1 Comentarios sobre las clases de MA–460

Esta es una versión en pdf de los comentarios en mensajes de texto que acompañaron las clases "virtuales" de MA–460, *Álgebra Lineal II*, en el primer semestre del 2020, a causa de la suspensión de clases presenciales.

1.1. Clase 1

Esta fue una clase presencial (hasta ahora, la única), el 10 de marzo del 2020. Lo que sigue es un corto resumen de esa clase.

Espacios vectoriales

Se dio un repaso a los conceptos de combinación lineal, independencia lineal, subespacios generados y bases.

Un resultado importante es el Escolio 1.9: cualquier base parcial puede ser prolongado a una base de todo el espacio vectorial, al agregar más vectores a los originales.

Aplicaciones lineales

Se repasó el concepto de aplicación lineal entre dos espacios vectoriales V y W; ellas forman un nuevo espacio vectorial $\mathcal{L}(V, W)$. Si W es el cuerpo $\mathbb{F}$ de escalares, $\mathcal{L}(V, \mathbb{F}) = V^*$ es el llamado **espacio dual** de V. A cada base de n vectores de V, le corresponde una base dual de n formas lineales en V^*.

Se definió la transpuesta de $S \in \mathcal{L}(V, W)$ como una aplicación $S^{\mathrm{t}} \in \mathcal{L}(W^*, V^*)$. Se habló de la **nulidad** de T, la dimensión $n(T)$ del núcleo de T, y del **rango** de T, la dimensión $r(T)$ del espacio imagen $T(V)$. Finalmente, se vio cómo la prolongación de bases establece *el teorema* importante *de rango y nulidad*:

$$n(T) + r(T) = n = \dim V.$$

1.2. Clase 2

Esta fue la primera clase virtual, el 13 de marzo del 2020. Se revisaron las secciones § 1.3: *Matrices* y § 1.4: *Ecuaciones lineales y eliminación gaussiana*.

Matrices

Una matriz A es un rectángulo $(m \times n)$ con entradas a_{ij} en $\mathbb{F}$. Cada *columna* $\boldsymbol{a}_j$ puede ser considerado como en vector en $\mathbb{F}^m$. En cambios, las filas $\boldsymbol{a}^i$ (p. 1-8) se consideran como formas lineales sobre $\mathbb{F}^n$ y así se forma la matriz producto $C = AB$ cuyas entradas son productos punto de una fila de A por una columna de B: fórmula (1.7).

Cada aplicación lineal $T\colon V \to W$ da lugar a una matriz, como sigue: primero se elige una *base* $\{\boldsymbol{x}_j\}$ del primer espacio vectorial V, y luego otra *base* $\{\boldsymbol{y}_i\}$ del segundo espacio W. Ahora, T está determinado por sus valores $T(\boldsymbol{x}_j)$ en la base de V, por linealidad (Escolio 1.14); ese vector en W tiene una expansión única en términos de las base $\{\boldsymbol{y}_i\}$ de W. Así, se llega a la "fórmula mágica":

$$T(\boldsymbol{x}_j) = \sum_{i=1}^{m} a_{ij}\,\boldsymbol{y}_i \tag{1.9}$$

donde se nota que los índices al lado derecho están "al revés": i, j, i.

La importancia de la fórmula (1.9) es que a cada aplicación lineal T le corresponde una sola matriz A, siempre y cuando se hayan fijado dos bases, una base $\mathcal{B}$ para V y otra base $\mathcal{C}$ para W.

Además, cada vector $\boldsymbol{x} \in V$ determina un vector de columna $\boldsymbol{c} \in \mathbb{F}^n$ (sus coeficientes en la base $\mathcal{B}$) y el vector $T\boldsymbol{x} \in W$ determina una columna $\boldsymbol{b} \in \mathbb{F}^m$ (sus coeficientes en la base $\mathcal{C}$). La fórmula (1.9) se transmuta en una multiplicación matriz-por-vector, $\boldsymbol{b} = A\boldsymbol{c}$. Eso es lo que dice la fórmula (1.10). Esa correspondencia conserva las operaciones algebraicas; en particular, la matriz de la *aplicación transpuesta* $T^{\mathrm{t}}\colon W^* \to V^*$ es la *matriz transpuesta* A^{t}.

Usamos T_A para referirnos a la aplicación $\boldsymbol{x} \mapsto A\boldsymbol{x}$, de $\mathbb{F}^n$ a $\mathbb{F}^m$. La imagen $T_A(\mathbb{F}^n)$, la cual es un subespacio de $\mathbb{F}^m$, es una combinación lineal de las *columnas* de la matriz A (que son vectores columna en $\mathbb{F}^m$). Esta es la fórmula clave (1.11). Entonces, el **rango** $r(A)$ de la matriz A, que es la dimensión del espacio imagen, es el número máximo de columnas de A que son linealmente independientes.

Si cambiamos las bases $\mathcal{B}$ de V y $\mathcal{C}$ de W, a otras bases $\mathcal{B}'$ de V y $\mathcal{C}'$ de W, las expansiones de unas en otras dan lugar a dos matrices P y Q, según la fórmula (1.12). Estas matrices son invertibles, porque siempre se puede volver a las bases originales. La nueva matriz de T, en las bases nuevas, no es A sino

$$B = QAP$$

un producto de tres matrices (P y Q son matrices cuadradas). Pero el subespacio imagen de $\mathbb{F}^m$ es la misma: entonces $r(B) = r(A)$. En otras palabras: la operación $A \mapsto QAP$ no cambia el rango.

Para el caso donde $V = W$ (un **operador lineal** T de V en sí mismo), la matriz A es cuadrada y se puede tomar Q como el inverso de P, escrito $Q = P^{-1}$. El cambio de matriz es $B = P^{-1}AP$ en este caso, y se dice que las matrices A y B son **semejantes**.

Tópico para el capítulo 2: dadas dos matrices cuadradas A y B, ¿cómo se puede saber si son semejantes o no?

Ecuaciones lineales y eliminación gaussiana

Las matrices A y los vectores columna $\boldsymbol{b}$ aparecen en los sistemas de ecuaciones "lineales" (esto es, de primer grado). Las incógnitan $x_1, \dots, x_n$ se organizan en otro vector de columna $\boldsymbol{x}$ y el sistema toma la forma $A\boldsymbol{x} = \boldsymbol{b}$: fórmulas (1.13a) y (1.13b).

Pero también se puede ver el sistema como una combinación lineal de las columnas de A, con coeficientes x_i, puesto igual a $\boldsymbol{b}$: esto es la fórmula (1.13c). Todo esto para decir que $\boldsymbol{b}$ queda en el espacio imagen de T_A, generado por las *columnas* de A.

Para la resolución práctica del sistema, se usan las tres famosas operaciones de fila, para cambiar (si es posible) la matriz A en una forma triangular. Si eso se logra, es fácil despejar la última ecuación; y en seguida, la penúltima; y así, de tras para adelante, se despejan las entradas del vector $\boldsymbol{x}$.

Es útil organizar el cálculo con la *matriz aumentada* $[A \mid \boldsymbol{b}]$, pegando $\boldsymbol{b}$ a la matriz A como una columna extra.

La Prop. 1.34 dice que cada operación de fila se efectúa por "premultiplicación" (a la izquierda) por una cierta matriz invertible. Si no se usa operaciones del tipo (c) – intercambio de filas – el producto L de esas matrices es *triangular inferior*, mientras la matriz original A se convierte en una matriz *triangular superior* V. En fin, al final se descubre la factorización $A = LV$. El ejemplo 1.35 explora ese proceso en detalle en un caso (3×3).

(En cambio, si es necesario cambiar filas, se obtiene $A = PLV$, donde la matriz P acumula todas los cambios de fila en una sola matriz.)

Prop. 1.36: En el caso homogéneo $A\boldsymbol{x} = \boldsymbol{0}$, el criterio para obtener una solución única es $n(A) = 0$. Por el teorema de rango y nulidad, eso es equivalente a $r(A) = n$. En el caso inhomogeneo, el criterio es $r([A \mid \boldsymbol{b}]) = r(A) = n$.

Una forma "práctica" de determinar el rango $r(A)$ es convertir A, con operaciones de fila, en la llamada **forma escalonada** (Defn 1.37). En esa forma, aparecen como columnas ciertos vectores de la base estándar: así se ve cuántas columnas linealmente independientes hay (esto es, la cantidad de columnas estándar), y ese es el rango.

Finalemente, si se usan *operaciones de columna* también (Prop. 1.41), eso se hace por "posmultiplicación" (a la derecha) por ciertas matrices invertibles; y así se puede "limpiar" la matriz original completamente para llegar a una matriz de bloques QAP con una copia de 1_k (la matriz identidad $k \times k$) en la esquina superior izquierda. Hecho eso, se concluye que el rango es k.

Sobre la primera lista de ejercicios, llamada "Tarea 1"

Ejercicio 1.8 Esto depende solamente de la definición de la aplicación transpuesta y del subespacio imagen. Es más fácil de lo que parece.

Ejercicio 1.9 La integral indefinida es un ejemplo de una aplicación lineal. ¿Por qué no es sobreyectiva? Piense en el "teorema fundamental del cálculo".

Ejercicio 1.11 Las matrices A y B están en "forma normal de Jordan", que veremos después. Es útil calcular sus potencias.

1.3. Clase 3

Esta fue la segunda clase virtual, el 17 de marzo del 2020. Se revisó la sección § 1.5: *Determinantes*.

Determinantes

Un sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$, fórmula (1.13b), tiene solución única

$$\boldsymbol{x} = A^{-1}\boldsymbol{b}$$

si la matriz de coeficientes A tiene una matriz inversa A^{-1}. Esto ocurre:

- ⋄ si y solo si la operación $T_A : \boldsymbol{x} \mapsto A\boldsymbol{x}$ es biyectiva;
- ⋄ si y solo si el rango de la matriz A es n (si A es una matriz $n \times n$);
- ⋄ si y solo si A tiene n columnas linealmente independientes;
- ⋄ si y solo si el proceso de eliminación gaussiana funciona.

Para estimar si todas estas condiciones se cumplen, se calcula un *número* (escalar), $\det A$, sabiendo que A^{-1} existe si y solo si $\det A \neq 0$.

Para matrices 2×2, se usa la fórmula (1.15):

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \qquad \det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Entonces el inverso de A está dado por la fórmula (1.16), que se debe memorizar.

Para matrices 3×3, hay una fórmula (1.17a) más complicada, que se puede calcular por *expansión en la primera fila*. (Se supone que esto ha sido visto en MA–360 y en otros cursos.) Los determinantes 2×2 que entran en esta fórmula se llaman **menores** de la matriz A. Así se llega a la fórmula (1.17b).

Para matrices más generales, $n \times n$, tenemos la Definición 1.42. Esto dice que la expansión con menores se puede hacer con cualquier fila, (1.18a) y (1.18b); o en cualquier columna, (1.18c).

¿Por qué estas expansiones siempre dan el mismo número $\det A$? Respuesta: si se sigue la expansión hasta llegar al determinante de A, se obtiene una fórmula maestra, la fórmula (1.19), donde la suma tiene $n!$ términos, etiquetados por todas las permutaciones de $(1, 2, \dots, n)$.

De la fórmula de Leibniz (1.19), se obtiene varias consecuencias. Una de ellas es la fórmula multiplicativa, Prop. 1.44:

$$\det(AB) = (\det A)(\det B).$$

Otra es la Prop. 1.46: $\det A = \det A^{\mathrm{t}}$, esto es: la transpuesta A^{t} tiene el mismo determinante que A (se cambian columnas por filas).

Otra consecuncia de la expansión por filas es que el determinante de una matriz *triangular* (Prop. 1.47) es fácil de calcular: es simplemente el producto de las entradas diagonales.

Igual de importante es la Prop. 1.45: el determinante de A *se puede calcular por eliminación gaussiana* (cuando esa funciona). Eso por cuanto las matrices (1.14 a,b,c) que efectúan las operaciones de fila tienen determinantes fáciles de calcular, que luego se multiplican [por la fórmula multiplicativa]. Es de notar que la operación de fila más usada (restar de una fila un múltiplo de otra) *no cambia el determinante*; y que un intercambio de filas solo cambia el signo del determinante.

La Prop. 1.49 habla del *mecanismo* de usar eliminación gaussiana para calcular determinantes: se trata de poner la matriz en forma triangular mediante las operaciones de fila de tipos (b) y (c) solamente, llevando la cuanta de la cantidad de intercambios de fila [tipo (c)] que se hace. En una etapa intermedia, la matriz tiene un aspecto de bloques:

$$A' = \begin{bmatrix} U' & X' \\ 0 & Y' \end{bmatrix}$$

donde el primer bloque U' ya es triangular y las filas debajo de U' ya están "limpias", presentando un bloque de ceros. El bloque Y' queda aún por procesar. Las entradas diagonales de U' son los llamados **pivotes** del proceso. Si todo termina bien, entonces $\det A$ es el producto de los pivotes y un (-1) por cada intercambio de filas: esto es la fórmula final (1.21).

Pero el proceso de eliminación gaussiana puede fallar, si la matriz A no era invertible: esto pasa cuando un pivote resulta ser igual a 0 y no queda más que detener el algoritmo. En este caso se concluye que $\det A = 0$ (porque el rango de A es menor que n).

Los determinantes también ayudan para calcular el rango de una matriz, que ahora puede ser rectangular. Se forman todas los menores $k \times k$, determinantes de las submatrices $k \times k$ presentes en A. Si al menos uno de ellos no es cero, el rango de A es mayor o igual que k; si todos son ceros, el rango de A es menor que k. Entonces (Prop. 1.50) el rango de A es el mayor tamaño k con un menor $k \times k$ no nulo.

Hay una fórmula (no muy útil) para expresar las soluciones del sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$ (cuando $\det A \neq 0$), llamado regla de Crámer. Se forma la matriz de **cofactores** de A, así:

(a) se calculan todos los menores m_{ij} como una matriz M;

(b) la transpuesta de M es la matriz M^{t} con entradas m_{ji};

(c) se cambian los signos de M^{t}, $(-1)^{i+j} m_{ji}$, según un "tablero de ajedrez" con casillas blancas, signos $+1$, en la diagonal principal; y casillas negras con signos -1.

La matriz de cofactores se llama $\operatorname{adj} A$, la **matriz adjugada**. La Prop. 1.52 dice que este producto de matrices da el determinante:

$$A\,(\operatorname{adj} A) = (\operatorname{adj} A)\,A = (\det A)\,1_n$$

donde 1_n es la matriz identidad $n \times n$.

Si $\det A \neq 0$, el inverso A^{-1} es la matriz $(\operatorname{adj} A)$ dividido por el determiante, fórmula (1.24). En el caso 2×2, esto da la fórmula anterior para A^{-1} que se pidió memorizar.

Esta manera de calcular el inverso no es práctica para matrices grandes, 45×45 por ejemplo, porque exige calcular $45^2 = 2025$ menores, cada uno de tamaño 44×44, y luego hacer una combinación lineal de algunos de ellos (expansión en una fila) para llegar a $\det A$. Sin embargo, sí sirve para justificar la regla de Crámer, fórmula (1.25), para expresar la solución única (si hubiera) de la ecuación $A\boldsymbol{x} = \boldsymbol{b}$, como cocientes de determinantes.

Dicho todo eso, la matriz adjugada tiene su utilidad, como se verá en algunos de los ejercicios.

Sobre la segunda lista de ejercicios, llamada "Tarea 2"

Ejercicio 1.13 Si no es obvio cómo "demostrar que $A^n = 0$", haga un ensayo con una matriz 4×4 de ese tipo, calculando A^2, A^3, A^4, para ver qué es lo que pasa con el patrón de los ceros.

Ejercicio 1.15 Un típico ejercicio de eliminación gaussiana. Fíjese que hay 4 ecuaciones para 5 variables, así que de antemano no se espera solución única. Pero hay que procesarlo para ver qué pasa.

Ejercicio 1.17 El determinante de Vandermonde (un contemporáneo de Lagrange) es muy enredado si se hace por expansión en filas (o columnas), pero es más fácil por eliminación gaussiana.

Ejercicio 1.18 Otro que es alucinante por expansión en filas, pero sale por eliminacion gaussiana.

Ejercicio 1.19 Teórico, pero susceptible a eliminacion gaussiana también.

Ejercicios 1.22 y 1.23 Ejemplos de cómo se puede emplear la matriz adjugada para conseguir resultados teóricos.

Ejercicio 1.24 Un poco de geometría analítica con determinantes. Aquí lo que rige es la circunstancia de que una matriz con dos filas iguales tiene determinante cero.

2 Comentarios sobre las clases de MA–460, parte 2

Esta es una versión en pdf de los comentarios en mensajes de texto que acompañaron las clases "virtuales" de MA–460 del 2020, *Álgebra Lineal II*, en el primer semestre, a causa de la suspensión de clases presenciales.

2.1. Clase 4

Esta es la clase virtual del 14 de abril *Álgebra Lineal II*, en el primer semestre , después de Semana Santa. La temática es la sección § 2.1: *Autovalores y autovectores*.

Autovalores y autovectores

El tema de este capítulo es la estructura de una aplicación lineal T de un espacio vectorial V en sí mismo, que también se llama un **operador lineal** sobre V. Dada una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, al operador T le corresponde una **matriz cuadrada**, A.

