ. Una clasificación no tan fina como la del teorema 223 pero cómoda para una descripción general, se basa en distinguir si es z nulo o no nulo (equivale a que Λ tenga o no puntos fijos) y luego examinar la dimensión del conjunto de puntos fijos de Γ, que coindide con la dimensión del espacio de puntos fijos de L. Advertimos que, a la hora de poner nombres, la identidad $\mathrm{id}_{\mathbb{E}}$ se considera tanto rotación como traslación. Por ello, dado que una correcta clasificación requiere que las clases distintas sean disjuntas, entenderemos que estamos clasificando las isometrías *distintas de la identidad*, si bien hablaremos laxamente de "clasificar las isometrías". Todo lo dicho vale en dimensión arbitraria, pero nosotros nos limitamos ahora al caso $n = \dim(\mathbb{E}) = 2, 3$.

10.13.1. El caso bidimensional

Siguiendo el esquema recién expuesto, clasificamos del modo siguiente:

1. *Caso $z = 0$, equivalente a que Λ tiene puntos fijos.* Según la dimensión del espacio de puntos fijos de Γ o L, que es la misma,

   a) Si $\dim(\mathbb{F}) = 2$, entonces Λ es la **identidad**, caso particularísimo de rotación y traslación.

   b) Si $\dim(\mathbb{F}) = 1$, entonces L es una **simetría** que fija una recta $\mathbb{F}$ y $\Lambda = \Gamma$ fija $\mathbf{F} = p + \mathbb{F}$.

   c) Si $\dim(\mathbb{F}) = 0$, entonces L es una **rotación distinta de la identidad**. Al ser $\mathbb{F} = 0$ y $\mathbf{F} = \{p\}$, solo hay un punto fijo p, que es el **centro de la rotación.**

2. *Caso $z \neq 0$, equivalente a que Λ no tiene puntos fijos.* Según la dimensión del espacio de puntos fijos de Γ o L, que es la misma,

   a) Si $\dim(\mathbb{F}) = 2$, entonces $\Lambda = T_z$ porque $\Gamma = \mathrm{id}$. Tenemos una **traslación que no es la identidad.**

   b) Si $\dim(\mathbb{F}) = 1$, entonces Γ es una simetría respecto a una recta $\mathbf{F} = p + \mathbb{F}$ y T_z es una traslación con vector z en la dirección de $\mathbb{F}$ (¡inevitablemente!). Este movimiento se llama **deslizamiento reflejado** o **traslación con simetría** porque Γ envía x a su simétrico x' respecto a una recta y luego lo traslada (se desliza x') según z en la dirección de esa recta.[13] Obsérvese que el hecho, señalado más arriba, de ser $T_z \circ \Gamma = \Gamma \circ T_z$ dice que es (al menos en este caso) lo mismo reflejar e igualar que igualar y reflejar.

   c) $\dim(\mathbb{F}) = 0$ es imposible porque $\mathbb{F} = \ker(\mathrm{id} - L)$ y no habría entonces vectores propios $z \neq 0$. Por supuesto puede componerse una traslación y una rotación, pero se obtiene alguno de los casos anteriores (¿cuál?).

El resumen es que, si $n = 2$ y excluimos la identidad, hay cuatro tipos de isometrías: traslaciones, rotaciones, simetrías, y deslizamientos reflejados.

10.13.2. El caso tridimensional

Siguiendo el esquema recién expuesto, clasificamos del modo siguiente:

1. *Caso $z = 0$, equivalente a que Λ tiene puntos fijos.* Según la dimensión del espacio de puntos fijos de Γ o L, que es la misma,

   a) Si $\dim(\mathbb{F}) = 3$, entonces Λ es la **identidad**, caso particularísimo de rotación y traslación.

   b) Si $\dim(\mathbb{F}) = 2$, entonces L es una **simetría (especular)** que fija un plano $\mathbb{F}$ y $\Lambda = \Gamma$ fija $\mathbf{F} = p + \mathbb{F}$.

   c) Si $\dim(\mathbb{F}) = 1$, entonces L es una **rotación distinta de la identidad**. Al ser $\mathbb{F} = \lg(u)$ y $\mathbf{F} = p + \mathbb{F}$, hay una recta de puntos fijos, que es el **eje de la rotación.**

   d) Si $\dim(\mathbb{F}) = 0$, entonces L es una **rotosimetría.** Al ser $\mathbb{F} = 0$ y $\mathbf{F} = \{p\}$, hay un solo punto fijo p invariante por $\Gamma = \Lambda$. Lo que hace Λ es rotar respecto a un eje ortogonal a un plano $\mathbb{P}$ y luego simetrizar respecto a este plano, conmutando estas dos transformaciones.

2. *Caso $z \neq 0$, equivalente a que Λ no tiene puntos fijos.* Según la dimensión del espacio de puntos fijos de Γ o L, que es la misma,

   a) Si $\dim(\mathbb{F}) = 3$, entonces $\Lambda = T_z$ porque $\Gamma = \mathrm{id}$. Tenemos una **traslación que no es la identidad.**

[13] En inglés *glide reflection* y en frances *glissage* o *symétrie glissée*. Nuestra elección de *deslizamiento reflejado* es un nombre largo, pero muy descriptivo. Advirtamos, no obstante, que casi nadie lo usa.

b) Si $\dim(\mathbb{F}) = 2$, entonces Γ es una simetría respecto a un plano $\mathbf{F} = p+\mathbb{F}$ y T_z es una traslación con vector z en la dirección de $\mathbb{F}$. Este movimiento se llama **deslizamiento reflejado** o **traslación con simetría** porque Γ envía x a su simétrico x' respecto a una recta y luego lo traslada (x se desliza a x') en la dirección de esa recta. Obsérvese que el hecho, señalado más arriba, de ser $T_z \circ \Gamma = \Gamma \circ T_z$ dice que es (al menos en este caso) lo mismo reflejar e igualar que igualar y reflejar. Hemos repetido palabra por palabra, excepto el cambio de "recta" por "plano" lo hecho en **2.b**.

c) Si $\dim(\mathbb{F}) = 1$, entonces L es una rotación distinta de la identidad. Al ser $\mathbb{F} = \lg(u)$ y $\mathbf{F} = p + \mathbb{F}$, hay una recta de puntos fijos por Γ (su eje de la rotación), y necesariamente $z \in \mathbb{F}$. Lo que hace Λ es componer la rotación Γ con una traslación en la dirección del eje de Γ. El nombre más común es **movimiento helicoidal**, pudiendo usarse otros nombres más complicados pero más descriptivos, como **deslizamiento rotado**, **rototraslación**, etc.

d) $\dim(\mathbb{F}) = 0$ es imposible porque $z \neq 0$ es vector propio de $\mathrm{id} - L$ y no puede ser $\mathbb{F} = 0$.