La Defn. 2.1 es vocabulario: la totalidad $\mathcal{L}(V)$ de tales operadores T cumple las mismas reglas algebraicas que el juego $M_n(\mathbb{F})$ de matrices $n \times n$. Por ahora, $\mathbb{F}$ es de tipo general (en la práctica, $\mathbb{R}$ o $\mathbb{C}$); más adelante, vamos a limitarnos a los escalares complejos, en $\mathbb{C}$.

Hay ciertos números λ en $\mathbb{F}$ y ciertos vectores $\boldsymbol{x} \neq \mathbf{0}$ en V que cumplen una ecuación de tipo (2.1):

$$T(\boldsymbol{x}) = \lambda\, \boldsymbol{x}. \tag{2.1}$$

En esos casos, λ se llama un **autovalor** (o "valor propio") de T y el vector $\boldsymbol{x}$ se llama un **autovector** (o "vector propio") de T. Es importante notar que $\boldsymbol{x} \neq \mathbf{0}$: el vector nulo cumpliría la ecuación anterior para cualquier λ y no impartiría información alguna.

Dada una base $\mathcal{B}$ de V, la fórmula mágica (1.9) que define la matriz A de T – usando la misma base en ambos lados, ver fórmula (2.2) – permite cambiar la ecuación anterior a $A\boldsymbol{x} = \lambda\, \boldsymbol{x}$ donde ahora $\boldsymbol{x}$ es un vector columna en $\mathbb{F}^n$. De este modo, también se habla de autovalores y autovectores de la matriz cuadrada A.

¿Cuál es la gracia de eso? Se puede escribir (2.1) como $T(\boldsymbol{x}) - \lambda\, \boldsymbol{x} = \mathbf{0}$, o bien:

$$(T - \lambda\, 1)\boldsymbol{x} = \mathbf{0},$$

y similarmente para matrices, $(A - \lambda\, 1_n)\boldsymbol{x} = \mathbf{0}$. Esto dice que $\boldsymbol{x}$ es un vector no nulo en el núcleo de $(T - \lambda\, 1)$, o que $(T - \lambda\, 1)$ no es invertible (Lema 2.4). Del mismo modo, Lema 2.5, la matriz $(A - \lambda\, 1_n)$ no es invertible y su determinante es cero.

Como $\det(A - \lambda\, 1_n) = \pm \det(\lambda\, 1_n - A)$ [el signo es $(-1)^n$] preferimos trabajar con la condición

$$\det(\lambda\, 1_n - A) = 0.$$

El lado izquierdo de esta ecuación es un "polinomio en λ" de grado n. Mejor dicho, si se define un polinomio de grado n con la fórmula (2.3):

$$p_A(t) := \det(t\, 1_n - A), \tag{2.3}$$

entonces las autovalores λ son *las raíces de este polinomio*. Este es el **polinomio característico** de A. Como un polinomio de grado n tiene, a lo sumo, n raíces, no hay más que n posibilidades para los autovalores (sea de la matriz A o del operador T).

Para obtener los autovalores, entonces, es necesario:

(a) escribir la matriz $(t\, 1_n - A)$, una matriz $n \times n$;

(b) calcular su determinante para obtener el polinomio $p_A(t)$;

(c) plantear la ecuación $p_A(t) = 0$;

(d) hallar todos las soluciones $t = \lambda$ de esa ecuación.

Hay un caso especial en el cual este trabajo es fácil; y a la postre, innecesario. Si A es *triangular* (Corolario 2.6), la matriz $(t\, 1_n - A)$ es también triangular, y las raíces de $p_A(t) = 0$ son $\lambda = a_{11}$, $\lambda = a_{22}$, etc. En breve: *si A es triangular, sus autovalores son sus elementos diagonales*.

Desafortunadamente, si A no es triangular, la tarea no es nada fácil. Pero hay al menos una buena noticia: si A y B son matrices semejantes, tienen el mismo polinomio característico (Lema 2.8 y Cor. 2.9); y por ende tienen los mismos autovalores. Esto significa que *un cambio de base no cambia los autovalores*, de manera que se puede hablar del polinomio característico de un operador T (Defn. 2.11).

El Ejemplo 2.13 introduce una mala noticia: hay matrices (no triangulares, obviamente) que no poseen autovalores – al menos en números reales. El malo de la película es una humilde matriz 2×2:

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.5}$$

Su polinomio característico es $p_J(t) = t^2 + 1$; y la ecuación cuadrática $t^2 + 1 = 0$ no tiene soluciones reales. Pero sí tiene soluciones complejas, $t = i$ y $t = -i$. En seguida, se obtienen dos autovectores en $\mathbb{C}^2$, por "bateo".

El Ejemplo 2.14 ofrece un caso más típico, esta vez 3×3. Primero se calcula $p_A(t) = t^3 - 9t$. Luego se factoriza $t^3 - 9t$ para obtener los autovalores $\lambda = 0, 3, -3$.

Para cada uno de esos valores de λ se debe buscar un autovector correspondiente. ¿Cómo proceder? Es cuestión de plantear el sistema $(\lambda\, 1_3 - A)\boldsymbol{x} = \mathbf{0}$ de ecuaciones lineales y resolverlo para $\boldsymbol{x}$; se rechaza la solución trivial $\boldsymbol{x} = \mathbf{0}$ que no puede ser un autovector. En cada caso, al aplicar la eliminación gaussiana a la matriz aumentada, se termina con una fila de ceros, lo cual dice que una de las coordenadas de $\boldsymbol{x}$ es *libre* y se puede encontrar las otras (por sustitución regresiva).

Ahora viene un paso importante, para no decir crucial: se arma una nueva matriz P, *cuyas columnas son los autovectores respectivos*. (Cada autovector viene con un factor de escala, la "variable libre" x_3, que se puede elegir para que las columnas de P sean más bellas.) Si se quiere, se puede calcular la matriz inversa P^{-1} también.

Para terminar el Ejemplo 3.14, se hace otro cálculo: resulta que $AP = PD$, donde D es la *matriz diagonal* cuyas entradas diagonales son los autovalores $\lambda = 0, 3, -3$. Dicho de otro modo, este cálculo establece la fórmula (2.6):

$$P^{-1}AP = D. \tag{2.6}$$

Esta fórmula justifica la búsqueda explícita de los autovectores: la conclusión es que *hay un cierto cambio de base* que convierte la matriz dada A en una matriz diagonal, D. Y la matriz D tiene los mismos autovalores que A: la matriz A es **diagonalizable**.

Eso nos deja con (al menos) dos inquietudes:

(1) P1/: ¿Habrá una *fórmula* para el polinomio característico $p_A(t)$ que no requiere la expansión directa de $\det(t\, 1_n - A)$?

R1/: **Sí**; pero hallar esa fórmula puede ser igual de doloroso. ☹

(2) P2/: ¿Toda matriz A de tamaño $n \times n$ es diagonalizable?

R2/: Por desgracia, **No**; pero se puede buscar un *criterio* de diagonalizabilidad.

Sobre la tercera lista de ejercicios, la "Tarea 3"

Ejercicio 2.3 La matriz del cuadrado mágico tiene un autovector "obvio": ¿sabe usted cuál es?

Ejercicio 2.6 Usar la definición misma de autovalor y autovector de una matriz.

Ejercicio 2.8 Algunos libros escriben la matriz compañera como la transpuesta de la matriz A; esto es, con los $-a_j$ en la última columna y las entradas 1 justo debajo de la diagonal principal. De todos modos, el polinomio característico $p_A(t)$ es el mismo.

2.2. Clase 5

Esta es la clase virtual del 17 de abril *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 2.1: *Autovalores y autovectores*, segunda parte.

Autovalores y autovectores (continuación)

Ahora se busca una *fórmula* para el polinomio característico de la matriz A, esto una expresión del polinomio

$$p_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$$

en la cual cada coeficiente c_k es una función explícita de los elementos de la matriz A.

Se escribe $c_{n-k} = (-1)^k \tau_k(A)$, con signos alternantes – esto es el efecto del signo menos en el desarrollo de $\det(t\, 1_n - A)$. Cada $\tau_k(A)$ es una suma de varios términos que son las llamadas **menores principales** de tamaño $k \times k$ de la matriz A. El primero de ellos es la **traza** de A, la suma de los elementos diagonales:

$$\operatorname{tr} A := a_{11} + \cdots + a_{nn}.$$

El último es el *determinante*, $\det A$. Los términos intermedios son más complicadas, pero pocas veces es necesario calcularlos desde el principio. (Para que los apuntes sean completos, se ofrece una demostración de la Prop. 2.16, cuya lectura es opcional.)

Como una cambio de base no altera el polinomio característico, estos coeficientes son invariantes bajo semejanza $A \mapsto P^{-1}AP$. En efecto, se sabe que

$$\begin{aligned} \operatorname{tr}(P^{-1}AP) &= \operatorname{tr}(APP^{-1}) = \operatorname{tr} A, \qquad \text{y} \\ \det(P^{-1}AP) &= (\det P)^{-1}(\det A)(\det P) = \det A. \end{aligned}$$

Solo por curiosidad, consideremos las menores principales 2×2 de una matriz 3×3. Dada la matriz

$$A := \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

hay tres submatrices principales (sus diagonales son porciones de la diagonal de A):

$$A_{12,12} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_{13,13} = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix}, \quad A_{23,23} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

La suma de los determinantes de estas matrices es el coeficiente

$$\tau_2(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}.$$

Para la matriz del Ejemplo 2.14, dada por

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

se ve que $\operatorname{tr} A = 1 - 1 + 0 = 0$; $\det A = (0 + 0 + 0) - (-4 + 4 + 0) = 0$; y además

$$\tau_2(A) = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 2 & 0 \end{vmatrix} + \begin{vmatrix} -1 & -2 \\ -2 & 0 \end{vmatrix} = (-1) + (-4) + (-4) = -9.$$

Entonces $p_A(t) = t^3 - 0\,t^2 + (-9)t - 0 = t^3 - 9t$. Esto coincide del desarrollo directo de $\det(t\,1_3 - A)$, como debería ser.

Moraleja: el cálculo de los coeficientes de $p_A(t)$ por la fórmula (2.7) no es muy práctico, salvo el segundo, $-\operatorname{tr} A$, y el último, $\pm \det A$. Pero hay una importante excepción: en el caso de la matriz triangular, donde los autovalores coinciden con los elementos diagonales, *todos los $\tau_k(A)$ son funciones explícitas de los autovalores*: esto es la fórmula (2.9).

Veremos un poco más adelante (en el apartado § 2.5) que toda matriz compleja es semejante a una matriz triangular. Entonces, si $\mathbb{F} = \mathbb{C}$, la fórmula (2.9) es válida para toda matriz en $M_n(\mathbb{C})$.

Al final de esta sección, se ofrece (sin demostración ni obligación) una fórmula para calcular determinantes, llamado *desarrollo de Laplace*. Es una generalización de la conocida expansión por filas y *consiste en tomar varias filas a la vez* para la expansión. Por ejemplo, para calcular un determinante 5×5, se puede expandir en las primeras dos filas, $I = \{1, 2\}$. Hay $\binom{5}{2} = 10$ maneras de elegir un par de columnas, J. Entonces la sumatoria (2.10) tendrá 10 términos, con una regla complicada para asignar los signos $(-1)^{s(I,J)}$. Con $I' = \{3, 4, 5\}$ y J' las columnas restantes, cada $\det A_{I'J'}$ es un determinante 3×3. Esta regla puede ser práctica si muchas entradas de A son ceros.

2.3. Clase 6

Esta es la clase virtual del 21 de abril *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 2.2: *El teorema de Cayley y Hamilton*.

El teorema de Cayley y Hamilton

El referido teorema, cuyo enunciado formal es el Teorema 2.26, dice que cualquier matriz A satisface su propia ecuación característica. Para poder apreciar el significado de esa afirmación, es necesario revisar primero la manipulación de polinomios.

Dos polinomios $f(t)$, $g(t)$, con coeficientes en un cuerpo $\mathbb{F}$ cualquiera, se pueden sumar, restar y multiplicar. El asunto que requiere cautela es la *división*. Pasa algo parecido con los números enteros: 28 divide 84, pero 35 no es un divisor de 84. Lo que sucede es que $84 = 2 \times 35 + 14$; en breve, esta es una *división con residuo*. (Como 14 es menor que 35, no hay más que hacer.)

Ahora, 84 y 35 tienen divisor común 7, y de hecho, 7 es su *máximo común divisor*: $\underline{\mathrm{mcd}}(84, 35) = 7$. Esto es obvio para quienes no han olvidado sus tablas de multiplicación, pero hay una forma más sistemática de hallar el mcd:

$$84 = 2 \times 35 + 14, \qquad 35 = 2 \times 14 + 7, \qquad 14 = 2 \times 7 + 0.$$

En cada paso después del primero, se toma el divisor y residuo anteriores como el dividendo y divisor actual, hasta que se llega al residuo 0; en cuyo caso, el penúltimo residuo es el mcd buscado. (Esta afirmación es la Proposición VII.2 de los *Elementos* de Euclides, y por eso se conoce como el *algoritmo euclidiano*.)

Otro ejemplo: $\mathrm{mcd}(2304, 1296) = 144$. En efecto,

$$\begin{aligned} 2304 &= 1 \times 1296 + 1008, & 1008 &= 3 \times 288 + 144, \\ 1296 &= 1 \times 1008 + 288, & 288 &= 2 \times 144 + 0. \end{aligned}$$

El algoritmo euclidiano también es aplicable a polinomios, con dos diferencias: la medida de tamaño de un polinomio es su *grado*, y el mcd debe ser un polinomio *mónico* (primer coeficiente igual a 1). La división con residuo se ve el la fórmula (2.11), donde el grado del residuo $r(t)$ debe ser *menor* que el grado del divisor $g(t)$, o bien $r(t) = 0$. (El polinomio nulo no tiene grado.)

Quienes han usado "fracciones parciales" para calcular integrales recordarán que el residuo de $f(t)$, al dividirlo por $(t - a)$, es el número $f(a)$: Lema 2.20. Como resultado, $(t - a)$ es un divisor de $f(t)$ – sin residuo – si y solo si $f(a) = 0$.

La Defn. 2.22 expresa formalmente que el mcd de dos polinomios debe ser (i) un divisor común; (ii) divisible par cualquier otro divisor común; y (iii) mónico. Con estos tres requisitos, $\mathrm{mcd}(f(t), g(t))$ siempre existe y es *único*. [Eso se demuestra al aplicar el algoritmo euclidiano: véase la fórmula (2.12).]

Si se persigue los detalles del algoritmo euclidiano, se llega a una fórmula importante (2.13): el máximo común divisor $\mathrm{mcd}(f(t), g(t))$ es una combinación – *no* lineal – de $f(t)$ y $g(t)$. De los polinomios "coeficientes" $a(t)$ y $b(t)$ que entran en esta fórmula, no sabemos nada de antemano – ni queremos saber nada, basta con que existen.

Una caso importante, Corolario 2.24, es cuando $f(t)$, $g(t)$ son *relativamente primos*, es decir, no tienen un factor común no trivial, es cuyo caso su mcd es el polinomio constante 1 [esto es el polinomio mónico de grado 0]. En ese caso, la fórmula de Bézout dice que hay otros polinomios $a(t)$, $b(t)$ tales que

$$a(t)\,f(t) + b(t)\,g(t) = 1.$$

► Después de ese repaso rápido del álgebra de polinomios, llegamos a lo que nos interesa: se puede formar nuevas matrices [o aplicaciones lineales] al *enchufar* una matriz cuadrada A [o un operador lineal T] en un polinomio cualquiera: véase las fórmulas (2.14). El único detalle notable es que al término constante c_0 del polinomio le corresponde la matriz $c_0\,1_n$ [o el operador $c_0\,1_V$]. Esta **evaluación** del polinomio $f(t)$ en A, para obtener otra matriz $f(A)$, respeta sumas y multiplicaciones.

Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada cualquiera, su polinomio característico es $p_A(t) := \det(t\,1_n - A) = t^n + c_{n-1}t^{n-1} + \cdots + c_1t + c_0$, como ya sabemos. Su **ecuación característica** es la ecuación $\underline{p_A(t) = 0}$, cuyas soluciones son los autovalores de A.

El teorema de Cayley y Hamilton ahora dice que A satisface esta ecuación. Es decir, si se evalúa el lado izquierdo para obtener la matriz $p_A(A)$, y el lado derecho la obtener la matriz nula 0, la igualdad se mantiene: $\underline{p_A(A) = 0}$.

La demostración, que no nos debe detener mucho rato, es una meditación sobre la fórmula (2.15) de la regla de Crámer:

$$(t\,1_n - A)\ \mathrm{adj}(t\,1_n - A) = p_A(t)\,1_n\,. \tag{2.15}$$

Se expande $\mathrm{adj}(t\,1_n - A)$ en un polinomio en t con coeficientes $B_{n-1}, \dots, B_1, B_0$. Un astuto truco telescópico con estos coeficientes termina con la fórmula final:

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1A + c_0\,1_n = 0.$$

Este es el teorema de Cayley y Hamilton.

► Acabamos de ver que una matriz $n\times n$ cumple una ecuación polinomial de grado n. Pero es posible, para algunas matrices A, que se cumple otra ecuación de grado menor. Entre todas las posibilidades para polinomios $f(t)$ tales que $f(A) = 0$, hay uno (y solo uno) que es mónico y de grado mínimo: este es el **polinomio mínimo**, $q_A(t)$.

Las consideraciones anteriores sobre división de polinomios con residuo llevan rápidamente a la conclusión de que $q_A(t)$ divide $p_A(t)$ – sin residuo. (Esta es la Prop. 2.28.) Un procedimiento práctico para hallar $q_A(t)$ es el siguiente:

(a) calcular el polinomio característico, $p_A(t)$,

(b) factorizar $p_A(t)$ y listar los polinomios mónicos que son divisores de $p_A(t)$;

(c) entre estos divisores, encontrar el $q_A(t)$ de menor grado tal que $q_A(A) = 0$.

El Ejemplo 2.31 ofrece una caso particular de este proceso.

Sobre la cuarta lista de ejercicios, la "Tarea 4"

Ejercicio 2.9 Es fácil calcular una potencia D^k de una matriz diagonal D. Para calcular potencias de una matriz no diagonal A, el primer paso es diagonalizar A (si resulta factible hacerlo).

Ejercicio 2.10 Si $f(t)$ es un polinomio, se obtiene la matriz $f(A)$ por sustituciones $t^k \mapsto A^k$ para cada k (y además $t^0 \mapsto 1_n$). Estas sustituciones también funcionan cuando $f(t)$ es una serie de potencias convergente.

Ejercicio 2.11 La fórmula para A^k se demuestra por inducción sobre k.

Ejercicio 2.13 El resultado $p_{AB}(t) = p_{BA}(t)$ se demuestra sin dificultad si A o B es invertible. En cambio, si $\det A = \det B = 0$, es necesario usar una prueba más sofisticada.

Ejercicios 2.14 y 2.15 En cada caso, el polinomio característico se calcula directamente de su definición. En seguida, se busca el polinomio mínimo entre los divisores del polinomio característico.

2.4. Clase 7

Esta es la clase virtual del 24 de abril *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 2.3: *Matrices diagonalizables* y el inicio de la sección § 2.4: *Descomposición primaria*.

Matrices diagonalizables

En el Ejemplo 2.14, cuyo objeto era la búsqueda de autovalores y autovectores de una cierta matriz A de tamaño 3×3, se construyó otra matriz invertible P y se mostró que $D := P^{-1}AP$ es una *matriz diagonal*. De hecho, las entradas diagonales de la matriz D son precisamente los tres autovalores de A; y las columnas de P son unos autovectores correspondientes. Además, las columnas de P (vistas como vectores en $\mathbb{R}^3$) son linealmente independientes; en virtud de eso, forman una *base* de $\mathbb{R}^3$. En consecuencia, el cambio $A \mapsto P^{-1}AP$ efectúa un *cambio de base* en $\mathbb{R}^3$ – de la base estándar $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ a la base $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$ de las columnas de P.

La independencia lineal de las columnas de P no es casual: significa que el rango de P es máximo, $r(P) = 3$; y por ende la matriz P es invertible, de manera que la expresión $P^{-1}AP$ tiene sentido. (Se debe recordar que el rango de una matriz es el número de columnas independientes que posee.) La propiedad clave de la matriz original A, entonces, es que posee $n = 3$ autovectores linealmente independientes.

Una matriz $n \times n$ se llama **diagonalizable** si es semejante a una matriz diagonal (por cambio de base). Antes de abordar criterios para decidir si una matriz es diagonalizable o no, vale la pena meditar por qué las matrices diagonales son especiales. Consideremos el cálculo (correcto):

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 0 \\ 0 \end{bmatrix}.$$

Si D es esta matriz 4×4, el cálculo se puede abreviar como $\underline{D\boldsymbol{e}_2 = 5\boldsymbol{e}_2}$. De igual manera, es fácil ver que:

$$D\boldsymbol{e}_1 = 2\boldsymbol{e}_1, \quad D\boldsymbol{e}_2 = 5\boldsymbol{e}_2, \quad D\boldsymbol{e}_3 = 4\boldsymbol{e}_3, \quad D\boldsymbol{e}_4 = 8\boldsymbol{e}_4 .$$

Esto dice simplemente que *los vectores de la base estándar son autovectores de la matriz diagonal* D. (Además, las entradas diagonales de D son sus autovalores.)

He aquí, entonces, una manera de enfocar el problema de diagonalizabilidad: una matriz $n \times n$ es diagonalizable si y solo si posee n autovectores linealmente independientes. El dilema, sin embargo, es: ¿cómo garantizar esa condición?

El Lema 2.32 da la primera pista: si $\{\lambda_1, \dots, \lambda_k\}$ son autovalores *distintos* de A, entonces los autovectores correspondientes son linealmente independientes.

La Prop. 2.33 saca la consecuencia: si una matriz $n \times n$ tiene *n autovalores distintos*, entonces $\mathbb{F}^n$ posee una base de sus autovectores. Al organizar esos autovectores como columnas de una matriz P, la cual es invertible, se puede chequear – con la fórmula (2.18) – que $AP = PD$; y por ende $P^{-1}AP = D$ es diagonal.

La Prop. 2.34 resume la discusión anterior.

► Es suficiente, pero *no es necesario*, que los autovalores sean distintos. Es importante tener una notación adecuada para los casos de autovalores repetidos. Se expone la situación general para una matriz diagonal D en la página 2.18. Se denota por $\{\alpha_1, \dots, \alpha_r\}$ el conjunto de autovalores distintos. En la diagonal de D la entrada α_1 viene repetida k_1 veces, la entrada α_2 viene repetida k_2 veces, etcétera. (Esas entradas pueden ocurrir en cualquier orden.) Entonces el *polinomio característico* de D se factoriza así, habida cuenta de que la matriz $(t\,1_n - D)$ es también diagonal:

$$p_D(t) = (t - \alpha_1)^{k_1}(t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r}.$$

En cambio, el *polinomio mínimo* no tiene factores repetidos:

$$q_D(t) = (t - \alpha_1)(t - \alpha_2) \cdots (t - \alpha_r). \tag{2.20}$$

Descomposición primaria

En general, el polinomio mínimo de una matriz A puede tener factores repetidos, como se ve en el Ejemplo 2.31. Ya es hora de mirar más de cerca a la relación entre los polinomios característico y mínimo.

Conviene regresar a los operadores lineales $T\colon V \to V$. Un concepto importante es el de un **subespacio invariante** para T; este es un subespacio $U \leqslant V$ tal que $T(U) \subseteq U$. (Los subespacios triviales $\{0\}$ y V mismo son invariantes; pero ese hecho no ayuda mucho.)

Supóngase, entonces, que $\dim V = n$ y $\dim U = m$ con $0 < m < n$. Tómese una base cualquiera $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ de U; como sabemos, ésta se puede prolongar en una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V. Con este tipo de base, el operador T tiene una "matriz en bloques", véase la fórmula (2.21):

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} A & X \\ 0 & B \end{bmatrix}. \tag{2.21}$$

Si $\boldsymbol{u} = a_1\boldsymbol{x}_1 + \cdots + a_m\boldsymbol{x}_m \in U$ y si $\boldsymbol{w} = a_{m+1}\boldsymbol{x}_{m+1} + \cdots + a_n\boldsymbol{x}_n$, se ve que

$$\begin{bmatrix} A & X \\ 0 & B \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} A\boldsymbol{u} + X\boldsymbol{w} \\ B\boldsymbol{w} \end{bmatrix},$$

y en el caso $\boldsymbol{w} = \mathbf{0}$ se obtiene $T\left(\begin{bmatrix} \boldsymbol{u} \\ \mathbf{0} \end{bmatrix}\right) = \begin{bmatrix} A\boldsymbol{u} \\ \mathbf{0} \end{bmatrix}$, mostrando que $T(U) \subseteq U$.

Si el bloque rectangular X fuera nulo, $X = 0$, se tendría además $T\left(\begin{bmatrix} \mathbf{0} \\ \boldsymbol{w} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{0} \\ B\boldsymbol{w} \end{bmatrix}$. En tal caso, T tiene un segundo subespacio invariante, el cual es

$$W := \operatorname{lin}\langle \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \rangle$$

Fíjese que los subespacios U y W son *suplementarios*: $\underline{V = U \oplus W}$. En este caso, el subespacio U **reduce** el operador lineal T (y W también lo reduce, desde luego). La matriz de T ahora es una **suma directa** de dos bloques cuadradas A y B, así:

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}. \tag{2.22}$$

► Debe ser obvio que una reducción a dos bloques, como ese (2.22), es el primer paso hacia una posible diagonalización de la matriz de T. Lo interesante es que esa reducción viene de una factorización del polinomio característico. Según la Prop. 2.38, si $\underline{p_T(t) = h(t)\,k(t)}$ se factoriza en dos polinomios *relativamente primos*, entonces se pueden formar dos operadores $h(T)$ y $k(T)$ (sobre V) al sustituir el operador T en los dos polinomios $h(t)$ y $k(t)$. Los núcleos de esos operadores son dos subespacios invariantes, $\underline{U := \ker h(T)}$ y $\underline{W := \ker k(T)}$ que reducen T, por ser $V = U \oplus W$.

2.5. Clase 8

Esta es la clase virtual del 5 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 2.4: *Descomposición primaria*.

Descomposición primaria: factorización de $p_T(t)$

Considérese la situación de la Prop. 2.38: hay un operador lineal $T : V \to V$ donde $\dim V = n$. Se busca un par de subespacios U y W suplementarios que reducen T. Esto quiere decir que $T(U) \subseteq U$, $T(W) \subseteq W$ y $V = U \oplus W$. Esto es posible si el polinomio característico de T es el producto de dos polinomios (mónicos) sin factor común: $p_T(t) = h(t)\,k(t) = k(t)\,h(t)$.

Se puede formar otros operadores sobre V al sustituir T en esos polinomios:

$$\underline{p_T(T) = h(T)\,k(T) = k(T)\,h(T)}.$$

La *prueba* de la Prop. 2.38 tiene cierta interés. En primer lugar, se usa la identidad de Bézout para polinomios para encontrar otras dos polinomios $a(t)$, $b(t)$ – no importa cuáles son – tales que

$$h(t)\,a(t) + k(t)\,b(t) = 1.$$

Si $\boldsymbol{x} \in V$, esto lleva a una ecuación entre vectores:

$$k(T)(b(T)\boldsymbol{x}) + h(T)(a(T)\boldsymbol{x}) = \boldsymbol{x}.$$

Si ahora se *define* U como el espacio imagen de $k(T)$ y W como el espacio imagen de $h(T)$, entonces $k(T)(b(T)\boldsymbol{x}) \in U$ y $h(T)(a(T)\boldsymbol{x}) \in W$. Se sigue que $\boldsymbol{x} \in U + W$ para cualquier $\boldsymbol{x} \in V$; o sea, que $V = U + W$.

Para ver que $U + W$ es una *suma directa*, es necesario mostrar que $U \cap W = \{0\}$. Ahí es donde entra el teorema de Cayley y Hamilton. Si $\boldsymbol{x} \in U \cap W$, hay un vector $\boldsymbol{y}$ tal que $\boldsymbol{x} = k(T)\boldsymbol{y} \in \operatorname{im} k(T) = U$, y otro vector $\boldsymbol{z}$ tal que $\boldsymbol{x} = h(T)\boldsymbol{z} \in \operatorname{im} h(T) = W$. Pero ahora el teorema muestra que

$$h(T)\boldsymbol{x} = h(T)k(T)\boldsymbol{y} = p_T(T)\boldsymbol{y} = \mathbf{0} \quad \text{y} \quad k(T)\boldsymbol{x} = k(T)h(T)\boldsymbol{z} = p_T(T)\boldsymbol{z} = \mathbf{0}.$$

Otra vez, con la identidad de Bézout, se reconstruye $\boldsymbol{x}$:

$$\boldsymbol{x} = a(T)(h(T)\boldsymbol{x}) + b(T)(k(T)\boldsymbol{x}) = a(T)(\mathbf{0}) + b(T)(\mathbf{0}) = \mathbf{0}.$$

(Cualquier operador lineal manda $\mathbf{0}$ en $\mathbf{0}$.) Se ha mostrado que $U \cap W = \{\mathbf{0}\}$ y por lo tanto $V = U \oplus W$.

Es fácil ver que U y W son espacios invariantes: por ejemplo, si $\boldsymbol{u} = k(T)\boldsymbol{x} \in U$, entonces $T\boldsymbol{u} = T(k(T)\boldsymbol{x}) = k(T)(T\boldsymbol{x}) \in U$ porque $t\,k(t) = k(t)\,t$ como polinomios. Del mismo modo, se obtiene $T(W) \subseteq W$.

Se ha logrado la reducción de T por U y W. Pero falta *mejorar la definición* de estos subespacios. Esto requiere el teorema de rango y nulidad! Nótese que $\boldsymbol{u} = k(T)\boldsymbol{x} \in U$ implica $h(T)\boldsymbol{u} = h(T)\,k(T)\boldsymbol{x} = p_T(T)\boldsymbol{x} = \boldsymbol{0}$ – Cayley y Hamilton de nuevo – así que $\boldsymbol{u} \in \ker h(T)$; en breve, $U \subseteq \ker h(T)$. Y del mismo modo, $W \subseteq \ker k(T)$. Pero ahora, al hacer un conteo de dimensiones, se obtiene

$$\dim U = n - \dim W = n - r(h(T)) = n(h(T)) = \dim(\ker h(T)),$$

así que nada sobra: $U = \ker h(T)$. Y también, $W = \ker k(T)$.

Eso termina la demostración de la Prop. 2.38: los subespacios U y W que reducen T son los *núcleos* de los operadores $h(T)$ y $k(T)$ obtenidos al factorizar $p_T(t)$. □

Descomposición primaria: reducción a dos bloques

Siempre que $p_T(t) = h(t)\,k(t)$ sea un producto de dos factores mónicos sin factor común, se puede expresar la matriz de T en dos bloques A y B, fórmula (2.22), y se obtiene

$$p_T(t) = \det \begin{bmatrix} t\,1_r - A & 0 \\ 0 & t\,1_s - B \end{bmatrix} = \det(t\,1_r - A)\det(t\,1_s - B) = p_A(t)\,p_B(t).$$

Es posible chequear que $p_A(t)$ y $p_B(t)$ son los polinomios característicos de $h(T)$ sobre U y $k(T)$ sobre W – esto es compatible con el teorema de Cayley y Hamilton porque $U = \ker h(T)$ y $W = \ker k(T)$.

Ese chequeo requiere examinar las raíces individuales λ_j de $p_T(t)$. Para tal efecto, se supondrá que el polinomio $p_t(t)$ *escinde*, esto es, se descompone completamente en factores de primer grado $(t - \lambda_j)$. Al recordar el caso de $t^2 + 1 = (t - i)(t + i)$, se observa la necesidad de trabajar con escalares complejos y no solamente reales.

Ahora no debe sorprender que el polinomio mínimo $q_t(t)$ se descompone de la misma forma: el Lema 2.40 dice que $q_T(t) = r(t)\,s(t)$ donde $r(t)$ es un factor de $h(t)$; y $s(t)$ es un factor de $k(t)$.

En particular (Cor. 2.41), en el caso $h(t) = (t - \lambda)^m$, se sigue que $r(t) = (t - \lambda)^l$ con $l \leqslant m$. Conclusión: *las raíces del polinomio $q_T(t)$* son las mismas raíces de $p_t(T)$, las cuales *son los autovalores de T*, pero quizás con multiplicidades más bajas.

Descomposición primaria: caso general

El Teorema 2.43 expone el caso general. Se factoriza el *polinomio mínimo* (no el polinomio característico) completamente: se expresa $q_T(t)$ como el producto de r polinomios mónicos $h_i(t)$ que no tiene factores comunes. Los subespacios invariantes son $\underline{U_i = \ker h_i(t)}$. Resulta entonces que $V = U_1 \oplus \cdots \oplus U_r$; y la matriz de T en una base conveniente consta de r bloques centradas en la diagonal.

Por ejemplo, si $\mathbb{F} = \mathbb{R}$, $\dim V \geqslant 4$ y si $q_T(t) = t^4 + t^3 - t - 1$, su factorización completa sobre es

$$q_T(t) = t^4 + t^3 - t - 1 = (t-1)(t+1)(t^2+t+1).$$

Este polinomio no escinde sobre $\mathbb{R}$: hay dos factores irreducibles de grado 2. He aquí una matriz real 6×6 con ese polinomio mínimo:

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

Es fácil verificar que su polinomio característico es

$$p_T(t) = p_A(t) = (t-1)^2(t+1)^2(t^2+t+1).$$

⟦ Indicación: el cálculo de $\det(t\,1_6 - A)$ se descompone en 3 bloques 2×2. ⟧

Pasando a números complejos, tómese $\underline{\omega = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})}$ como en el Ejercicio 2.7. Resulta que $\omega^3 = 1$ y además $1 + \omega + \omega^2 = 0$. Así se comprueba que

$$t^2 + t + 1 = (t-\omega)(t-\omega^2).$$

Esto dice que $q_T(t)$ *sí escinde sobre* $\mathbb{C}$:

$$q_T(t) = t^4 + t^3 - t - 1 = (t-1)(t+1)(t-\omega)(t-\omega^2).$$

Los autovalores de A son $(1, 1, -1, -1, \omega, \omega^2)$; dos autovalores repetidos.