Tenemos $T_z \circ \Gamma = \Gamma \circ T_z$ luego se puede simetrizar o rotar primero y luego trasladar, o hacerlo al revés. *El resumen es que, excluyendo la identidad, hay seis tipos de isometrías para $n = 3$, que son traslaciones, rotaciones, simetrías, rotosimetrías, deslizamientos reflejados (respecto a un plano), y deslizamientos rotados (movimientos helicoidales).* Es curioso que mientras la clasificación para $n = 3$ la realizó Euler en 1776, fue Chasles en 1832 quien lo hizo en el caso más sencillo de $n = 2$.

10.14. Problemas sobre isometrías afines

Ya hemos dicho muchas veces que al implementar un teorema que nos dice que en una determinada base o referencia un ente matemático (función lineal, bilineal, isometría, etc,) tiene una expresión normalizada, canónica o estándar, es mucho más sencillo el determinar esta expresión que la base o referencia para la cual el ente se representa de esta forma distinguida. Por poner un ejemplo, si L es un endomorfismo diagonalizable, $C_L(X)$, conocidas sus raíces, nos permite encontrar la matriz que tendrá L en una base diagonalizadora. Otra posibilidad, todavía más sencilla es utilizar el teorema 214 cuando $n = 3$. Si tenemos ahora $\Lambda(x) = t + L(x)$ como isometría afín, podemos encontrar su expresión bien sea factorizada como $\Lambda = T_z \circ \Gamma$ o bien como en el teorema 223, conociendo tan solo las raíces de $C_L(X)$ y el módulo del vector z proyectado de t en $\ker(\mathrm{id} - L)$. Si $t = y + z \in \mathrm{im}(\mathrm{id} - L) \oplus \ker(\mathrm{id} - L)$ tenemos que el módulo de desplazamiento μ es calculable por

$$\mu^2 = \|z\|^2 = d(t, \mathrm{im}(\mathrm{id} - L))^2 = \frac{\Gamma(t, b_1, \ldots, b_m)}{\Gamma(b_1, \ldots, b_m)}$$

siendo $(b_1, \ldots, b_m)$ una base de cualquier tipo de $\mathrm{im}(\mathrm{id} - L)$.

Problema 436 *En $\mathbb{R}^3$ con el producto euclidiano estándar ε nos dan $\Lambda : \mathbb{R}^3 \to \mathbb{R}^3$ con matriz*

$$\Lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2}\sqrt{2} & \frac{1}{2} \\ \frac{1}{2}\sqrt{2} & 0 & -\frac{1}{2}\sqrt{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

en la referencia estándar. Dar la matriz de Λ de acuerdo con el teorema 223. Describir de modo más o menos intuitivo qué movimiento representa.

Problema 437 *Se considera $\Lambda : \mathbb{E} \to \mathbb{E}$ dada por $\Lambda(x) = t - x$ con $t \neq 0$. Dar su matriz, de acuerdo con el teorema 223, y el conjunto de sus puntos fijos. El producto euclidiano ω es arbitrario. ¿Depende la respuesta de cómo se elija ω? Describir con palabras qué hace Λ.*

Queremos estudiar las **simetrías afines** Σ respecto a hiperplanos $\mathbf{H}$ que no son hiperplanos vectoriales. Primero hay que definirlas. Escribamos $\mathbf{H} = p + \mathbb{H}$, siendo $p \in \mathbf{H}$ y $\mathbb{H}$ un hiperplano vectorial. Tomemos la simetría *vectorial* $S : \mathbb{E} \to \mathbb{E}$ respecto de $\mathbb{H}$. Definimos la simetría *afín* $\Sigma : \mathbb{E} \to \mathbb{E}$ respecto de $\mathbf{H}$ por la condición $\Sigma(x) = p + S(x - p)$. Existen otros p' tales que $\mathbf{H} = p + \mathbb{H} = p' + \mathbb{H}$ y hemos de probar que si $\Sigma'(x) = p' - S(x - p')$ se tiene $\Sigma = \Sigma'$. Restando,

$$\Sigma(x) - \Sigma'(x) = (p - p') - S(p - p') = (p - p') - (p - p') = 0$$

ya que $p - p' \in \mathbb{H}$ que está fijado por S. La fórmula $\Sigma(x) = p + S(x - p)$ equivale a $\Sigma = T_p \circ S \circ (T_p)^{-1}$, no dependiendo Σ de la elección de p en $\mathbf{H}$.

Problema 438 *Sea $\mathbb{H}$ un hiperplano vectorial y $p \neq 0$ un vector ortogonal a él. Consideramos el hiperplano afín $\mathbf{H} = p + \mathbb{H}$. Dar en términos de p la fórmula de la simetría especular Σ respecto de $\mathbf{H}$.* ♦

Solución. Es un puro cálculo. Escribimos primero la fórmula de la simetría vectorial respecto al plano ortogonal a p, que es

$$S(x) = x - \frac{2\langle x, p\rangle}{\|p\|^2} p.$$

A continuación, $\Sigma = T_p \circ S \circ T_p^{-1}$; o sea,

$$\Sigma(x) = T_p \circ S(x-p) = T_p \circ \left(x - p - \frac{2\langle x-p, p\rangle}{\|p\|^2} p \right) = x - \frac{2\langle x-p, p\rangle}{\|p\|^2} p = S(x) + 2p.$$

Esta fórmula de Σ tiene una sencilla descripción verbal: se ejecuta primero la simetría lineal S respecto a $\mathbb{H}$ y luego se traslada con vector $2p$ en dirección perpendicular al plano. Obsérvese que $\mathbf{H}$ no pasa por el origen y que $p \in \mathbf{H}$ es el punto de $\mathbf{H}$ más próximo al origen. Otro modo de dar la respuesta es que si se usa la fórmula tipo $\Lambda = T_t \circ L$ se tiene que el vector t de la traslación es $2p$ y la parte lineal es la simetría vectorial S respecto al plano ortogonal a p. ♦

Problema 439 *Para el producto euclidiano estándar de $\mathbb{R}^n$, calcular el simétrico de $u = (1, 1, \ldots, 1)$ respecto al hiperplano $\mathbf{H}$ de ecuación $x^1 + \ldots + x^n = 1$.*