En la próxima clase se verá cómo esta factorización del polinomio mínimo conduce a un criterio concreto de diagonalizabilidad.

2.6. Clase 9

Esta es la clase virtual del 8 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 2.4: *Descomposición primaria* y la sección § 2.5: *La forma de Jordan*.

Descomposición primaria: consecuencias

Resumen de la situación hasta ahora: se considera un operador lineal $T\colon V \to V$ cuyo polinomio característico escinde:

$$p_T(t) = (t - \alpha_1)^{k_1}(t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r}. \tag{2.24}$$

Esto sucede siempre *si los escalares son complejos*, $\mathbb{F} = \mathbb{C}$. Aquí $\{\alpha_1, \dots, \alpha_r\}$ son los autovalores *distintos* de T; los enteros $k_1, \dots, k_r$ (cuya suma es n) son sus *multiplicidades*.

Se sabe (Corolario 2.41) que cada α_i es también una raíz del polinomio mínimo, así que cada término $(t - \alpha_i)$ es un factor del polinomio mínimo:

$$q_T(t) = (t - \alpha_1)^{l_1}(t - \alpha_2)^{l_2} \cdots (t - \alpha_r)^{l_r}, \tag{2.25}$$

con $1 \leqslant l_i \leqslant k_i$ en cada caso. Al sustituir $t \mapsto T$, se obtiene unos r operadores cuyos núcleos son los subespacios que reducen T: esto es, $U_i := \ker\bigl((T - \alpha_i\, 1)^{l_i}\bigr)$.

Por fin llegamos al añorado *criterio de diagonalizabilidad* de una matriz A, Prop. 2.45: que el polinomio *mínimo* de A escinde, sin factores repetidos:

$$\boxed{q_A(t) = (t - \alpha_1)(t - \alpha_2) \cdots (t - \alpha_r).}$$

En primer lugar, si esto ocurre, los bloques de A corresponden a espacios de columnas en $\mathbb{F}^n$ dados por $U_i = \ker(T_A - \alpha_i\, 1)$ y en cada bloque se obtiene una matriz escalar $\alpha_i\, 1$. Al volver al polinomio característico $p_A(t)$ de la fórmula (2.24), se nota que ese bloque es de tamaño $k_i \times k_i$. Luego de una cambio de base, T_A tiene una matriz diagonal.

Es importante que $q(t)$ escinda. Un contraejemplo es la matriz real $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, con $p_J(t) = t^2 + 1$ irreducible sobre $\mathbb{R}$. Como $q_J(t)$ divide $p_J(t)$, es obligatoria que $q_J(t) = t^2 + 1$ también. Este polinomio mínimo es irreducible pero de segundo grado.

Sin embargo, si se pasa a números complejos, el polinomio mínimo sí escinde: $p_J(t) = q_J(t) = (t - i)(t + i)$. Los autovalores son distintos y de multiplicidad uno: $\{i, -i\}$ y J es diagonalizable sobre $\mathbb{C}$ con estas entradas diagonales.

La forma de Jordan de una matriz compleja

¿Qué pasa si el polinomio mínimo tiene factores repetidos? Bueno, de todos modos la descomposición primaria permite juntar bases de los subespacios U_i para presentar T como una matriz de bloques:

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_r \end{bmatrix}. \tag{2.26}$$

El polinomio mínimo define cuáles son los bloques (tienen un solo autovalor α_i); el polinomio característico ofrece sus tamaños: el bloque número i es $(k_i \times k_i)$ y sus columnas están contados por $\dim U_i = k_i$.

Si el bloque es diagonal, entonces $A_i = \alpha_i \, 1_{k_i}$ es una "matriz escalar". En cambio, si el bloque *no es* diagonal, hay un sobrante $\underline{N_i} := A_i - \alpha_i \, 1_{k_i}$. Por la mera definición del subespacio U_i, esta matriz es **nilpotente**:

$$N_i^{l_i} = (A_i - \alpha_i \, 1_{k_i})^{l_i} = 0.$$

⟦ En los apuntes, se dice que $N_i^{k_i} = 0$ por el teorema de Cayley y Hamilton. Esa diferencia no es crucial: el polinomio mínimo del bloque es $(t - \alpha_i)^{l_i}$. Lo importante es que *alguna* potencia de N_i se anula. ⟧

Está claro que N_i conmuta con el bloque escalar $\alpha_i \, 1_{k_i}$. Al juntar los bloques escalares, se obtiene una matriz diagonal B; al juntar los sobrantes, se obtiene una matriz nilpotente N tal que $BN = NB$. La Prop. 2.49 formaliza esta esquema: cada matriz cuadrada A es de la forma $\underline{A = B + N}$, con B diagonalizable, N nilpotente y $BN = NB$.

Ahora sigue un argumento largo que culmina con la prueba de la Prop. 2.51: dada una matriz nilpotente $N \in M_n(\mathbb{C})$, *se puede hallar una base $\mathcal{B}$ de V en la cual el operador T_N tiene una matriz triangular*. Y no solo triangular: en la diagonal hay ceros (se debe recordar el Ejercicio 1.13) y en la primera subdiagonal superior, solo aparecen entradas 1 o 0. En efecto, en esta *trigonalización*, el resultado es un juego de bloques $k \times k$ de la siguiente forma (para $k = 5$):

$$J_5(0) := \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.30}$$

La matriz N entonces es *semejante* a una suma directa de este tipo de bloques.

Falta sumarle, al bloque nilpotente, una matriz escalar para llenar la diagonal. Si $\lambda \in \mathbb{C}$ (representando un autovalor α_i de una matriz dada), entonces se forma el siguiente **bloque de Jordan**:

$$\underline{J_5(\lambda)} := \lambda\, 1_5 + J_5(0) = \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}. \tag{2.31}$$

Para este bloque, se puede notar que su polinomio característico es $p(t) = (t - \lambda)^5$: es fácil calcular el determinante de una matriz triangular. Su polinomio mínimo es *también* $q(t) = (t - \lambda)^5$, porque $J_5(0)^5 = 0$ pero $J_5(0)^4 \neq 0$ por cálculo directo.

En contraste, la matriz

$$M := \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

tiene $p_M(t) = (t - \lambda)^5$ pero $q_M(t) = (t - \lambda)^3$. En efecto, se ve que $M = J_3(\lambda) \oplus J_2(\lambda)$ es una suma directa de *dos* bloques; el exponente de $q_M(t)$ está dado por el tamaño del bloque más grande.

El Teorema 2.54 es el sumario final: cualquier matriz $A \in M_n(\mathbb{C})$ es semejante a una suma directa de bloques de Jordan. En la diagonal de esta **forma normal de Jordan** aparecen los autovalores de A; la matriz A es diagonalizable si y solo si todos sus bloques de Jordan son (1×1). ☺

Sobre la quinta lista de ejercicios, la "Tarea 5"

Ejercicio 2.18 Una matriz B es *idempotente* si $B^2 = B$. Sus autovalores son 0 y 1. Pero B no tiene que ser diagonal – o sí?

Ejercicio 2.20 Un *algoritmo* para poner una matriz triangular en su forma de Jordan.

Ejercicios 2.24, 2.25 Hay una *fórmula* para el polinomio mínimo; solo se ve en el viejo libro de Gantmacher. Se basa en el cálculo de menores de polinomios matriciales.

Ejercicio 2.26 Un ejemplo del uso de la fórmula de Gantmacher para calcular $q_A(t)$. Nótese que esta matriz 5×5 tiene 25 menores 4×4, pero varios de ellos son ceros.

3 Comentarios sobre las clases de MA–460, parte 3

Esta es una versión en pdf de los comentarios en mensajes de texto que acompañaron las clases "virtuales" de MA–460 del 2020, *Álgebra Lineal II*, en el primer semestre, a causa de la suspensión de clases presenciales.

3.1. Clase 10

Esta es la clase virtual del 12 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.1: *Productos escalares reales y complejos*.

Productos escalares reales y complejos

En este capítulo se estudiará los fenómenos de álgebra lineal ligados con productos escalares (a veces llamados "productos internos"). Por ese contexto, el cuerpo de escalares siempre será los números reales $\mathbb{R}$ o bien los números complejos $\mathbb{C}$.

Un número complejo es un $\alpha = s + it$ con $s, t \in \mathbb{R}$. El **conjugado complejo** de α es $\bar{\alpha} = s - it$. Nótese que $\alpha \in \mathbb{R}$ si y solo si $\bar{\alpha} = \alpha$.

La Defn. 3.1 introduce el producto escalar $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ de dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$. El ejemplo principal es el **producto punto** en $\mathbb{R}^n$, dado por

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x} \cdot \boldsymbol{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{3.1a}$$

También hay un producto punto en $\mathbb{C}^n$, dado por

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle \equiv \bar{\boldsymbol{z}} \cdot \boldsymbol{w} := \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n. \tag{3.1b}$$

Fíjese que en ambos casos el producto punto es lineal en la *segunda* variable, pero solo *semilineal* (antilineal) en la primera variable. En ambos casos el producto escalar es *positivo*, Defn 3.1(d), en el sentido de que $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{x} = \boldsymbol{0}$.

Otros ejemplos de productos escalares son: (i) la integral de funciones continuas $f : [a, b] \to \mathbb{C}$, dado por

$$\langle f, g \rangle := \int_a^b \overline{f(t)}\, g(t)\, dt.$$

También hay (ii) un producto escalar de matrices rectangulares: si $A, B \in \mathbb{R}^{m \times n}$, se define $\langle A, B \rangle := \operatorname{tr}(A^{\mathsf{t}} B)$.

Con esos ejemplos, queda claro que vale la pena estudiar productos escalares en general, sin limitarse a productos punto.

Un concepto importante es la **norma** o **longitud** de un vector, dado por

$$\|x\| := \sqrt{\langle x, x\rangle}. \tag{3.2}$$

La Prop. 3.9 señala las propiedades generales compartidas por esa y por otras normas posibles:

$$\|\alpha x\| = |\alpha|\,\|x\|, \qquad \|x + y\| \leqslant \|x\| + \|y\|, \qquad \|x\| \geqslant 0,$$

con $\|x\| = 0$ solo si $x = \mathbf{0}$. Una propiedad que es particular a las normas de tipo (3.2) – y que se usa para probar la desigualdad triangular en este caso – es la llamada **desigualdad de Schwarz**:

$$\big|\langle x, y\rangle\big| \leqslant \|x\|\,\|y\| \quad \text{para todo} \quad x, y \tag{3.3}$$

con igualdad si y solo si x, y son *proporcionales* (uno de ellos es un múltiplo del otro).

La demostración de (3.3) – ver Prop. 3.7 – es interesante. Primero, si $\mathbb{F} = \mathbb{R}$, la expresión cuadrática

$$f(t) := \|x + ty\|^2 = \|y\|^2\,t^2 + 2\langle x, y\rangle\,t + \|x\|^2 \equiv at^2 + bt + c$$

cumple $f(t) \geqslant 0$ para todo $t \in \mathbb{R}$. El *grafo* de la función f es una *parábola* que no corta el eje horizontal más de una vez. Por lo tanto, su discriminante no puede ser positivo: $b^2 - 4ac \leqslant 0$. Al mirar esta desigualdad más de cerca, se obtiene la desigualdad (3.3).

El caso complejo de la desigualdad de Schwarz requiere un pequeño truco, usando la *forma polar* de un número complejo:

$$\alpha = s + it = r\cos\theta + ir\,\mathrm{sen}\,\theta \equiv r\,e^{i\theta},$$

donde $r := |\alpha| \equiv \sqrt{\alpha\bar{\alpha}} \geqslant 0$.

El Lema 3.10 menciona la *ley del paralelogramo*:

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

cuya prueba es un cálculo sencillo con los productos escalares. Ahora bien, es posible estudiar otros tipos de norma (aunque no en este curso) que *no* provienen de un producto escalar, por cuanto no cumplen esta ley del paralelogramo. El nombre se debe a una relación entre las longitudes y diagonales de un paralelogramo con vértices $\mathbf{0}, x, x + y, y$.

En presencia de dicha ley, es posible recuperar el producto escalar de la norma. En el caso complejo, por ejemplo:

$$4\,\langle y, x\rangle = \|x + y\|^2 - \|x - y\|^2 + i\,\|ix + y\|^2 - i\,\|ix - y\|^2. \tag{3.5}$$

Esta se llama una *fórmula de polarización*.

La desigualdad de Schwarz permite definir el **ángulo entre dos vectores** porque implica que hay un único ángulo θ con $0 \leqslant \theta \leqslant \pi$ tal que:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \cos \theta.$$

En particular, $\theta = \pi/2$ si y solo si $\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0}$, en cuyo caso se dice que $\boldsymbol{x}, \boldsymbol{y}$ son vectores **ortogonales**, escrito $\boldsymbol{x} \perp \boldsymbol{y}$.

El Lema 3.14 señala que un conjunto de vectores ortogonales entre sí debe ser *linealmente independiente*: si $\boldsymbol{x}_1 \perp \boldsymbol{x}_j$ para $j = 1, \dots, m$, una combinación lineal nula de esos vectores obedece

$$0 = \langle \boldsymbol{x}_j, \mathbf{0} \rangle = \langle \boldsymbol{x}_j, c_1 \boldsymbol{x}_1 + \cdots + c_m \boldsymbol{x}_m \rangle = \langle \boldsymbol{x}_j, c_j \boldsymbol{x}_j \rangle = c_j \, \|\boldsymbol{x}_j\|^2.$$

Como $\boldsymbol{x}_j \neq \mathbf{0}$ implica $\|\boldsymbol{x}_j\| > 0$, esto dice que cada $c_j = 0$, mostrando la independencia lineal.

La Defn. 3.15 introduce el **complemento ortogonal** $M^\perp$ de un subespacio $M \leqslant V$ al definir:

$$M^\perp := \{\, \boldsymbol{y} \in V : \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \text{ para todo } \boldsymbol{x} \in M \,\},$$

esto es, la totalidad de vectores ortogonales a cada vector en M. No es difícil comprobar que $(M^\perp)^\perp = M$, y es obvio que $M \cap M^\perp = \{\mathbf{0}\}$. De hecho, V es la suma directa (Defn 1.10) de estos dos subespacios: $\underline{M \oplus M^\perp = V}$. Esta se llama una **suma directa ortogonal**.

3.2. Clase 11

Esta es la clase virtual del 15 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.2: *Bases ortonormales*.

Bases ortonormales

Lo que distingue los espacios vectoriales con producto escalar es la existencia de ciertas bases muy cómodas y bien adaptadas a dichos espacios. Una **base ortonormal** (Definición 3.16) así se llama porque (a) **orto-** dos vectores distintos son ortogonales; y (b) **-normal** cada vector es normalizado, por tener norma igual a 1. En resumen:

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \begin{cases} 0 & \text{si } j \neq k, \\ 1 & \text{si } j = k. \end{cases} \tag{3.6}$$

Si $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de un espacio vectorial V, cada vector $\boldsymbol{e}_k$ define una *forma lineal* $f_k \colon V \to \mathbb{C}$ (en el caso complejo) dado por el producto escalar de $\boldsymbol{e}_k$ y otro vector; esto es, $\underline{f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle}$. (Obsérvese que $\boldsymbol{e}_k$ está puesto al lado izquierdo y $\boldsymbol{x}$ al lado derecho, para que esta expresión sea lineal en $\boldsymbol{x}$.)

Concretamente, es $f_k(\boldsymbol{x}) = \alpha_k$, donde $\boldsymbol{x} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n$ es la expansión de $\boldsymbol{x}$ en la base $\mathcal{E}$ de V. Entonces $\mathcal{F} := \{f_1, \dots, f_n\}$ es la base dual del espacio dual V^*, de la misma dimensión n que V. La correspondencia $J\boldsymbol{e}_k := f_k$ entonces se extiende a una biyección $J\colon V \to V^*$. Pero aquí hay un detalle importante: esta J no es lineal, sino **semilineal** (o *antilineal*), porque los múltiplos escalares se transforman en sus conjugados complejos: si $\boldsymbol{y} = \beta_1 \boldsymbol{e}_1 + \cdots + \beta_n \boldsymbol{e}_n \in V$, entonces

$$J\boldsymbol{y} = J(\beta_1 \boldsymbol{e}_1 + \cdots + \beta_n \boldsymbol{e}_n) := \bar{\beta}_1 f_1 + \cdots + \bar{\beta}_n f_n.$$

Aquí $J\boldsymbol{y} \in V^*$ es otra forma lineal; su efecto sobre $\boldsymbol{x}$ es

$$J\boldsymbol{y}(\boldsymbol{x}) = \bar{\beta}_1 f_1(\boldsymbol{x}) + \cdots + \bar{\beta}_n f_n(\boldsymbol{x}) = \bar{\beta}_1 \alpha_1 + \cdots + \bar{\beta}_n \alpha_n = \langle \boldsymbol{y}, \boldsymbol{x} \rangle. \tag{3.7}$$

En breve, $J\boldsymbol{y}$ es la forma lineal $\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x} \rangle$. Después de todo, J no depende de la base ortonormal específica de V.

La biyección semilineal $J\colon V \to V^*$ tiene otros usos. Por ejemplo, al identificar V^* con V mediante esta J, el *anulador* $M^\perp \leqslant V^*$ de un subespacio $M \leqslant V$ (Defn. 1.21) se identifica con el *complemento ortogonal* $M^\perp \leqslant V$ (Defn. 3.15). Este es el Lema 3.17. Entonces la notación $M^\perp$ no es ambigua. Por conteo de dimensiones, ya se puede chequear la afirmación de que $V = M \oplus M^\perp$ (suma directa ortogonal).

Un problema práctico de gran importancia es la *construcción* de una base ortonormal para V. Esto se hace por un proceso inductivo (Prop. 3.18), bautizado el **algoritmo de Gram y Schmidt** – aunque fue usado muchos años antes por Laplace. Tiene dos pasos que se repiten alternadamente: un paso es dividir un vector no nulo por su longitud, para normalizarlo a un vector de norma 1. El otro paso es proyectar un vector sobre el complemento ortogonal de los vectores ya obtenidos.

El algoritmo necesita empezar con una surtido de vectores $\boldsymbol{x}_k$ que son *linealmente independientes*. Esta condición es necesaria para que el algoritmo no se detenga; y es también suficiente si los $\boldsymbol{x}_k$ generan V. Luego, se debe *empezar con una base cualquiera* $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V. Los primeros pasos son:

$$\begin{aligned} & && \boldsymbol{e}_1 := \boldsymbol{x}_1/\|\boldsymbol{x}_1\|; \\ \boldsymbol{y}_2 &:= \boldsymbol{x}_2 - \langle \boldsymbol{e}_1, \boldsymbol{x}_2 \rangle\, \boldsymbol{e}_1, && \boldsymbol{e}_2 := \boldsymbol{y}_2/\|\boldsymbol{y}_2\|; \\ \boldsymbol{y}_3 &:= \boldsymbol{x}_3 - \langle \boldsymbol{e}_1, \boldsymbol{x}_3 \rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_3 \rangle\, \boldsymbol{e}_2, && \boldsymbol{e}_3 := \boldsymbol{y}_3/\|\boldsymbol{y}_3\|. \end{aligned} \tag{3.8}$$

En palabras:

- ◇ se normaliza $\boldsymbol{x}_1$ a $\boldsymbol{e}_1$,
- ◇ se proyecta $\boldsymbol{x}_2$ sobre $\operatorname{lin}\langle \boldsymbol{e}_1 \rangle^\perp$, obteniendo $\boldsymbol{y}_2$,

- ⋄ se normaliza $\boldsymbol{y}_2$ a $\boldsymbol{e}_2$,
- ⋄ se proyecta $\boldsymbol{x}_2$ sobre $\operatorname{lin}\langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle^\perp$, obteniendo $\boldsymbol{y}_3$,
- ⋄ se normaliza $\boldsymbol{y}_3$ a $\boldsymbol{e}_3$,

y se continua así hasta llegar a $\boldsymbol{e}_n$, donde el proceso termina.

En los ejercicios se verán ejemplos de convertir una base inicial $\mathcal{B}$ en una base ortonormal $\mathcal{E}$ con este algoritmo. También permite sacar conclusiones teóricas. El Corolario 3.19 refina el Escolio 1.9 (compleción de una base parcial) a la compleción de una base *ortonormal* parcial. Si $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ son vectores ortogonales y normalizados, pero $m < \dim V$, primero se obtiene una base $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\}$ de V a como haya lugar; después, con el algoritmo se endereza esta base a una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$. (Como se puede sospechar, el algoritmo tiene un efecto trivial sobre un surtido de vectores que ya son ortonormales.)

Sobre la sexta lista de ejercicios, la "Tarea 6"

Ejercicio 3.1 Es un ejemplo concreto de cómo hallar un complemento ortogonal.

Ejercicio 3.5 Aquellos productos escalares que son definidos por integrales dan lugar a familias de *polinomios ortogonales*. El caso más sencillo es el de los *polinomio de Chebyshev*, $T_n(t)$, definidos por la fórmula $T_n(\cos\theta) \equiv \cos n\theta$. (Es un resultado "conocido" de la trigonometría que $\cos n\theta$ es un polinomio en la variable $\cos\theta$.) Para mostrar al ortogonalidad – y para resolver el Ejercicio 2.22 – es útil recordar estas fórmulas trigonométricas:

$$\begin{aligned} \cos\alpha + \cos\beta &= 2\cos\bigl(\tfrac{1}{2}(\alpha+\beta)\bigr)\cos\bigl(\tfrac{1}{2}(\alpha-\beta)\bigr), \\ \operatorname{sen}\alpha + \operatorname{sen}\beta &= 2\operatorname{sen}\bigl(\tfrac{1}{2}(\alpha+\beta)\bigr)\cos\bigl(\tfrac{1}{2}(\alpha-\beta)\bigr). \end{aligned}$$

Ejercicio 3.6 Este ejercicio es "geométricamente obvio" al echar una mirada a la Figura 3.1. Para *demostrarlo* algebraicamente, se recomiendo tomar los cuadrados de las normas en la fórmula.

Ejercicio 3.7 Esta "ley del paralelepípedo" es una extensión de la ley del paralelogramo a tres vectores. Para demostrarlo, *no se debe usar* productos escalares – de todos modos, no ayudan. La Figura 3.2 del paralelepípedo incluye muchos paralelogramos.

Ejercicios 3.8, 3.9 El teorema de Jordan y von Neumann recupera un producto escalar de la ley del paralelogramo. El paso difícil es la *aditividad* del producto escalar en sus dos variables: este es la parte 3.8(a), que requiere el ejercicio anterior.

3.3. Clase 12

Esta es la clase virtual del 19 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.3: *Aplicaciones transpuestas y adjuntas*.

Aplicaciones transpuestas y adjuntas

Una aplicación lineal $T\colon V \to W$ entre dos espacios vectoriales cualesquiera tiene una *aplicación transpuesta* entre los espacios duales, $T^{\mathrm{t}}\colon W^* \to V^*$. Esto fue discutido brevemente en la sección 1.2 (la Defn. 1.18). Si V y W admiten productos escalares, hay biyecciones $J\colon V \to V^*$ y $J'\colon W \to W^*$ que permiten simplificar la presentación de esa aplicación transpuesta.

Es conveniente separar los casos de escalares reales y complejos. En el caso real $\mathbb{F} = \mathbb{R}$, cuando V y W son espacios $\mathbb{R}$-vectoriales y $\langle -,- \rangle$ es un producto escalar real, entonces J y J' también son $\mathbb{R}$-lineales y se puede identificar $T^{\mathrm{t}}\colon W^* \to V^*$ con $J^{-1}T^{\mathrm{t}}J'\colon W \to V$. Entonces *se redefine la transpuesta* de T como la aplicación lineal $\underline{T^{\mathrm{t}}\colon W \to V}$ que cumple la "fórmula mágica":

$$\langle T^{\mathrm{t}}\boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{y}, T\boldsymbol{x} \rangle \quad \text{para todo} \quad \boldsymbol{x} \in V,\ \boldsymbol{y} \in W. \tag{3.10}$$

(Fíjese que el producto escalar al lado izquierdo es el de V, mientras el producto escalar al lado derecho es el de W.)

Como el producto escalar real es simétrico, la igualdad (3.10) también se puede escribir como

$$\langle T\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, T^{\mathrm{t}}\boldsymbol{y} \rangle.$$

En fin: *la T puede pasar de un lado a otro de un producto escalar real, convirtiéndose en T^{t}.*

En el lenguaje de matrices, si $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$, y $A \in \mathbb{R}^{m\times n}$ es la matriz de $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$, su matriz transpuesta es $A^{\mathrm{t}} \in \mathbb{R}^{n\times m}$. El producto escalar relevante es el *producto punto* (en $\mathbb{R}^n$ o en $\mathbb{R}^m$) y la fórmula (3.10) se transforma en

$$A^{\mathrm{t}}\boldsymbol{y} \cdot \boldsymbol{x} = \boldsymbol{y} \cdot A\boldsymbol{x} \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{y} \in \mathbb{R}^m. \tag{3.11}$$

En el *caso complejo* $\mathbb{F} = \mathbb{C}$, cuando V y W son espacios $\mathbb{C}$-vectoriales y $\langle -,- \rangle$ es un producto escalar complejo, ahora las biyecciones J y J' solo son *semilineales*. Ahora la pareja de una aplicación $\mathbb{C}$-lineal $S\colon V \to W$ es la llamada **aplicación adjunta** $\underline{S^*\colon W \to V}$ dada por $S^* := J^{-1}S^{\mathrm{t}}J'$. Esta S^* es también una aplicación $\mathbb{C}$-lineal y cumple una nueva "fórmula mágica":

$$\langle S^*\boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{y}, S\boldsymbol{x} \rangle \quad \text{para todo} \quad \boldsymbol{x} \in V,\ \boldsymbol{y} \in W. \tag{3.12}$$

(Otra vez, el producto escalar al lado izquierdo es el de V, mientras el producto escalar al lado derecho es el de W.)

El producto escalar complejo ya no es simétrico, pero se puede tomar el conjugado complejo de ambos lados de (3.12), llegando a

$$\langle S\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, S^*\boldsymbol{y}\rangle.$$

La S puede pasar de un lado a otro de un producto escalar complejo, convirtiéndose en S^*.

Una consecuencia importante de la fórmula (3.12) es la relación entre las matrices B de S y B^* de S^*. Si $B = [b_{ij}] \in \mathbb{C}^{m\times n}$, entonces $B^* = [\bar{b}_{ji}] \in \mathbb{C}^{n\times m}$. En otras palabras, la **matriz adjunta** B^* es *la transpuesta conjugada* de B. Nótese que $(B^*)^* = B$.

Vocabulario: una matriz cuadrada real $A \in M_n(\mathbb{R})$ es **simétrica** si $\underline{A^t = A}$. Una matriz cuadrada compleja $B \in M_n(\mathbb{C})$ es **hermítica** si $\underline{B^* = B}$.

Ejemplo 1: si $A \in \mathbb{R}^{m\times n}$, las matrices $\underline{A^tA} \in M_n(\mathbb{R})$ y $\underline{AA^t} \in M_m(\mathbb{R})$ son simétricas.

Ejemplo 2: si $B \in \mathbb{C}^{m\times n}$, las matrices $\underline{B^*B} \in M_n(\mathbb{C})$ y $\underline{BB^*} \in M_m(\mathbb{C})$ son hermíticas.

Las propiedades más importantes de las matrices hermíticas (y de las matrices reales simétricas, que de hecho son hermíticas) *son:*

(a) *sus autovalores son reales;*

(b) *los autovectores para autovalores distintos son ortogonales entre sí;*

(c) *cada matriz hermítica es diagonalizable.*

La realidad de los autovalores se ve así: si $\lambda \in \mathbb{C}$ es un autovalor y si $\boldsymbol{x} \neq \boldsymbol{0}$ es un autovector para λ, entonces

$$\lambda\,\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{x}, \lambda\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \lambda\boldsymbol{x}, \boldsymbol{x}\rangle = \bar{\lambda}\,\langle \boldsymbol{x}, \boldsymbol{x}\rangle.$$

Como $\langle \boldsymbol{x}, \boldsymbol{x}\rangle > 0$, esto dice que $\lambda = \bar{\lambda}$, o sea, $\lambda \in \mathbb{R}$. (Un cálculo similar establece la ortogonalidad de los autovectores.)

La prueba de diagonalizabilidad (Prop. 3.26) usa la descomposición primaria del capítulo 2, mejorada con un truco de "inducción al revés" que termina mostrando que todos los bloques de Jordan son 1×1.

Si $A^* = A$, la propiedad (b) dice que su descomposición primaria $\mathbb{C}^n = U_1 \oplus \cdots \oplus U_r$ es una *suma directa ortogonal*. Se puede encontrar una base ortonormal en cada U_k por separado; su unión es entonces una base ortonormal para $\mathbb{C}^n$. Por lo tanto:

(d) *cada matriz hermítica tiene una base ortonormal de autovectores.*

3.4. Clase 13

Esta es la clase virtual del 22 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.4: *Matrices ortogonales, unitarias y positivas*.

Matrices ortogonales, unitarias y positivas

En esta sección se continuará con el desarrollo paralelo de los casos real y complejo. Se trata de identificar tres clases importantes de matrices cuadradas y de resaltar las propiedades especiales de cada clase.

Matrices ortogonales Una matriz real $Q \in M_n(\mathbb{R})$ es **ortogonal** si *sus columnas forman una base ortonormal* de $\mathbb{R}^n$.

Presentación algebraica [Prop. 3.28(a) y Cor. 3.29(a)]: Q es ortogonal si y solo si

$$\underline{Q^{\mathrm{t}}\,Q = 1_n = Q\,Q^{\mathrm{t}}}, \quad \text{esto es:} \quad Q^{-1} = Q^{\mathrm{t}}.$$

En efecto, la entrada (i,j) de la matriz $C = Q^{\mathrm{t}}\,Q$ es

$$c_{ij} = \sum_{k=1}^{n} q_{ki}\,q_{kj} = \langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = [\![i = j]\!].$$

Matrices unitarias Una matriz compleja $U \in M_n(\mathbb{C})$ es **unitaria** si *sus columnas forman una base ortonormal* de $\mathbb{C}^n$.

Presentación algebraica [Prop. 3.28(b) y Cor. 3.29(b)]: U es ortogonal si y solo si

$$\underline{U^*\,U = 1_n = U\,U^*}, \quad \text{esto es:} \quad U^{-1} = U^*.$$

La entrada (i,j) de la matriz $M = U^*U$ es

$$m_{ij} = \sum_{k=1}^{n} \overline{u_{ki}}\,u_{kj} = \langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = [\![i = j]\!].$$

Para seguir, se debe recordar del teorema de rango y nulidad (Tma. 1.23) que $r(T^{\mathrm{t}}) = r(T)$ para una aplicación lineal cualquiera; y para matrices, vale lo mismo: $r(A^{\mathrm{t}}) = r(A)$. $[\![$ El número de filas linealmente independientes coincide con el número de columnas linealmente independientes. $]\!]$ La conjugación compleja tampoco cambia el rango de una matriz: el Lema 3.30 dice que: $r(A) = r(\overline{A}) = r(A^*) = r(A^{\mathrm{t}})$.

Algo más sorprendente es que las matrices cuadradas $A^{\mathrm{t}}A$ y A^*A también comparten ese mismo rango: en particular, $\underline{r(A^*A) = r(A)}$.

Esta es la Prop. 3.31, que se basa en un cálculo llamativo: si $\boldsymbol{x} \in \mathbb{C}^n$, entonces

$$A^*A\boldsymbol{x} = \boldsymbol{0} \implies \langle \boldsymbol{x}, A^*A\boldsymbol{x} \rangle = 0 \implies \langle A\boldsymbol{x}, A\boldsymbol{x} \rangle = 0 \implies A\boldsymbol{x} = \boldsymbol{0}. \tag{3.14}$$

Esto dice que los operadores T_{A^*A} y T_A, ambos con dominio $\mathbb{C}^n$, tiene el mismo núcleo. Al contar dimensiones, tiene la misma nulidad: $n(A^*A) = n(A)$. Por ende, tienen el mismo rango:

$$r(A^*A) = n - n(A^*A) = n - n(A) = r(A).$$

► El cálculo (3.14) evidencia que la combinación de matrices con vectores aporta información valiosa. Vale la pena estudiar una matriz cuadrada A a través de los números $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle$, para todo $\boldsymbol{x} \in \mathbb{C}^n$.

(a) $\underline{A = A^*}$ si y solo si $\underline{\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \in \mathbb{R}}$ para cada $\boldsymbol{x} \in \mathbb{C}^n$. (Esta es la Prop. 3.35.)

(b) A se llama una **matriz positiva** cuando $\underline{\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geqslant 0}$ para todo $\boldsymbol{x}$. (Defn. 3.36.)

(c) Si además $\underline{\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0}$ para $\boldsymbol{x} \neq \boldsymbol{0}$, dícese que A es **positiva definida**.

Ejemplo: si $\underline{A = B^*B}$, entonces $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \langle \boldsymbol{x}, B^*B\boldsymbol{x} \rangle = \langle B\boldsymbol{x}, B\boldsymbol{x} \rangle \geqslant 0$.

Prop. 3.38: *A es definida positiva si y solo si es positiva e invertible*. La demostración pasa por notar que A tiene una base ortonormal de autovectores (por ser hermítica); y que en los dos casos un autovalor nulo $\lambda_i = 0$ está excluido.

Dada una matriz $A \in M_n(\mathbb{R})$, ¿cómo se puede *determinar si es definida positiva o no*? Primero hay que chequear, por simple inspección, que A es simétrica. Hecho eso, es cuestión de factorizarla como $\underline{A = LDU}$ *por eliminación gaussiana simple*. Desde luego, como $A^{\mathsf{t}} = A$, se obtendrá $U = L^{\mathsf{t}}$; y como A debe ser invertible, no se espera intercambios de filas. Todo depende, entonces, de si el factor diagonal D es definida positiva. En resumen: solo hace falta que todos los pivotes cumplen $\underline{a_{kk}^{(k)} > 0}$.