Problema 440 *Consideramos la factorización $\Lambda = T_z \circ \Gamma$ del teorema 221. Probar que Λ, T_z y Γ conmutan entre sí. ¿Es cierto que si es Λ fija p también fija $T_z(p)$?*

Problema 441 *Probar que toda traslación $T_v(x) = x + v$, $v \neq 0$, es expresable como composición de dos simetrías especulares respecto a planos ortogonales a v pudiendo elegir una libremente.* ♦

Solución. Sea $\mathbb{H}$ es el hiperplano vectorial ortogonal a v y $u = v/\|v\|$. Simplifica el cálculo comprobar que para $\gamma \in \mathbb{R}$ se tiene para $S(x) = x - 2\langle x, u\rangle u$ que $T_{\gamma u} \circ S = S \circ T_{-\gamma u}$ (fácil, pero ojo al signo). Tomamos puntos $p = \alpha u$ y $q = \beta u$ y vamos a calcular la composición de Σ_p y Σ_q, que es la composición de las simetrías especulares respecto a los planos $\mathbf{H}_p$ y $\mathbf{H}_q$ con dirección $\mathbb{H}$ y que pasan por p y q. Sabemos que $\Sigma_p = T_{\alpha u} \circ S \circ T_{-au}$ y $\Sigma_q = T_{\beta u} \circ S \circ T_{-\beta u}$. Entonces

$$\Sigma_p \circ \Sigma_q = T_{\alpha u} \circ S \circ T_{-au} \circ T_{\beta u} \circ S \circ T_{-\beta u} = T_{2\alpha u} \circ S^2 \circ T_{-2\beta u} = T_{2(\alpha-\beta)u}.$$

Eligiendo α y β a nuestro gusto de modo que sea $2(\alpha - \beta)u = v$ queda $\Sigma_p \circ \Sigma_q(x) = T_v(x)$. ♦

Dados p y q distintos, $m = \frac{p+q}{2}$ es el **punto medio del segmento entre** p y q. El hiperplano $\mathbf{M}$ que pasa por m y cuya dirección $\mathbb{M}$ es el hiperplano vectorial ortogonal a $p - q = v$ se llama el **hiperplano mediatriz** (del segmento entre p y q). La intuición nos dice que si Σ es la simetría especular respecto de $\mathbf{M}$, se tiene $\Sigma(p) = q$ y $\Sigma(q) = p$.

Problema 442 *Confirmar la intuición que acabamos de describir*[14] *y la fórmula de Σ en términos de p y q. Probar además que Σ es lineal si y solo si $\|p\| = \|q\|$.*

Solución parcial. Sea S la simetría vectorial respecto al plano ortogonal a $v = p - q$. Sabemos por el teorema 222 que $\Sigma = T_m \circ S \circ T_{-m}$. Entonces,

$$\Sigma(p) = T_m \circ S\left(p - \frac{p+q}{2}\right) = T_m \circ S\left(\frac{p-q}{2}\right) \overset{*}{=} T_m\left(-\frac{p-q}{2}\right) = \frac{p+q}{2} - \frac{p-q}{2} = q.$$

Usamos en $\overset{*}{=}$ que $p - q = 2v$, ortogonal a $\mathbb{M}$, luego $S(v) = -v$. Como $\Sigma^2 = \mathrm{id}$, $\Sigma(q) = \Sigma^2(p) = p$. ♦

Un **deslizamiento reflejado** es una isometría de la forma $\Delta = T_w \circ \Sigma$, siendo Σ una simetría especular (no necesariamente vectorial) y $T_w \neq \mathrm{id}$ una traslación con w en la dirección del hiperplano que fija Σ. Por el teorema 221 se puede factorizar de modo único $\Delta = T_z \circ \Gamma$ con Γ fijando puntos, y al ser esta factorización única, $w = z$ y $\Sigma = \Gamma$. Como $T_z \circ \Gamma = \Gamma \circ T_z$ (teorema 224) tenemos $\Delta^2 = T_w \circ \Sigma \circ \Sigma \circ T_w = T_{2w}$; o sea, Δ^2 es una traslación.

[14] Es curioso que esto vale en *cualquier* espacio euclidiano, si bien es cierto que al cambiar el producto ω, cambia también el hiperplano mediatriz.

Problema 443 *Nos dan la isometría $\Lambda(x) = t + S(x)$ siendo S una simetría lineal de plano $\mathbb{H}$. Dar condiciones necesarias y suficientes para que Λ sea un deslizamiento reflejado y, cuando lo sea, calcular el vector de desplazamiento en función de t y S. Indicación: Queda una bonita respuesta si se observa que* $\ker(\mathrm{id} - S) = \mathrm{im}(\mathrm{id} + S)$.

Problema 444 *Consideramos una isometría $\Lambda(x) = t + S(x)$ siendo $S(x) = x - 2\langle x, u\rangle u$ con $\|u\| = 1$. Calcular la referencia $\mathcal{R}$ del teorema 223. (No cuesta mucho e intervienen t y u en la respuesta.)*

10.15. Factorización de isometrías afines

Querríamos tener un teorema similar al teorema de Cartan-Dieudonne (teorema 215) que asegure que toda isometría afín es factorizable con simetrías especulares (afines). Factorizamos $\Lambda = T_z \circ \Gamma$ como en el teorema 221. Si $z = 0$ tenemos $\Lambda = \Gamma$ y se intuye que $\Gamma = \Lambda$ será factorizable con tantas simetrías como se precisen para factorizar su parte lineal L. Si $z \neq 0$ y factorizamos T_z con dos simetrías (ver el problema 441), tendremos Λ también factorizable. En el enunciado del teorema, que confirma todo esto, se estudia también el número de simetrías posibles y qué propiedades pueden tener. Para no hacer el enunciado demasiado pesado adelantamos notación. Tenemos $\Lambda = T_z \circ \Gamma$, una isometría afin factorizada como en el teorema 221 con L como aplicación lineal asociada. Será $\mathbb{F}$ el espacio de vectores fijos de L, de dimensión m. Recordemos que $z \in \mathbb{F}$ pues $(\mathrm{id} - L)(z) = 0$. Intervendrá cuando sea $z \neq 0$ (o sea, cuando Λ no tenga puntos fijos).

Teorema 225 (de factorización afín) *Con las notaciones expuestas*

1. *Si Λ tiene conjunto de puntos fijos $\mathbf{F} = p + \mathbb{F}$,*
   a) *Se puede factorizar $\Lambda = \Sigma_1 \circ \cdots \circ \Sigma_{n-m}$ con $n - m$ simetrías fijando también $\mathbf{F}$ y $n - m$ es el mínimo número posible de isometrías necesarias.*
   b) *Dada la simetría Σ que fije $\mathbf{F}$, se puede elegir una factorización $\Lambda = \Sigma_1 \circ \cdots \circ \Sigma_{n-m}$ con simetrías fijando también $\mathbf{F}$ siendo a voluntad $\Sigma = \Sigma_1$ o $\Sigma = \Sigma_{n-m}$.*[15]
2. *Si Λ no tiene ningún punto fijo se puede factorizar $\Lambda = \Sigma_1 \circ \cdots \circ \Sigma_{n-m+2}$ con $n - m + 2$ simetrías, nunca con $n - m + 1$, ni estrictamente menos de $n - m$.*

Demostración. Supongamos que $p = \Lambda(p)$ y escribamos $\Lambda = T_p \circ L \circ (T_p)^{-1}$ (teorema 222). Aplicamos el teorema 215 a L y factorizamos como $L = S_1 \circ \cdots \circ S_{n-m}$ con $n - m$ simetrías vectoriales S_j con hiperplano conteniendo a $\mathbb{F}$ y el número $n - m$ mínimo posible. Entonces

$$\Lambda = T_p \circ (S_1 \circ \cdots \circ S_{n-m}) \circ T_p^{-1} = \left[T_p \circ S_1 \circ T_p^{-1}\right] \circ \cdots \circ \left[T_p \circ S_{n-m} \circ T_p^{-1}\right] = \Sigma_1 \circ \cdots \circ \Sigma_{n-m}.$$

En general, si Λ es composición de funciones afines $\Lambda_1, \ldots, \Lambda_k$ con partes lineales $L_1, \ldots, L_k$, la parte lineal L de Λ es $L_1 \circ \cdots \circ L_k$. Si se pudiera factorizar $\Lambda = \Sigma'_1 \circ \cdots \circ \Sigma'_k$ con $k < n - m$ y todas las Σ'_j fijando p, las partes lineales S'_j de cada Σ'_j, que son simetrías especulares lineales, verificarían

$$L = T_p^{-1} \circ \Lambda \circ T_p^{-1} = \left[T_p \circ \Sigma'_1 \circ T_p^{-1}\right] \circ \cdots \circ \left[T_p \circ \Sigma'_k \circ T_p^{-1}\right] = S'_1 \circ \ldots \circ S'_k$$

y L sería factorizable con $k < n - m$ simetrías lineales especulares, contradiciendo el teorema 215.

Sea Σ una simetría que fija el hiperplano $\mathbf{H} = p + \mathbb{H}$ y $\mathbf{H} \supset \mathbf{F}$. Si $v \in \mathbb{F}$ se tiene $p + v = T_p(v) \in \mathbf{H}$ y $\Sigma(T_p(v)) = T_p(v)$; o sea $T_p^{-1} \circ \Sigma \circ T_p(v) = v$. Pero $\Sigma = T_p \circ S \circ T_p^{-1}$ y $S = T_p^{-1} \circ \Sigma \circ T_p(v)$, luego hemos mostrado que $\mathbf{H} \supset \mathbf{F}$ implica que $\mathbb{H} \supset \mathbb{F}$. Al factorizar $L = S_1 \circ \cdots \circ S_{n-m}$ podemos, por esto recién probado, tomar $S = S_1$ aplicando el teorema 215. Es ya inmediato que

$$\Lambda = T_p \circ (S \circ S_2 \circ \cdots \circ S_{n-m}) \circ T_p^{-1} = \Sigma \circ \Sigma_2 \circ \cdots \circ \Sigma_{n-m}.$$

Para conseguir $\Sigma = \Sigma_{n-m}$ se hace una demostración análoga. Queda totalmente probado **1**.

Si Λ no tiene puntos fijos, escribimos $\Lambda = T_z \circ \Gamma$ y, como Γ los tiene y la parte lineal de Γ es también L, podemos factorizar $\Gamma = \Sigma_1 \circ \cdots \circ \Sigma_{n-m}$ de acuerdo con **1**. Por el problema 441 también se puede factorizar $T_z = \Sigma_{-1} \circ \Sigma_0$, y $\Lambda = \Sigma_{-1} \circ \Sigma_0 \circ \Sigma_1 \circ \cdots \circ \Sigma_{n-m}$ queda factorizada con $n - m + 2$ simetrías.

[15] Hay que insistir, porque es un punto sutil, que no se pretende simplemente que S aparezca en la factorización de L, sino que aparezca en la que tiene el mínimo número de factores.

Estudiemos $\Lambda = \Sigma_1 \circ \ldots \circ \Sigma_k$ con $k < n - m + 2$. No puede ser $k = n - m + 1$ porque Λ tendría una parte lineal composición de $n - m + 1$ simetrías, por el párrafo anterior también sería composición de $n - m + 2$, y los determinantes serían opuestos. Supongamos $k < n - m$ y sean $v_1, \ldots, v_k$ son los vectores de las simetrías S_j, las partes lineales de las Σ_j. De manera análoga a como se ha hecho en el caso lineal (teorema 215), sería $L = S_1 \circ \ldots \circ S_k$ y $\mathbb{F}$ contendría a $\lg(v_1, \ldots, v_k)^\perp$. Entonces,

$$m = \dim(\mathbb{F}) \geq \dim \lg(v_1, \ldots, v_k)^\perp = n - \dim \lg(v_1, \ldots, v_k) \geq n - k > n - (n - m) = m,$$

que es contradictorio. ♣

Llama la atención que no hemos descartado en **2** que Λ sin puntos fijos sea factorizable con $n - m$ isometrías. Excepcionalmente esto es posible. En el plano tomamos dos simetrías lineales S_1 y S_2 respecto a rectas $\mathbb{D}_1$ y $\mathbb{D}_2$ perpendiculares y otra simetría Σ_0 respecto a una recta $p + \mathbb{D}_1$ que no pasa por el origen. Se cumple que $T_z = \Sigma_0 \circ S_1$ es una traslación no nula y $\Gamma = S_1 \circ S_2 = -\operatorname{id}$ es una simetría central (lineal). La composición $\Lambda = T_z \circ S_1 \circ S_2 = T_z \circ (-\operatorname{id})$ es una simetría central afín con parte lineal $-\operatorname{id} = L$. Claramente L tiene $\mathbb{F} = 0$ y $n - m + 2 = 2 - 0 + 2 = 4$. Sin embargo, $\Lambda = \Sigma_0 \circ S_1 \circ S_1 \circ S_2 = \Sigma_0 \circ S_2$ factorizable tan solo con $2 = n - m$ simetrías. Posiblemente exponer un caso general.