La Prop. 3.39 verifica el algoritmo anterior, con una modificación de la eliminación gaussiana que aprovecha la simetría de A. Después de cada operación de fila, se hace *la operación de columna análoga* en seguida, "limpiando" con ceros no solo debajo de la diagonal, sino también a su derecha. Si un pivote resulta ser negativo o cero, el algoritmo termina sin éxito. De los contrario, se termina con una factorización:

$$A = LDL^{\mathsf{t}} = (LD^{1/2})(D^{1/2}L^{\mathsf{t}}) = CC^{\mathsf{t}},$$

donde $\boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} = \boldsymbol{x}^{\mathsf{t}}CC^{\mathsf{t}}\boldsymbol{x} = \|C^{\mathsf{t}}\boldsymbol{x}\|^2 > 0$ para $\boldsymbol{x} \neq \boldsymbol{0}$. (La matriz C^{t} es invertible.) Esta se llama la **factorización de Cholesky** de la matriz definida positiva A.

Sobre la sétima lista de ejercicios, la "Tarea 7"

Ejercicios 3.10 a 3.13 Diversos ejemplos de matrices ortogonales.

Ejercicio 3.17 Este ejercicio indica la relación entre matrices positivas y formas cuadráticas (un tema del Capítulo 4) en dos variables.

Ejercicio 3.18 Una manera (no muy eficiente, pero a veces útil) de establecer independencia lineal de vectores en $\mathbb{R}^n$ es la positividad de su determinante de Gram. Este ejercicio busca la *justificación* de ese criterio de independencia.

Ejercicio 3.19 En la Prop. 3.39 se aprende cómo verificar la positividad definida de una matriz por eliminación gaussiana (simple). Este ejercicio ofrece una manera alternativa de hacer lo mismo en el espíritu de la regla de Crámer, esto es, al calcular ciertos determinantes.

Ejercicios 3.20 Esta matriz 8×8 fue descubierta en un trabajo de tesis de la UCR hace unos años; había que determinar las condiciones sobre los parámetros b, c para que fuera definida positiva. El reto del ejercicio es buscar el modo más eficiente de hacer esa verificación.

3.5. Clase 14

Esta es la clase virtual del 26 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.5: *Operadores autoadjuntos*.

Operadores autoadjuntos

Las secciones anteriores trataron varias clases de matrices reales y complejas. Es hora de hacer un pequeño cambio de enfoque, para estudiar aplicaciones lineales en vez de matrices.

Ahora vamos a enfocar los espacios vectoriales sobre $\mathbb{C}$, dotados de productos escalares complejos: estos se llaman **espacios de Hilbert**. En esta sección y la que sigue, V y W serán espacios de Hilbert finitodimensionales.

Si $T \in \mathcal{L}(V, W)$ es una aplicación lineal, tiene una aplicación lineal adjunta $T^* \in \mathcal{L}(W, V)$. Un resultado importante (Prop. 3.41) es que el núcleo de uno es el complemento ortogonal de la imagen del otro; así, $\underline{\ker T^* = T(V)^\perp}$ y $\ker T = T^*(W)^\perp$. El primero de ellos se ve de esta manera:

$$T^* \boldsymbol{y} = \mathbf{0} \iff \langle T^* \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \text{ para todo } \boldsymbol{x} \iff \langle \boldsymbol{y}, T\boldsymbol{x} \rangle = 0 \text{ para todo } \boldsymbol{x}.$$

En el caso $V = W$, con T y T^* en $\mathcal{L}(V)$, se dice que T es **autoadjunto** si $\underline{T^* = T}$. Una matriz A de T es hermítica, $A^* = A$.

Para $T \in \mathcal{L}(V)$ cualquiera, se puede expresar $T = T_1 + i\,T_2$ donde T_1, T_2 son autoadjuntos. Las fórmulas son obvias:

$$T_1 = \tfrac{1}{2}(T^* + T), \qquad T_2 = \tfrac{i}{2}(T^* - T). \tag{3.18}$$

Sabemos que una matriz hermítica tiene *autovalores reales*. Sucede lo mismo con los operadores autoadjuntos (Lema 3.44) por el mismo cálculo:

$$\lambda\,\langle x, x\rangle = \langle x, \lambda x\rangle = \langle x, Tx\rangle = \langle Tx, x\rangle = \langle \lambda x, x\rangle = \bar{\lambda}\,\langle x, x\rangle.$$

También es cierto que cada $\langle x, Tx\rangle \in \mathbb{R}$, si T es autoadjunto.

► Una clase de operadores autoadjuntos de enorme importancia son los **proyectores** (algunos los llaman "proyectores ortogonales"). $P \in \mathcal{L}(V)$ es un proyector si es autoadjunto, $P^* = P$, y a la vez *idempotente*, $P^2 = P$. En breve: $\underline{P^2 = P = P^*}$.

Si $Px \in P(V)$, entonces $P(Px) = P^2x = Px$; y si $y \in \ker P$, entonces $Py = \mathbf{0}$.

Pero además $\ker P = \ker P^* = P(V)^\perp$ y $V = P(V) \oplus P(V)^\perp$.

Entonces P actúa como 1 sobre $P(V)$ y actúa como 0 sobre $P(V)^\perp$. En resumen: la descomposición $V = P(V) \oplus P(V)^\perp$ *reduce* P a una matriz de bloques:

$$[P]^{\mathcal{E}}_{\mathcal{E}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

¿Cuántos proyectores tiene $\mathcal{L}(V)$? Sucede que hay un único proyector $\underline{P_M}$ para cada subespacio $M \leqslant V$ tal que $P_M(V) = M$ (y también $\ker P_M = M^\perp$). Esto es la Prop. 3.46: se forma la base ortonormal $\mathcal{E}$ al unir bases ortonormales de M y de $M^\perp$.

► En un espacio de Hilbert, la forma normal de Jordan de una matriz no es el concepto más apropiado, porque no toma en cuenta el producto escalar. En vez de la relación de semejanza $A \sim P^{-1}AP$ por cambios de bases arbitrarias, se prefiere *cambios de bases ortonormales*. Entonces la relación "correcta" es la **semejanza unitaria**, es decir, por la relación $\underline{A \sim U^*AU}$ donde la matriz U es unitaria, $U^{-1} = U^*$.

La Prop. 3.48 y el Cor. 3.49 dicen que para cualquier $A \in M_n(\mathbb{C})$ hay alguna matriz unitaria U tal que U^*AU es *triangular*; o sea, la matriz A es **trigonalizable**.

El triángulo se forma por una cadena de subespacios $R_k \leqslant V$ con $\dim R_k = k$ y $T_A(R_k) \subseteq R_k$. Esto se demuestra por inducción; el paso clave es definir $R_{n-1} := (\mathbb{C}y)^\perp$ donde y es un autovector del operador adjunto T_A^* (que siempre existe).

3.6. Clase 15

Esta es la clase virtual del 29 de mayo *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 3.6: *El teorema espectral y sus consecuencias*.

El teorema espectral

A la luz de la Prop. 3.48 (conocido como el teorema de Schur), ya se sabe que para cualquier operador $T \in \mathcal{L}(V)$ existe una base ortonormal de V en la cual T tiene una matriz triangular. Como corolario, cualquier matriz $A \in M_n(\mathbb{C})$ es unitariamente semejante a una matriz triangular.

Se debe, entonces, buscar las condiciones sobre T o A que garanticen que esta matriz triangular es *diagonal*. Una condición suficiente – pero no necesaria – es que A sea una *matriz hermítica*; porque por la Prop. 4.36 se sabe que una matriz hermítica posee una base ortonormal de autovectores. Si U es una matriz unitaria cuyas columnas son esta base, entonces $U^*AU = D$ es diagonal.

El **teorema espectral** es la versión "operatorial" de esa circunstancia. Viene en dos versiones: la primera (el Teorema 3.51) se refiere a un operador *autoadjunto* $\underline{T = T^*} \in \mathcal{L}(V)$. Si $\{\alpha_1, \dots, \alpha_r\}$ son los autovalores *distintos* de T (que son reales), se obtiene una expansión

$$T = \alpha_1 P_1 + \cdots + \alpha_r P_r, \tag{3.22}$$

donde $P_1, \dots, P_r$ son *proyectores* no nulos tales que

$$P_i P_j = 0 \quad \text{si} \quad i \neq j; \quad \text{y} \quad P_1 + \cdots + P_r = 1_V \,. \tag{3.23}$$

El conjunto de autovalores $\alpha_1, \dots, \alpha_r\}$ se llama el **espectro** de T.

Las condiciones $P_i P_j = 0$ dicen las imágenes $M_j := P_j(V)$ son *subespacios ortogonales* de V (por el Lema 3.47); y la condición $P_1 + \cdots + P_r = 1_V$ dice que la suma directa de estos subespacios es todo V; o sea, $M_1 \oplus \cdots \oplus M_r = V$. En contraste con la "descomposición primaria" del capítulo anterior, esto es una suma directa *ortogonal*.

Para probar el teorema, solo hay que obtener estos subespacios M_j e invocar el (único) proyector P_j tal que $P_j(V) = M_j$, dado por la Prop. 3.46. Como $T = T^*$, se toma una base ortonormal de autovectores para T; se define M_j como el subespacio generado por los autovectores de esta base cuyo autovalor es α_j.

Una primera consecuencia de este teorema es la Prop. 3.53: dos matrices hermíticas $A = A^*$ y $B = B^*$ son *simultáneamente diagonalizables* por una matriz unitaria U si y solo si $\underline{AB = BA}$. Se aplica el teorema espectral al operador T_A; y se comprueba que $T_B(M_j) \subseteq M_j$ para cada j cuando A y B conmutan.

La segunda versión del teorema espectral (Teorema 3.56) es más sofisticado. La conclusión es la misma: el operador T tiene una expansión (3.22) donde $\{\alpha_1, \dots, \alpha_r\}$ son sus autovalores distintos y $P_1, \dots, P_r$ son proyectores asociados. Pero la *hipótesis* es más general: el operador T debe ser **normal**, es decir, T y T^* *conmutan* ($\underline{T^*T = TT^*}$) pero no son necesariamente iguales. Por ejemplo, una matriz unitaria U es normal, $U^*U = 1_n = UU^*$, pero rara vez es autoadjunto. En esta circunstancia, los autovalores α_j de T no son reales necesariamente.

La tarea esencial de la prueba es producir una base ortonormal de autovectores. Ahora, en principio solo existe una base ortonormal para la cual T tiene una matriz triangular *superior* A; y con la misma base T^* tiene la matriz A^* que es triangular *inferior*. Estas matrices obedecen $A^*A = AA^*$. Y ahora, un truco astuto: se comparan los elementos diagonales de estas matrices positivas. Cuando $n = 4$, se obtiene:

$$
\begin{aligned}
|a_{11}|^2 &= |a_{11}|^2 + |a_{12}|^2 + |a_{13}|^2 + |a_{14}|^2, \\
|a_{12}|^2 + |a_{22}|^2 &= |a_{22}|^2 + |a_{23}|^2 + |a_{24}|^2, \\
|a_{13}|^2 + |a_{23}|^2 + |a_{33}|^2 &= |a_{33}|^2 + |a_{34}|^2, \\
|a_{14}|^2 + |a_{24}|^2 + |a_{34}|^2 + |a_{44}|^2 &= |a_{44}|^2.
\end{aligned}
$$

Al examinar estas ecuaciones renglón por renglón, se ve que:

$$a_{12} = a_{13} = a_{14} = 0; \quad \text{luego} \quad a_{23} = a_{24} = 0; \quad \text{finalmente} \quad a_{34} = 0.$$

La matriz triangular A es de hecho *diagonal*. El resto sigue como en la primera versión del teorema espectral.

Conviene reescribir el teorema espectral en términos de matrices: si A es una *matriz normal*, esto es, $A^*A = AA^*$, entonces se obtiene

$$A = \alpha_1 E_1 + \cdots + \alpha_r E_r, \tag{3.26}$$

donde las matrices E_j se comportan como las P_j de (3.23); ante todo, $\underline{E_j^2 = E_j = E_j^*}$.

Una segunda consecuencia del teorema espectral es la Prop. 3.57: una matriz $A \in M_n(\mathbb{C})$ es *positiva* si y solo si A es hermítica con autovalores $\alpha_j \geqslant 0$. Esta matriz hermítica A es *definida positiva* si y solo si cada $\alpha_j > 0$.

La prueba depende de la siguiente propiedad de los proyectores:

$$\langle \boldsymbol{x}, E_j\boldsymbol{x}\rangle = \langle \boldsymbol{x}, E_j^2\boldsymbol{x}\rangle = \langle \boldsymbol{x}, E_j^*E_j\boldsymbol{x}\rangle = \langle E_j\boldsymbol{x}, E_j\boldsymbol{x}\rangle = \|E_j\boldsymbol{x}\|^2 \geqslant 0;$$

en seguida:

$$\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \sum_{j=1}^r \alpha_j \langle \boldsymbol{x}, E_j\boldsymbol{x}\rangle = \sum_{j=1}^r \alpha_j \,\|E_j\boldsymbol{x}\|^2. \tag{3.27}$$

Esta propiedad permitirá probar que A es positiva si y solo si $A = B^*B$ para alguna B.

Sobre la octava lista de ejercicios, la "Tarea 8"

Ejercicio 3.21 Los operadores autoadjuntos comparten las propiedades algebraicas de las matrices hermíticas.

Ejercicio 3.22 Un operador autoadjunto es la diferencia de dos operadores positivas: análogo a lo que sucede con una función real, que es la diferencia de dos funciones no negativas.

Ejercicio 3.25 Los operadores unitarios y normales tienen "caracterizaciones vectoriales". (Esto no requiere el teorema espectral.)

Ejercicio 3.27 La *descomposición polar* de una matriz invertible se puede calcular directamente.

Ejercicios 3.28 Hay un operador integral F, la *transformación de Fourier*, de gran utilidad en la física e ingeniería eléctrica, por ser unitaria y por cumplir $F^4 = 1$. Para hacer cálculos aproximados, se requiere una versión finitodimensional, la F de este ejercicio. Basado en raíces enésimas (complejas) de 1, esta F también es unitaria con $F^4 = 1_n$: entonces su espectro es $\{1, -1, i, -i\}$. Resulta que las *multiplicidades* de estos cuatro autovalores no son obvias: para hallarlas, se requiere un estudio de sumatorias de números complejos hecho originalmente por Gauss.

3.7. Clase 16

Esta es la clase virtual del 2 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la segunda parte de la sección § 3.6: *El teorema espectral y sus consecuencias*.

El teorema espectral (continuación)

Una tercera consecuencia del teorema espectral es la expresión de una matriz positiva A como $A = B^*B$ para alguna matriz B. Este es el Teorema 3.58, que se puede resumir como (a) $\iff$ (b) $\iff$ (c) $\iff$ (d), donde:

(a) A es positiva: $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \geqslant 0$ para todo $\boldsymbol{x}$;

(b) $A = A^*$ y sus autovalores α_j cumplen $\alpha_j \geqslant 0$;

(c) $\underline{A = B^*B}$ para alguna matriz B;

(d) $\underline{A = C^2}$ para alguna matriz hermítica $C = C^*$.

Para empezar, (d) $\implies$ (c) al tomar $B = C$.
Además, (c) $\implies$ (a) porque $\langle \boldsymbol{x}, B^*B\boldsymbol{x}\rangle = \langle B\boldsymbol{x}, B\boldsymbol{x}\rangle \geqslant 0$.
La Prop. 3.57 mostró (a) $\implies$ (b) usando el teorema espectral.

Solo falta (b) $\implies$ (d). El teorema espectral ha mostrado que (b) implica:

$$A = \alpha_1 E_1 + \cdots + \alpha_r E_r, \quad \text{con cada} \quad \alpha_j \geqslant 0.$$

Ahora se define una C particular, la siguiente matriz R que es *positiva*:

$$R := \sqrt{\alpha_1}\, E_1 + \sqrt{\alpha_2}\, E_2 + \cdots + \sqrt{\alpha_r}\, E_r.$$

Cada E_j es un proyector, con $E_i E_j = 0$ para $i \neq j$. Entonces

$$R^2 = \alpha_1 E_1^2 + \cdots + \alpha_r E_r^2 = \alpha_1 E_1 + \cdots + \alpha_r E_r = A.$$

Esto verifica (b) $\implies$ (d) al tomar $C = R$.

Pero R no es solo hermítica, es positiva (y si A es definida positiva, R también es definida positiva). Se escribe $\underline{R = A^{1/2}}$; esto es la **raíz cuadrada positiva** de A.

Una cuarta consecuencia del teorema espectral es la Prop. 3.60: una matriz U es *unitaria* si U es normal ($U^*U = UU^*$) y sus autovalores (complejos) α_j cumplen $|\alpha_j| = 1$.

► Si $T \in \mathcal{L}(V)$ es un operador cualquiera, T^*T es un operador positivo; su **módulo** es el operador positivo

$$|T| := (T^*T)^{1/2}. \tag{3.28}$$

Esto se hace por analogía con los números complejos: si $\underline{z = re^{i\theta}} = r\cos\theta + ir\operatorname{sen}\theta$, entonces $\underline{r = |z|} := \sqrt{\bar{z}z} = (\bar{z}z)^{1/2}$.