Teorema 226 *Con las notaciones del teorema anterior, suponemos que $\mathbb{F} = 0$, luego Γ solo fija un punto p. Entonces $\Lambda = T_z \circ \Gamma$ es factorizable con solo n simetrías, aunque sea $z \neq 0$. Como consecuencia, cualquier isometría es factorizable con $n + 1$ simetrías como máximo.*

Demostración. La traslación T_z con $z \neq 0$ es factorizable como $T_z = \Sigma_0' \circ \Sigma_1'$ con Σ_1' fijando p (problema 441) y Γ como $\Gamma = \Sigma_1 \circ \cdots \circ \Sigma_n$ fijando todas las Σ_j el punto p y con libertad de elección de Σ_1 con tal de que fije p (**1** en el teorema 225 aplicado a Γ). Pues bien, tomamos $\Sigma_1 = \Sigma_1'$ y $\Lambda = \Sigma_0' \circ \Sigma_2 \circ \cdots \circ \Sigma_n$ pues $\Sigma_1 \circ \Sigma_1 = \operatorname{id}$.

Si $m = \dim(\mathbb{F}) \neq 0$ el que Λ sea factorizable con $n - m + 2$ isometrías y $n - m + 2 \leq n + 1$ da en estos casos la segunda afirmación. El párrafo anterior la implica también como cierta si $m = 0$. ♣

Problema 445 *Suponiendo $n = 3$, hacer una recuento (justificado) del* mínimo *de simetrías necesarias para factorizar una isometría Λ que no fije puntos.*

10.16. Más problemas sobre isometrías afines

Problema 446 *En $\mathbb{R}^2$ nos dan el producto euclidiano ω que en la base estándar tiene matriz*

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

Dar todos los deslizamientos reflejados (pedimos la ecuación) que fijen, no punto a punto pero sí globalmente, la recta **D** *que pasa por $(1,1) = u$ con dirección $\lg(v)$ y $v = (0,1)^\top$.* ♦

Solución. Sea Σ la simetría respecto de **D**. Todos los deslizamientos Δ serán de la forma $\Delta = T_z \circ \Sigma$ con z no nulo en la dirección de **D**. La ecuación de Σ es $\Sigma = T_u \circ S \circ T_u^{-1}$ y S la simetría ortogonal respecto a $\lg(v)$. Para empezar necesitamos w ortogonal a $(0,1)^\top$, que calculamos con

$$0 = (0,1)\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = x + 2y.$$

Vale por tanto $(2,-1)^\top = w$ con $\|w\|^2 = 2$. Seguimos calculando

$$S\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - 2\frac{1}{2}\left\langle \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - x\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -x \\ x + y \end{pmatrix},$$

$$\Sigma\begin{pmatrix} x \\ y \end{pmatrix} = T_u \circ S \circ T_u^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = T_u \circ S\begin{pmatrix} x - 1 \\ y - 1 \end{pmatrix} = T_u\begin{pmatrix} -(x-1) \\ (x-1) + (y-1) \end{pmatrix} = \begin{pmatrix} 2 - x \\ x + y - 1 \end{pmatrix}.$$

Un vector arbitrario en la dirección de **D** es $z = (0, \lambda)^\top$ con $\lambda \in \mathbb{R}$. Podemos finalizar con

$$\Delta\begin{pmatrix} x \\ y \end{pmatrix} = T_z \circ \Sigma\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda \end{pmatrix} + \begin{pmatrix} 2 - x \\ x + y - 1 \end{pmatrix} = \begin{pmatrix} 2 - x \\ x + y + \lambda - 1 \end{pmatrix}.$$

Si se quiere en forma matricial

$$\Delta\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ \lambda - 1 \end{pmatrix},$$

y, en todos los casos, como $z \neq 0$, ha de ser $\lambda \neq 0$. ♦

Problema 447 *En $\mathbb{R}^2$ tomamos un producto ω no estándar de matriz ω y dos matrices b_i que son isometrías para ω. los datos son*

$$b_1 = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 1 & 0 \\ -1 & -1 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Clasificar las isometrías $\Lambda_i(x) = t + b_i(x)$ siendo $t = (1,-3)^\top$ y dar sus ecuaciones normalizadas.

Problema 448 *Consideramos dimensión* 2. *Decir si en* alguna *circunstancia pueden ser ciertas las afirmaciones siguientes, donde las rotaciones y traslaciones se suponen que* no son id .

1. *La composición de dos rotaciones puede ser una traslación.*
2. *La composición de una simetría y una rotación es otra simetría.*
3. *Si Δ es un deslizamiento con simetría, Δ^2 es una rotación.*

Problema 449 *En $\mathbb{R}^2$ con el producto estándar se considera la colección de isometrías*

$$\Lambda_h\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} h \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Probar que todas son rotaciones, calcular sus centros y determinar si las ecuaciones normalizadas son iguales. (Si esto último tuviese respuesta afirmativa, se deduciría que todas las Λ_h "hacen esencialmente lo mismo".) ♦

Solución. Visto cómo es L y que tiene determinante 1, todas son rotaciones. También es inmediato que $\mathrm{id} - L$ es un isomorfismo, luego siempre $z = 0$ y todas las Λ_h tienen la misma ecuación canónica

$$\Lambda\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

Las coordenadas (u, v) no son las naturales y tienen como origen de la referencia $\mathcal{R}_h$ el único punto fijo, el centro. Se calcula como la solución de $(\mathrm{id} - L)(x) = t$; en concreto

$$\begin{pmatrix} h \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \qquad c = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{-1}\begin{pmatrix} h \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} h-1 \\ h+1 \end{pmatrix}.$$

Para todas las $\mathcal{R}_h$ sirve una misma base ortogonal positiva. Lo más sencillo es elegir la base estándar. ♦

Problema 450 *Se pide la ecuación normalizada de la isometría Λ de $\mathbb{R}^3$*

$$\Lambda\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2/3 & -1/3 & 2/3 \\ -1/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Nota: le vamos a dar algunos datos al lector para ahorrarle el trabajo monótono. Identificando L con la matriz 3×3 decimos que $\det(L) = -1$ y que $(-1,1,0)^\top$ y $(2,0,1)^\top$ están fijos por L.

Los problemas que siguen pretenden evitar un cálculo matricial de espacios propios y resolución de sistemas pues ya hemos hecho muchos. Los cálculos son algo conceptuales pero muy rápidos.

Problema 451 *Sea $\Lambda(x) = t + L(x)$ una rotosimetría y $v \neq 0$. ¿Qué tipo de isometría es $\Delta = T_v \circ \Lambda$?*

Problema 452 *En $\mathbb{E}$ de dimensión arbitraria n nos dan dos vectores unitarios u y v que son ortogonales, y $S(x) = x - 2\langle x, u\rangle u$ la simetria lineal que fija el plano ortogonal a u. Decir qué tipos de isometría son $\Lambda_1(x) = u + S(x)$ y $\Lambda_2(x) = v + S(x)$ con la misma S para ambas.*

Problema 453 *Suponemos $\dim(\mathbb{E}) = 3$. Nos dan dos vectores v_1 y v_2 unitarios, independientes y ortogonales a un tercer vector unitario v_3, siendo $\langle v_1, v_2\rangle = 1/2$ (por consiguiente, $\mathcal{V} = (v_1, v_2, v_3)$ no es base ortonormal, pero "le falta poco"). Consideramos las simetrías vectoriales $S_i(x) = x - 2\langle x, v_i\rangle v_i$, $i = 1, 2$ y tenemos la función lineal $L(x) = S_1 \circ S_2(x)$. Clasificar las isometrias afines*

$$\Lambda_i(x) = v_i + L(x), \quad i = 1, 2, 3. \blacklozenge$$

Solución primera. El primer procedimiento es un puro cálculo. Se tiene en general, para $\mathrm{id} - L$,

$$L(x) = S_1(x - 2\langle x, v_2\rangle v_2) = (x - 2\langle x, v_2\rangle v_2) - 2\langle (x - 2\langle x, v_2\rangle v_2), v_1\rangle v_1 = x - 2\langle x, v_2\rangle v_2 - 2\langle x, v_1\rangle v_1,$$

$$(\mathrm{id} - L)(x) = x - L(x) = 2\langle x, v_2\rangle v_2 + 2\langle x, v_1\rangle v_1, \quad \ker(\mathrm{id} - L) = \lg(v_3), \quad \mathrm{im}(\mathrm{id} - L) = \lg(v_1, v_2).$$

El vector z_i de cada Λ_i es la proyección de v_i sobre $\ker(\mathrm{id} - L) = \lg(v_3)$ por lo que $z_1 = z_2 = 0$ y $z_3 = v_3$. Para tener los puntos fijos para $i = 1, 2$ hay que resolver $v_i = (\mathrm{id} - L)(x) = 2\langle x, v_2\rangle v_2 + 2\langle x, v_1\rangle v_1$. Claramente $p_i = \frac{1}{2}v_i$ es un punto fijo, y el conjunto de todos los puntos fijos es $\mathbf{D}_i = p_i + \mathbb{D}$ siendo $\mathbb{D}$ el espacio de soluciones de $2\langle x, v_2\rangle v_2 + 2\langle x, v_1\rangle v_1$, que es $\lg(v_3)$. Así pues, Λ_1 y Λ_2 son rotaciones.

Como $v_3 \neq 0$, Λ_3 no tiene puntos fijos pero su parte lineal L tiene a $\mathbb{D} = \lg(v_3)$ como espacio de puntos fijos. L es pues una rotación de eje $\mathbb{D}$ y $\Lambda_3 = T_{v_3} \circ L$, que es un movimiento helicoidal.

Solución segunda. Se trata de factorizar. L, al ser composición de dos simetrías, es una rotación de eje $\mathbb{D}$ ortogonal a v_1 y v_2, luego debe ser $\mathbb{D} = \lg(v_3)$. Factorizamos $T_{v_1} = \Sigma_1' \circ \Sigma_2'$ siendo Σ_2' una simetría especular lineal cuyo plano contiene a $\mathbb{D}$. Pero L se puede factorizar como composición de dos simetrías con planos conteniendo T_{v_1}a $\mathbb{D}$, con libre elección de una de ellas, por lo que elegimos $L = \Sigma_2' \circ \Sigma_3'$ y $\Lambda_1 = T_{v_1} \circ L = \Sigma_1' \circ \Sigma_2' \circ \Sigma_2' \circ \Sigma_3' = \Sigma_1' \circ \Sigma_3'$ y Λ_1 es una rotación. El caso de Λ_2 es análogo y el caso de Λ_3 es como en la solución de más arriba. $\blacklozenge$

Problema 454 *Sean (p_1, q_1) y (p_2, q_2) dos pares de puntos de un plano euclidiano $\mathbb{E}$ tales que $\|p_1 - q_1\| = \|p_2 - q_2\|$. Probar que hay exactamente dos isometrías Λ tales que $\Lambda(p_1) = p_2$ y $\Lambda(q_1) = q_2$, siendo una rotación o traslación y la otra una simetría o deslizamiento con simetría.*

Una parte muy interesante, pero que no vamos a tratar, consiste en tomar un subconjunto F de $\mathbb{R}^2$ (o de $\mathbb{E}$ bidimensional si se quiere generalizar) y considerar el conjunto $\mathcal{F}$ de las isometrías afines Λ que cumplen $\Lambda(F) = F$, admitiendo que $x \in F$ puede ser movido, pero sin salir de F. La idea es estudiar $\mathcal{F}$, que es un grupo con la operación de composición. Casi siempre, cuando se quiere comparar una identidad de isometrías $\Lambda_1 = \Lambda_2$ de $\mathcal{F}$ sabiendo, por ejemplo, que ambas son rotaciones, se elige un segmento $\mathcal{S}$ con extremos en F que ha de ir a otro $\mathcal{S}'$ con extremos en F también, y si Λ_1 y Λ_2 tienen el mismo efecto en el segmento, serán iguales por el problema 454. Este trabajo con isometrías, que deja de lado los pesados cálculos con matrices, es muy atractivo y la tentación de dejar de lado la fundamentación es grande e incluso necesaria. Si un día lo descubre el lector (estamos pensando en la clasificación de grupos cristalográficos) esperamos que nos disculpe por hacerle calcular tanto. Hay no obstante problemas que no se resuelven con un puro algoritmo.