► Defn. 3.61: $W \in \mathcal{L}(V)$ es una **isometría parcial** si $\underline{WW^*W = W}$.
Es de notar que tanto W^*W como WW^* son proyectores:

$$(W^*W)^2 = W^*(WW^*W) = W^*W,$$
$$(WW^*)^2 = (WW^*W)W^* = WW^*.$$

Sus imágenes en V son: $WW^*(V) = \underline{W(V)}$ y $W^*W(V) = W^*(V) = \underline{(\ker W)^\perp}$.

► El Teorema 3.62 dice que cualquier operador $T \in \mathcal{L}(V)$ tiene una **descomposición polar** análogo a $re^{i\theta} = \underline{z = e^{i\theta}r}$ en $\mathbb{C}$. Hay una *única* isometría parcial W tal que

$$\ker W = \ker T \quad \text{y} \quad \underline{T = W|T|}. \tag{3.29}$$

Para mostrarlo, se define $W(|T|\boldsymbol{x}) := T\boldsymbol{x}$ y $W\boldsymbol{y} := \mathbf{0}$ si $T\boldsymbol{y} = \mathbf{0}$, aprovechando la suma directa ortogonal $V = |T|(V) \oplus \ker T$. En seguida se comprueba que $WW^*W = W$.

4 Comentarios sobre las clases de MA–460, parte 4

Esta es una versión en pdf de los comentarios en mensajes de texto que acompañaron las clases "virtuales" de MA–460 del 2020, *Álgebra Lineal II*, en el primer semestre, a causa de la suspensión de clases presenciales.

4.1. Clase 17

Esta es la clase virtual del 5 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 4.1: *Formas bilineales y sus matrices*.

Formas bilineales y sus matrices

En este capítulo se estudiará las funciones de dos variables sobre un espacio vectorial, que son lineales en cada variable; es decir, funciones bilineales con valores escalares. Todo esto se resume en el término "formas bilineales". Al inicio se va a abandonar los productos escalares del capítulo anterior, para empezar desde cero. Pero paulatinamente los productos escalares volverán al escenario por una puerta lateral, porque son ejemplos concretos de algunas formas bilineales.

Por ahora, entonces, $\mathbb{F}$ es nuevamente un cuerpo cualquiera. (El cuerpo $\mathbb{Q}$ de los números racionales es un ejemplo importante para las discusiones que siguen.) El espacio $\mathbb{F}$-vectorial V será siempre de dimensión finita.

La Defn. 4.1 dice que una **forma bilineal** es una función $\underline{f\colon V \times V \to \mathbb{F}}$ que es lineal en cada variable:

$$\begin{aligned} f(a\boldsymbol{x} + b\boldsymbol{z}, \boldsymbol{y}) &= a\, f(\boldsymbol{x}, \boldsymbol{y}) + b\, f(\boldsymbol{z}, \boldsymbol{y}), \\ f(\boldsymbol{x}, c\boldsymbol{y} + d\boldsymbol{w}) &= c\, f(\boldsymbol{x}, \boldsymbol{y}) + d\, f(\boldsymbol{x}, \boldsymbol{w}). \end{aligned} \tag{4.1}$$

Un ejemplo obvio es el *producto punto* de dos vectores $\boldsymbol{c}, \boldsymbol{d} \in \mathbb{F}^n$,

$$f(\boldsymbol{c}, \boldsymbol{d}) := \underline{\boldsymbol{c}^{\mathsf{t}}\boldsymbol{d}} = c_1 d_1 + \cdots + c_n d_n \,.$$

Esto es también el producto matricial del vector de fila $\boldsymbol{c}^{\mathsf{t}}$ con el vector de columna $\boldsymbol{d}$. Se obtienen otros ejemplos al intercalar una matriz $n \times n$ entre la fila y columna:

$$f(\boldsymbol{c}, \boldsymbol{d}) := \underline{\boldsymbol{c}^{\mathsf{t}} A\, \boldsymbol{d}} = \sum_{i,j=1}^{n} c_i\, a_{ij}\, d_j \,.$$

Este ejemplo resulta ser *universal*: si $\boldsymbol{x}, \boldsymbol{y} \in V$ y si se toma una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, con $\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$, $\boldsymbol{y} = d_1\boldsymbol{x}_1 + \cdots + d_n\boldsymbol{x}_n$, se define una matriz cuadrada $\underline{A = [f]^{\mathcal{B}}_{\mathcal{B}}}$ por

$$a_{ij} := f(\boldsymbol{x}_i, \boldsymbol{x}_j), . \tag{4.2}$$

Entonces

$$f(\boldsymbol{x}, \boldsymbol{y}) = f\left(\sum_{i=1}^{n} c_i\boldsymbol{x}_i, \sum_{j=1}^{n} d_j\boldsymbol{x}_j\right) = \sum_{i,j=1}^{n} c_i\, a_{ij}\, d_j = \boldsymbol{c}^t A\, \boldsymbol{d}.$$

► Hasta ahora, una matriz cuadrada A representa un operador lineal $T\colon V \to V$; pero aquí no hay operador lineal a la vista. Lo que sí se puede definir, a partir de una forma bilineal f, es una aplicación lineal $\underline{F\colon V \to V^*}$, dada por $F(\boldsymbol{y}) := f_{\boldsymbol{y}}$ donde $f_{\boldsymbol{y}}$ es la forma *lineal* con $f_{\boldsymbol{y}}(\boldsymbol{x}) := f(\boldsymbol{x}, \boldsymbol{y})$. En resumen, $F(\boldsymbol{y})(\boldsymbol{x}) := f(\boldsymbol{x}, \boldsymbol{y})$.

Por ser $f(\boldsymbol{x}, \boldsymbol{y})$ lineal en la *primera* variable $\boldsymbol{x}$, se sigue que $f_{\boldsymbol{y}}$ es lineal; o sea, $f_{\boldsymbol{y}} \in V^*$. Y por ser $f(\boldsymbol{x}, \boldsymbol{y})$ lineal en la *segunda* variable $\boldsymbol{y}$, se sigue que $F(\boldsymbol{y})$ es lineal en $\boldsymbol{y}$, esto es, que $\underline{F \in \mathcal{L}(V, V^*)}$.

Si $\mathcal{B}^* := \{f_1, \dots, f_n\}$ es la base de V^* dual a la base $\mathcal{B}$, se relacionan f, F y la matriz A por la fórmula

$$[F]^{\mathcal{B}^*}_{\mathcal{B}} = A = [f]^{\mathcal{B}}_{\mathcal{B}}.$$

Se debe recordar que, por conteo de dimensiones:

$$\dim \mathcal{L}(V, V^*) = (\dim V)(\dim V^*) = (\dim V)^2 = n^2 = \dim M_n(\mathbb{F}).$$

Este conteo resalta que la correspondencias $F \leftrightarrow A \leftrightarrow f$ son biyecciones.

Se introduce la aplicación F no (solo) por complicar la notación, sino porque un cambio de base de V no modifica A en $P^{-1}AP$ (semejanza), sino en $\underline{P^{\mathrm{t}}AP}$.

Para verlo, sea $\boldsymbol{x}'_s =: \sum_{j=1}^{n} p_{js}\boldsymbol{x}_j$ un cambio de base dada por la matriz invertible P. Esto muda la matriz A, con $a_{ij} := f(\boldsymbol{x}_i, \boldsymbol{x}_j)$, en la matriz B, con entradas $b_{rs} := f(\boldsymbol{x}'_r, \boldsymbol{x}'_s)$. Ahora la Prop. 4.4 usa la bilinealidad de F para comprobar que

$$b_{rs} = \sum_{i,j=1}^{n} p_{ir}\, a_{ij}\, p_{js}\ ,$$

y a la derecha se ve la entrada (r, s) de la matriz $P^{\mathrm{t}}AP$. En breve:

$$\underline{B = P^{\mathrm{t}}AP}. \tag{4.3}$$

Se dice que las matrices A y B son **matrices congruentes**, $A \doteq B$, en vez de *semejantes*. (Estas palabras nada tienen que ver con la clasificación de triángulos en la geometría euclidiana.)

Obviamente, habrá que buscar la manera de clasificar matrices por congruencia. Esto resultará bastante más fácil que la clasificación por semejanza, donde era necesario factorizar el polinomio mínimo, etc., etc. Algo que si es posible afirmar de una vez es que las matrices congruentes *tienen el mismo rango*, $\underline{r(A) = r(B)}$.

► Antes de abordar esa tarea, hay un concepto más de importancia para las formas bilineales. Se dice (Defn. 4.6) que f es **no degenerada** si

$$f(\boldsymbol{x}, \boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{x} \implies \boldsymbol{y} = \mathbf{0}.$$

El Lema 4.7 se puede resumir así:

$$f \text{ es no degenerada} \iff F \text{ es biyectiva} \iff A \text{ es invertible.}$$

► El epígrafe del Capítulo 4 es una cita del gran geómetra italiano *Federigo Enriques*. Su libro *Lezioni di Geometria Proiettiva* (Bologna, 1904) termina con esta linda perorata:

> *Las diversas ramas de la matemática pura y aplicada se entrelazan y conectan entre sí por vías inesperadas; y las ideas que sacan su origen de problemas elementales de la práctica, parece que deben madurarse por una larga elaboración del pensamiento, antes de que puedan descender fecundamente al campo de actividad de la vida.*

4.2. Clase 18

Esta es la clase virtual del 9 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 4.2: *Formas bilineales simétricas*.

Formas bilineales simétricas

Una forma bilineal $d \colon V \times V \to \mathbb{F}$ es **simétrica** si $\underline{d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$.

Su matriz $A = [d]_{\mathcal{B}}^{\mathcal{B}}$ es simétrica: $\underline{A^{\mathsf{t}} = A}$. Esta condición es preservada por la relación de congruencia: $(P^{\mathsf{t}}AP)^{\mathsf{t}} = P^{\mathsf{t}}A^{\mathsf{t}}P = P^{\mathsf{t}}AP$.

De modo similar, se dice que una forma bilineal s es **alternada** o **antisimétrica** si $\underline{s(\boldsymbol{x}, \boldsymbol{y}) = -s(\boldsymbol{y}, \boldsymbol{x})}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$. Estas formas tiene matrices $B = [s]_{\mathcal{B}}^{\mathcal{B}}$ antisimétricas: $\underline{B^{\mathsf{t}} = -B}$.

Cualquier forma bilineal es una suma $f(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y}) + s(\boldsymbol{x}, \boldsymbol{y})$ de una forma simétrica y una forma alternada. Hay una fórmula "casi" obvia (Lema 4.9):

$$\begin{aligned} d(\boldsymbol{x}, \boldsymbol{y}) &:= \tfrac{1}{2}\bigl(f(\boldsymbol{x}, \boldsymbol{y}) + f(\boldsymbol{y}, \boldsymbol{x})\bigr), \\ s(\boldsymbol{x}, \boldsymbol{y}) &:= \tfrac{1}{2}\bigl(f(\boldsymbol{x}, \boldsymbol{y}) - f(\boldsymbol{y}, \boldsymbol{x})\bigr). \end{aligned} \tag{4.4}$$

Hay una excepción si $1+1 = 0$ en $\mathbb{F}$, como ocurre con el cuerpo binario $\mathbb{F}_2 = \{\bar{0}, \bar{1}\}$; porque en esos cuerpos no hay distinción entre simétrica y alternada, y además no hay división por 2: como $2 = 0$, la "mitad" $\frac{1}{2}$ no existe.

Entonces supondremos que $1 + 1 \neq 0$ en $\mathbb{F}$.

Volviendo a una forma bilineal simétrica d, se define el **subespacio ortogonal** (con respecto a d) de $M \leqslant V$:

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{x} \in V : d(\boldsymbol{x}, \boldsymbol{y}) = 0 \ \text{ para todo } \boldsymbol{y} \in M \,\} \\ &= \{\, \boldsymbol{y} \in V : d(\boldsymbol{x}, \boldsymbol{y}) = 0 \ \text{ para todo } \boldsymbol{x} \in M \,\}. \end{aligned} \tag{4.5}$$

Está claro que $\{\mathbf{0}\}^\perp = V$; pero $V^\perp = \{\mathbf{0}\}$ si y solo si d es no degenerada.

La Prop. 4.11 dice que si d es *no degenerada* y $M \cap M^\perp = \{\mathbf{0}\}$, entonces $\underline{M \oplus M^\perp = V}$.

► La **forma cuadrática** $q \colon V \to \mathbb{F}$ asociada a d es la función $\underline{q(\boldsymbol{x}) := d(\boldsymbol{x}, \boldsymbol{x})}$.

En términos de la matriz simétrica A de d, q es una función cuadrática:

$$q(\boldsymbol{x}) = \sum_{i,j=1}^{n} a_{ij}\, x_i x_j \,.$$

El Ejemplo 4.14 exhibe esta función en $\mathbb{F}^3$,

$$q(x, y, z) := \underline{ax^2 + 2hxy + by^2 + 2gxz + 2fyz + cz^2} = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

(Nótese que $2hxy = hxy + hyx$; el coeficiente $2h$ se reparte en un h en ambos lados de la diagonal.)

La Prop. 4.15 da una "fórmula de polarización" para recuperar d a partir de q:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \tfrac{1}{2}\big(q(\boldsymbol{x} + \boldsymbol{y}) - q(\boldsymbol{x}) - q(\boldsymbol{y})\big). \tag{4.6}$$

► Ya es hora de buscar una *forma canónica* para d bajo la relación de congruencia. Es decir, dada $A = A^{\mathrm{t}}$, se busca un cambio de base P tal que $P^{\mathrm{t}}AP$ sea diagonal. La Prop. 4.16 dice que hay una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ de V, donde $k = r(A)$ es el rango de d, tal que

$$[d]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} \quad \text{con } B = \operatorname{diag}[b_1, \dots, b_k] \ \text{ y cada } b_i \neq 0.$$

La prueba es parecido al algoritmo de Gram y Schmidt, pero menos complicado porque no hay producto escalar a la vista.

El Corolario 4.17 saca la consecuencia para la forma cuadrática:

$$q(\boldsymbol{x}) = b_1 f_1(\boldsymbol{x})^2 + b_2 f_2(\boldsymbol{x})^2 + \cdots + b_k f_k(\boldsymbol{x})^2 \tag{4.8}$$

donde cada f_i es una forma *lineal*: $\underline{f_i(\boldsymbol{x}) := b_i^{-1} d(\boldsymbol{x}, \boldsymbol{x}_i)}$.

El Ejemplo 4.19 calcula un ejemplo de (4.8) en detalle: para $\boldsymbol{x} \in \mathbb{Q}^4$, sea

$$q(\boldsymbol{x}) := x_1^2 - 4x_1x_2 + 2x_1x_3 - 2x_1x_4 + 3x_2^2 - 6x_2x_3 + 8x_2x_4 + 2x_3x_4 + 2x_4^2.$$

Para $\boldsymbol{x}_1$ de $\mathcal{B}$, tomamos $\boldsymbol{e}_1$ de la base estándar. Entonces $\underline{b_1 := q(\boldsymbol{e}_1) = 1}$ y se calcula

$$f_1(\boldsymbol{x}) := \underline{b_1^{-1} d(\boldsymbol{x}, \boldsymbol{e}_1)} = x_1 - 2x_2 + x_3 - x_4.$$

Ahora se obtiene una nueva forma cuadrática de 3 variables:

$$\begin{aligned} \underline{q'(\boldsymbol{x})} := q(\boldsymbol{x}) - b_1 f_1(\boldsymbol{x})^2 &= q(\boldsymbol{x}) - (x_1 - 2x_2 + x_3 - x_4)^2 \\ &= -x_2^2 - 2x_2x_3 + 4x_2x_4 - x_3^2 + 4x_3x_4 + x_4^2\,. \end{aligned}$$

En el siguiente paso, $\underline{b_2 := q'(\boldsymbol{e}_2) = -1}$ y $f_2(\boldsymbol{x}) := \underline{b_2^{-1} d(\boldsymbol{x}, \boldsymbol{e}_2)} = x_2 + x_3 - 2x_4$. Así,

$$\underline{q''(\boldsymbol{x})} := q(\boldsymbol{x}) - b_1 f_1(\boldsymbol{x})^2 - b_2 f_2(\boldsymbol{x})^2 = q'(\boldsymbol{x}) - (x_2 + x_3 - 2x_4)^2 = 5x_4^2\,.$$

Con $\boldsymbol{x}_3 = \boldsymbol{e}_4$, se obtiene $b_3 = q''(\boldsymbol{e}_4) = 5$. En fin:

$$q(\boldsymbol{x}) = (x_1 - 2x_2 + x_3 - x_4)^2 - (x_2 + x_3 - 2x_4)^2 + 5x_4^2.$$

4.3. Clase 19

Esta es la clase virtual del 12 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la segunda parte de la sección § 4.2: *Formas bilineales simétricas*.

Formas bilineales simétricas: la reducción de Lagrange

El Ejemplo 4.19, vista en la clase anterior, indica el proceso de simplificar una forma cuadrática real $q(\boldsymbol{x})$ cuando está presentado como una función cuadrática real de n variables. El objetivo es convertir $q(\boldsymbol{x})$ es *una combinación lineal de cuadrados* de funciones de primer grado en dichas variables.