Como ejemplo vamos a mostrar como se calcula visualmente el centro, si existe, de una composición de rotaciones con diferentes centros. Es esencial expresar las rotaciones Λ_1 y Λ_2 como producto de simetrías respecto de rectas, una de las cuales, une los centros c_1 y c_2 de Λ_1 y Λ_2. Sea $\Lambda_3 = \Lambda_1 \circ \Lambda_2$ y c_3 su centro. Factorizamos $\Lambda_1 = \Sigma_1 \circ \Sigma$ y $\Lambda_2 = \Sigma \circ \Sigma_2$, siendo Σ la simetría respecto de la recta que une c_1 y c_2. Claramente, $\Lambda_3 = \Lambda_1 \circ \Lambda_2 = \Sigma_1 \circ \Sigma \circ \Sigma \circ \Sigma_2 = \Sigma_1 \circ \Sigma_2$. Aquí usamos algo esencial, y es que c_3 es el único punto fijo de Λ_3. Si las rectas fijas (ejes) de Σ_1 y Σ_2 no son paralelas, se cortan en un punto, que ha de ser el centro c_3. Si los ejes no se cortasen, serían paralelos y Λ_3 sería una traslación de dirección ortogonal a los ejes.

Problema 455 *Si tenemos dos rotaciones Λ y Λ' que se relacionan con la simetría Σ por $\Lambda \circ \Sigma = \Sigma \circ \Lambda'$, probar que sus centros c y c′ son simétricos respecto al eje de Σ. (Fácil. Es jugar con una ecuación.) Utilizar esto para mostrar que si con rotaciones Λ_1 y Λ_2 tales que $\Lambda_1 \circ \Lambda_2 \neq \mathrm{id}$ construimos nuevas rotaciones $\Lambda_3 = \Lambda_1 \circ \Lambda_2$ y $\Lambda_4 = \Lambda_1^{-1} \circ \Lambda_2^{-1}$ (¡ojo!, no decimos $\Lambda_4 = \Lambda_2^{-1} \circ \Lambda_1^{-1}$), los centros c_3 u c_4 son simétricos respecto de la recta $\mathbf{D}$ que pasa por c_1 y c_2.*

La utilidad de esto es que si en una figura F detectamos dos rotaciones de centros c_1 y c_2 que preservan F, tras calcular c_3 como se explicó antes del problema, se puede encontrar un nuevo centro de simetría rotacional c_4, que será el simétrico de c_3 respecto de la recta que une c_1 y c_2.

El teorema que sigue es laborioso de probar, pero si se admite, da cómo plantearse muchos problemas puramente numéricos, cuyas limitaciones ya hemos comentado, pero que hay que saber hacer. Se explica tras el enunciado.

Supongamos que se tiene $\Lambda(x) = t + L(x)$ con vector de desplazamiento z, y otra isometría $\Gamma(x) = r + G(x)$. Consideramos una nueva isometría $\Theta = \Gamma \circ \Lambda \circ \Gamma^{-1}$ y nos preguntamos cuál es el vector de desplazamiento de Θ.

Teorema 227 *Con las notaciones anteriores, $G(z)$ es el vector de desplazamiento de $\Theta = \Gamma \circ \Lambda \circ \Gamma^{-1}$*

La demostración requiere cálculos pesados, pero si se admite el teorema se ven enseguida sus consecuencias. La principal es que, como G es una isometría lineal, se verifica que $\|G(z)\| = \|z\|$ y deducimos que *los módulos de desplazamiento de $\Theta = \Gamma \circ \Lambda \circ \Gamma^{-1}$ y Λ son iguales.* Si se buscan problemas de tipo estrictamente numérico, se puede elegir Λ en forma normalizada, tomar como Γ otra isometría, y calcular $\Theta = \Gamma \circ \Lambda \circ \Gamma^{-1}$. Si el problema es hallar la ecuación normalizada de Θ, sabemos que al tener el mismo módulo de desplazamiento y polinomio característico, esta ecuación es la de Λ.

El lector decidirá si quiere ver la demostración del teorema 227.

Demostración. Al ser $\Theta = \Gamma \circ \Lambda \circ \Gamma^{-1}$, la parte lineal M de Θ es $M = G \circ L \circ G^{-1}$. Abreviaremos $\bar{L} = \mathrm{id} - L$ y $\bar{M} = \mathrm{id} - M$. Observemos primero que con $\bar{L} = \mathrm{id} - L$,

$$\bar{M} = \mathrm{id} - G \circ L \circ G^{-1} = G \circ \mathrm{id} \circ G^{-1} - G \circ L \circ G^{-1} = G \circ \bar{L} \circ G^{-1}$$

Se tiene pues $\bar{M} = \mathrm{id} - M = G \circ \bar{L} \circ G^{-1}$. Lo siguiente es calcular

$$\begin{aligned}\Gamma \circ \Lambda \circ \Gamma^{-1}(x) &= \Gamma \circ \Lambda \circ \left(-G^{-1}(r) + G^{-1}(x)\right) = \Gamma\left(t + L\left(-G^{-1}(r) + G^{-1}(x)\right)\right)\\ &= \Gamma\left(t - L \circ G^{-1}(r) + L \circ G^{-1}(x)\right) = r + G\left(t - L \circ G^{-1}(r) + L \circ G^{-1}(x)\right)\\ &= G(t) + \left(r - G \circ L \circ G^{-1}(r)\right) + G \circ L \circ G^{-1}(x) = G(t) + \bar{M}(r) + \bar{M}(x)\end{aligned}$$

El vector de traslación de Θ es $w = G(t) + \bar{M}(r)$. Descomponemos $\mathbb{E} = \mathrm{im}\left(\bar{M}\right) \oplus \ker\left(\bar{M}\right)$ y el vector de desplazamiento v de Θ será la componente de w en $\ker\left(\bar{M}\right)$, que es la que tenga $G(t)$. Descomponemos $t = \bar{L}(p) + z$ y

$$G(t) = G\left(\bar{L}(p) + z\right) = G \circ \bar{L} \circ G^{-1}(G(p)) + G(z) = \bar{M}(G(p)) + G(z)$$

y es obvio que la componente de w y $G(t)$ en $\ker\left(\bar{M}\right)$ es la de $G(z)$. Pero $\bar{M} \circ G(z) = G \circ \bar{L}(z) = 0$ porque por construcción $z \in \ker\left(\bar{L}\right)$. En definitiva, el vector de desplazamiento de Θ es la componente en $\ker\left(\bar{M}\right)$ de $w = G(t) + \bar{M}(r)$, que se ha probado que es $G(z)$ como queríamos. ♣

Para concluir ofrecemos al lector que verifique las tablas de clasificación de las isometrías para $n = 2, 3$. Son un puro recetario pero sirven como repaso de conceptos. *Es un gran problema final (sin número).*

Fijamos una isometría $\Lambda = t + L(x)$ o $\Lambda = T_z \circ \Gamma$. Hay diversos criterios para clasificar las isometrías hilando más o menos fino. Todos usan el que Λ determina de modo unívoco varios entes matemáticos, y, según sean estos, podemos ir clasificando. Estos entes son, sin analizar de momento su relevancia:

1. La parte lineal L, y como consecuencia, cualquier otro objeto determinado por L como la traza $\mathrm{tr}(L)$, determinante $\det(L)$, polinomio característico $C(X)$, y subespacios propios, incluida su dimensión.

2. La traslación de vector t,[16] y sobre todo su descomposición $t = (\mathrm{id} - L)(p) + z$ en $\mathbb{E} = \mathrm{im}(\mathrm{id} - L) \oplus \ker(\mathrm{id} - L)$, que permite quizás varios p pero determina sin ambigüedad z y su módulo $\|z\|$. Son el vector y módulo de desplazamiento o deslizamiento.

3. Los puntos fijos de Λ, si existen, son las soluciones de $t = (\mathrm{id} - L)(x)$. El que sea $z \neq 0$ equivale a que Λ no tenga puntos fijos; no obstante, Γ siempre los tiene y son los p que cumplen $t' = (p - L(p))$, siendo t' la proyección de t sobre $\mathrm{im}(\mathrm{id} - L)$ en la descomposición ortogonal $\mathbb{E}$ referida.

[16] Rozamos el peligro. El que usemos el espacio vectorial $\mathbb{E}$ como modelo del espacio afín permite tener $t = \Lambda(0)$ de modo unívoco. Si se estudia el espacio afín general, Λ y el módulo de deslizamiento vienen fijados por Λ, pero no t, y la construcción se complica.

4. Como aunque sea $z \neq 0$, se tiene $(\mathrm{id} - L)(z) = 0$, la traslación T_z mueve en paralelo al espacio de puntos fijos de Γ.

5. La ecuación estándar, normalizada o canónica para una referencia euclidiana, que a veces no es preciso conocer, es de la forma

$$\begin{pmatrix} \Lambda(x)^1 \\ \vdots \\ \Lambda(x)^m \end{pmatrix} = \begin{pmatrix} a_1^1 & \cdots & a_n^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \cdots & a_n^n \end{pmatrix} \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix} + \begin{pmatrix} \mu \\ \vdots \\ 0 \end{pmatrix},$$

siendo a la matriz estándar en una base ortonormal de la isometría lineal L, que es construible con el polinomio característico (teorema 213) y el vector $(\mu, 0, \ldots, 0)^\top$ tiene $\mu = \|z\|$.

Nos vamos a limitar a los casos con $n = 2, 3$. Para $n = 2$ hay dos tipos generales

$$\Lambda_{(1)} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p & -q \\ q & p \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \quad \Lambda_{(2)} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \mu \\ 0 \end{pmatrix},$$

con $p^2 + q^2 = 1$ y $q, \mu \geq 0$. El primer tipo corresponde a rotaciones y traslaciones, según sea $\mu = 0$ o $\mu \neq 0$ y el segundo a simetrías y deslizamientos reflejados, según sea $\mu = 0$ o $\mu \neq 0$. La vía rápida para llegar a las ecuaciones es esta:

1. Las ecuaciones $\Lambda_{(1)}$ corresponden al caso $\det(L) = 1$. Si no es traslación, algo sencillo de comprobar, debe ser $\mu = \|z\| = 0$ porque no tiene puntos fijos, y $p = \frac{1}{2}\operatorname{tr}(L)$, $q = \sqrt{1 - p^2}$.

2. Las ecuaciones $\Lambda_{(2)}$ corresponden al caso $\det(L) = -1$. Como en este caso $\Lambda^2 = T_{2z}$ y la matriz ya se conoce, se aplica, $\Lambda^2(0) = 2z$ y luego $\mu = \|z\|$. Claramente, si Λ es una simetría, $z = 0$.

Las dos tablas que siguen permiten clasificar Λ en estos cuatro tipos de modo más geométrico o algebraico.

$\Lambda(x) = T_z \circ \Gamma$	Γ fija un punto	Γ fija una recta
Con puntos fijos	Rotaciones	Simetrías
Sin puntos fijos	Traslaciones	Deslizamientos reflejados

.

$\Lambda(x) = t + L(x)$	$\det(L) = 1$	$\det(L) = -1$
$\mu = 0$	Rotaciones	Simetrías
$\mu > 0$	Traslaciones	Deslizamientos reflejados

.

Vamos a seguir una vía similar al caso $n = 2$ pero centrándonos en la factorización $\Lambda = T_z \circ \Gamma$. Interesa primero conocer si es o no es $z = 0$ para lo que hay que examinar $(\mathrm{id} - L)(x) = t$. Si tiene soluciones p, ha de ser $z = 0$, $\Lambda = \Gamma$, y quedan fijos los puntos p solución de $(\mathrm{id} - L)(x) = t$. Si p es uno de ellos, podemos escribir, cosa frecuentemente útil, $\Lambda = T_p \circ L \circ T_p^{-1}$. El tipo de isometría es simetría, rotación (incluida id), o rotosimetría, correspondiente a las dimensiones de $\ker(\mathrm{id} - L)$ como expone la tabla más abajo.

Si $(\mathrm{id} - L)(x) = t$ no tiene solución, ha de calcularse z con $t = (\mathrm{id} - L)(p) + z$, aunque esto puede evitarse a veces para la ecuación normalizada. La posibilidad de abreviar aparece si Λ no es una traslación ni un movimiento helicoidal porque se tiene $\Lambda^2 = T_{2z}$ y $\mu = \|z\|$ se calcula como para $n = 2$. El caso del movimiento helicoidal se distingue porque hay que calcular p y z (es el caso más laborioso) y entonces $\mathbf{D} = p + \lg(z)$ es el eje del movimiento.

El recetario para una clasificación algebraica es

$\Lambda = T_d \circ \Phi$	$\dim \ker(\mathrm{id}_{\mathbb{E}} - L)$	nombre	raíces de $C(X)$
Con puntos fijos	3	identidad	$(1, 1, 1)$
Con puntos fijos	2	simetría (especular)	$(1, 1, -1)$
Con puntos fijos	1	rotación (no identidad)	$(1, p \pm qi)$, $q > 0$
Con puntos fijos	0	rotosimetría	$(-1, p \pm qi)$, $q \geq 0$
Sin puntos fijos	3	traslación	$(1, 1, 1)$
Sin puntos fijos	2	deslizamiento con simetría	$(1, 1, -1)$
Sin puntos fijos	1	movimiento helicoidal	$(1, p \pm qi)$, $q > 0$