Los *coeficientes* de esta combinación lineal son números reales (muchas veces, enteros): algunos pueden ser positivos y otros negativos. En principio, el cuerpo de escalares $\mathbb{F}$ podría ser otro (los racionales $\mathbb{Q}$, por ejemplo), toda vez que $1 + 1 \neq 0$ en $\mathbb{F}$.

Lo que se requiere ahora es un procedimiento general, sistemático, de efectuar esta simplificación con una forma cuadrática cualquiera sobre $\mathbb{F}^n$ (siempre bajo la hipótesis $1 + 1 \neq 0$). Esta **reducción de Lagrange** expresa $q(\boldsymbol{x})$ como la suma de uno o dos cuadrados (multiplicados por coeficientes) y una forma cuadrática residual en $(n-1)$ o $(n-2)$ variables. Al repetir la receta con las formas cuadráticas residuales hasta agotar los variables remanentes, se llega a la expansión (4.8) deseada.

La Prop. 4.20 ofrece la receta para la reducción. Estas son las fórmulas para el primer término, o el primer par de términos, de la expansión:

$$q(\boldsymbol{x}) = \frac{1}{a_{11}} \biggl(\sum_{j=1}^n a_{1j} x_j \biggr)^2 + q_1(\boldsymbol{x}), \tag{4.9a}$$

$$q(\boldsymbol{x}) = \frac{1}{2a_{12}} \biggl(\sum_{j=1}^n (a_{1j} + a_{2j}) x_j \biggr)^2 - \frac{1}{2a_{12}} \biggl(\sum_{j=1}^n (a_{1j} - a_{2j}) x_j \biggr)^2 + q_2(\boldsymbol{x}). \tag{4.9b}$$

Para aplicar – con preferencia – la primera fórmula (4.9a), es necesario que $a_{11} \neq 0$; o al menos que algún otro $a_{kk} \neq 0$, en cuyo caso se puede permutar $x_k \leftrightarrow x_1$ y luego aplicar la fórmula. Al lado derecho, el coeficiente de x_1^2 es $a_{11}^2/a_{11} = a_{11}$ y el coeficiente de $x_1 x_j$ es $2a_{11}a_{1j}/a_{11} = 2a_{1j}$, los mismos del $q(\boldsymbol{x})$ original. Al restar ese primer término, el resto $q_1(\boldsymbol{x})$ solo depende de $(x_2, \dots, x_n)$.

Pero si todas las entradas diagonales de A son ceros, la receta (4.9a) no funciona. Entonces se puede aplicar la fórmula (4.9b) si $a_{12} \neq 0$; o al menos si algún otro $a_{ij} \neq 0$, en cuyo caso se permuta $x_i \leftrightarrow x_1$ y $x_j \leftrightarrow x_2$ antes de invocar la fórmula. ⟦ Y si *todas* las entradas de A son ceros, entonces $q(\boldsymbol{x}) \equiv 0$ es la fórmula cuadrática nula; y no hace falta hacer nada. ⟧ Se puede chequear ahora que al lado derecho los coeficientes de x_1^2, x_2^2, $x_1 x_j$, $x_2 x_j$ son los de $q(\boldsymbol{x})$ y la forma cuadrática residual $q_2(\boldsymbol{x})$ solo depende de $(x_3, \dots, x_n)$.

► Volviendo a la expansión (4.8):

$$q(\boldsymbol{x}) = b_1\, f_1(\boldsymbol{x})^2 + b_2\, f_2(\boldsymbol{x})^2 + \cdots + b_k\, f_k(\boldsymbol{x})^2 \tag{4.8}$$

si el coeficiente b_j tiene un factor repetido, este se puede absorber el la forma lineal $f_j(\boldsymbol{x})$. Por ejemplo, si $b_1 = 9c_1$, entonces

$$b_1\, f_1(\boldsymbol{x})^2 = 9c_1\, f_1(\boldsymbol{x})^2 = c_1\, f_1(3\boldsymbol{x})^2.$$

Cuando $\mathbb{F} = \mathbb{Q}$, por ejemplo, así se puede modificar los coeficientes de (4.8) para que sean $b_j = r_j/s_j$ con $r_j, s_j \in \mathbb{Z}$ “libres de cuadrados”. Detrás de esa cortina hay muchas problemas de la teoría de números.

Formas bilineales simétricas: la ley de inercia

La reducción de $q(\boldsymbol{x})$ es mucho más sencilla si $\mathbb{F} = \mathbb{R}$ o $\mathbb{F} = \mathbb{C}$. En este caso, la reducción sigue la llamada **ley de inercia** de Sylvester.

Cabe recordar que la expansión (4.8) expresa que la matriz A de la forma bilineal simétrica d en alguna base $\mathbb{C}$ apropiada es *diagonal*: $A \doteqdot \operatorname{diag}[b_1, \dots, b_k, 0, \dots, 0]$, con $k = r(A)$.

Ahora se trata de quitar factores cuadrados de los b_j. Hay dos casos:

(a) Si $\mathbb{F} = \mathbb{C}$, *todo* $b_j \neq 0$ es un cuadrado, $b_j = a_j^2$ y $\underline{c_j = 1}$. Entonces

$$[d]_{\mathcal{C}}^{\mathcal{C}} = \begin{bmatrix} 1_k & 0 \\ 0 & 0_{n-k} \end{bmatrix}. \tag{4.10a}$$

(b) Si $\mathbb{F} = \mathbb{R}$, solo los b_j positivos son cuadrados. Así, $b_j = \pm a_j^2$ y $\underline{c_j = \pm 1}$. Entonces

$$[d]_{\mathcal{C}}^{\mathcal{C}} = \begin{bmatrix} 1_p & 0 & 0 \\ 0 & -1_q & 0 \\ 0 & 0 & 0_{n-p-q} \end{bmatrix} \tag{4.10b}$$

donde $p + q = k = r(A)$.

En el caso real (b), los números p y q son invariantes bajo congruencia de matrices (Prop. 4.24), al igual que el rango $\underline{p + q} = r(A)$. La *diferencia* $\underline{s(A) := p - q} \in \mathbb{Z}$ es la **signatura** de A.

Para resumir, la **ley de inercia** dice lo siguiente: una matriz real simétrica $A = A^t \in M_n(\mathbb{C})$, al ser congruente con la matriz diagonal en (4.10b), clasifica la forma cuadrática $q(\boldsymbol{x}) = \boldsymbol{x}^t A \boldsymbol{x}$ por su rango $r(A) = p + q$ y su signatura $s(A) = p - q$.

► El caso $q = 0$ corresponde a formas bilineales **positivas**: $d(\boldsymbol{x}, \boldsymbol{x}) \geqslant 0$ para todo $\boldsymbol{x}$. El subcaso $q = 0$, $p = k = n$ corresponde a formas bilineales **definidas positivas**: $d(\boldsymbol{x}, \boldsymbol{x}) > 0$ para todo $\boldsymbol{x} \neq \mathbf{0}$.

4.4. Clase 20

Esta es la clase virtual del 16 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 4.3: *Formas hermíticas*.

Formas hermíticas

En la discusión de formas bilineales $f\colon V \times V \to \mathbb{F}$, hasta ahora se ha omitido el producto escalar complejo que *no es* bilineal: es lineal en una variable pero solo *semi*lineal (esto es, antilineal) en la otra. Cuando se trata de escalares complejos, $\mathbb{F} = \mathbb{C}$, es necesario retroceder para considerar funciones de dos variables que son "$\frac{3}{2}$-lineales". Se dice que $h\colon V \times V \to \mathbb{C}$ es una **forma sesquilineal** si h es aditiva en las dos variables pero la multiplicación escalar funciona así:

$$h(z, \alpha w) = \alpha\, h(z, w), \qquad h(\alpha z, w) = \bar{\alpha}\, h(z, w), \qquad z, w \in V,\ \alpha \in \mathbb{C}.$$

Un producto escalar complejo $\langle z, w\rangle$ tiene estas propiedades, y además es **hermítica**:

$$h(z, w) = \overline{h(w, z)} \quad \text{para} \quad z, w \in V.$$

Esta es la "versión de $\frac{3}{2}$" de una forma bilineal *simétrica*.

Dada una base $\mathcal{B} = \{z_1, \dots, z_n\}$ de V, se define la matriz de h como antes:

$$A = [h]_{\mathcal{B}}^{\mathcal{B}} \quad \text{donde} \quad a_{ij} := h(z_i, z_j).$$

Es h es una forma hermítica si y solo si $\underline{A^* = A}$.

Bajo un cambio de base $\mathcal{B} \mapsto \mathcal{B}'$ usando una matriz invertible P, la nueva matriz de h, $B = [h]_{\mathcal{B}'}^{\mathcal{B}'}$, está relacionada con A por una relación de **$*$-congruencia**:

$$B = P^* A P. \tag{4.12}$$

Nótese la $P^* = \overline{P^{\mathsf{t}}}$ en vez de la P^{t} de congruencia: las relaciones de congruencia y $*$-congruencia son diferentes.

En contraste con el producto escalar complejo, una forma hermítica no tiene que ser definida positiva. En lugar de eso, hay una *ley de inercia* (de Sylvester) directamente análoga al caso de formas bilineales simétricas reales. Por ejemplo:

$$\underline{h(z, w)} := \bar{z}_1 w_1 + \cdots + \bar{z}_p w_p - \bar{z}_{p+1} w_{p+1} - \cdots - \bar{z}_{p+q} w_{p+q}$$

con $p + q \leqslant n$, es una forma hermítica de **rango** $\underline{k = p + q}$ y **signatura** $\underline{s = p - q}$. Su matriz en la base estándar $\mathcal{E}$ de $\mathbb{C}^n$ es

$$[h]_{\mathcal{E}}^{\mathcal{E}} = \operatorname{diag}[1, \dots, 1, -1, \dots, -1, 0, \dots, 0] = \begin{bmatrix} 1_p & 0 & 0 \\ 0 & -1_q & 0 \\ 0 & 0 & 0_{n-p-q} \end{bmatrix}. \tag{4.13}$$

Prop. 4.32: cada forma hermítica tiene una base $\mathcal{C}$ tal que la matriz $[h]_{\mathcal{C}}^{\mathcal{C}}$ tiene este aspecto, con p, q determinadas por h. La forma h es *positiva* si y solo si $q = 0$; es *negativa* si y solo si $p = 0$.

La prueba de la Prop. 4.32 es diferente del caso simétrico real: se usa el resultado (Lema 3.50) de que una matriz hermítica $A = A^*$ es diagonalizable por una matriz unitaria U; así que Prop. 4.32

En consecuencia, al contar autovalores con multiplicidad, se obtiene:

(a) $\underline{p}$ es *el número total de autovalores positivos* de A;

(b) $\underline{q}$ es *el número total de autovalores negativos* de A;

(c) $n(A) = n - p - q$ es *el número total de autovalores ceros* de A.

Sobre la novena lista de ejercicios, la "Tarea 9"

Ejercicio 4.5 Para una forma bilineal simétrica d [o una forma hermítica h] el concepto de subespacio ortogonal (4.5) se comporta como el complemento ortogonal con respecto a un producto escalar; pero cuando d es degenerado se debe manejarlo con cautela.

Ejercicio 4.6 Un resultado no visto en el curso dice que una matriz simétrica real puede ser diagonalizada por una matriz ortogonal. Este ejercicio exhibe un ejemplo concreto de ese fenómeno.

Ejercicios 4.8 y 4.9 Dada una lista de formas cuadráticas específicas, se les aplica la reducción de Lagrange para así determinar sus rangos y signaturas.

Ejercicios 4.10 y 4.11 Se trata de hacer lo mismo con algunas formas cuadráticas de tipo más general. En el Ejercicio 4.11, el rango y la signatura pueden depender de los parámetros a y b.

Ejercicio 4.12 Las formas cuadráticas se usan para analizar la geometría de curvas y superficies dadas por ecuaciones de segundo grado. En el caso plano, las curvas cuadráticas – las llamadas *cónicas* – tienen rectas tangentes que los cortan una sola vez. (Las demás rectas las cortan dos veces o ninguna.) Este ejercicio es un ejemplo del uso geométrico de las formas cuadráticas.

4.5. Clase 21

Esta es la clase virtual del 19 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 5.4: *Formas bilineales alternadas*.

Formas bilineales alternadas

Después de estudiar formas bilineales simétricas, conviene abordar formas bilineales antisimétricas o *alternadas*. Aquí de nuevo se considera un cuerpo de escalares $\mathbb{F}$ cualquiera con la sola condición de que $1 = 1 \neq 0$.

La siguiente forma bilineal alternada sobre $\mathbb{F}^2$:

$$s_0(\boldsymbol{x}, \boldsymbol{y}) := \underline{x_1 y_2 - x_2 y_1} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \equiv \underline{\boldsymbol{x}^{\mathrm{t}} J_2\, \boldsymbol{y}} \qquad J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

es el ejemplo básico. En cierto sentido, cualquier forma bilineal alternada es congruente a una suma directa de copias de s_0.

Vocabulario: una forma bilineal alternada *no degenerada* se llama **simpléctica**. Dada una base de V, su matriz A es antisimétrica e invertible.

Análogo al complemento ortogonal para una forma bilineal simétrica $\underline{d}$, se puede definir el **complemento simpléctico** con respecto a $\underline{s}$ con la misma fórmula:

$$M^{\perp} = \{\, \boldsymbol{x} \in V : s(\boldsymbol{x}, \boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{y} \in M \,\}.$$

Pero en contraste con la versión simétrica, en este caso podría ocurrir que $\underline{M \subseteq M^{\perp}}$ debido a la antisimetría.

La Prop. 4.37 enuncia la "forma normal" para la matriz de una forma bilineal alternada. Se define J_{2m} como suma directa de m bloques J_2:

$$\underline{J_{2m}} := \begin{bmatrix} J_2 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_2 \end{bmatrix} = \underbrace{J_2 \oplus \cdots \oplus J_2}_{m \text{ veces}}.$$

Entonces existe una base $\mathcal{B} = \{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_m, \boldsymbol{y}_m, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$ de V tal que:

$$[s]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} J_{2m} & 0 \\ 0 & 0_{n-2m} \end{bmatrix}. \tag{4.16}$$

La prueba es por inducción. Primero se obtiene un subespacio $\mathrm{lin}\langle \boldsymbol{x}_1, \boldsymbol{y}_1\rangle \simeq \mathbb{F}^2$ en donde s tiene matriz J_2. Luego se acumulan otros pares $\{\boldsymbol{x}_j, \boldsymbol{y}_j\}$ tales que

$$s(\boldsymbol{x}_j, \boldsymbol{y}_j) = 1, \quad s(\boldsymbol{y}_j, \boldsymbol{x}_j) = -1. \tag{4.17}$$

Después de unos m pares de esos vectores, se agota el rango de s, y se rellena la matriz de s con un bloque de ceros.

Durante la construcción de la matriz (4.16), se descubre que el *rango* de s es necesariamente *par*: $\underline{k = 2m}$.

(En particular, si $\dim V = 2m + 1$, V *no posee* una forma bilineal simpléctica.)

Pfaffianos de matrices antisimétricas

La forma normal (4.16) tiene una consecuencia inesperada. Consideremos el caso donde s es no degenerada: $n = 2m$ y $[s]_{\mathcal{B}}^{\mathcal{B}} = J_{2m}$. En otra base, s tiene una matriz A tal que $\underline{A^{\mathrm{t}} = -A}$ y A^{-1} existe.

Entonces A y J_{2m} son congruentes: hay una matriz R tal que $\underline{A = R^{\mathrm{t}} J_{2m} R}$. Como $\det J_{2m} = (\det J_2)^m = 1$ y $\det R^{\mathrm{t}} = \det R$, se ve que

$$\det A = (\det R)^2. \tag{4.18}$$

Ahora la sorpresa: la cantidad $\underline{\det R}$ resulta ser *un polinomio en las entradas* de A (con coeficientes enteros, de grado m). Este polinomio se llama el **pfaffiano** de la matriz antisimétrica A; y se escribe $\underline{\operatorname{Pf} A} := \det R$.

Antes de ver por qué, aquí están los dos ejemplos básicos. Con $2m = 2$,

$$A = \begin{bmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{bmatrix}, \qquad \det A = a_{12}^2, \qquad \operatorname{Pf} A := a_{12}.$$

El caso 4×4 es más interesante:

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{bmatrix} \implies \boxed{\operatorname{Pf} A = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}.} \tag{4.21}$$

La fórmula básica es: $\underline{\det A = (\operatorname{Pf} A)^2}$. Esto determina $\operatorname{Pf} A$ hasta un signo. Se debe agregar una regla para fijar el signo; esa regla es $\underline{\operatorname{Pf}(J_{2m}) = +1}$.

Otra fórmula importante para los pfaffianos es la siguiente:

$$\operatorname{Pf}(R^{\mathrm{t}} A R) = (\det R)(\operatorname{Pf} A). \tag{4.23}$$

4.6. Clase 22

Esta es la clase virtual del 23 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 5.4: *Formas bilineales alternadas* (segunda parte).

Polinomios y pfaffianos

Una matriz antisimétrica invertible A de lado $2m$ es congruente con la matriz J_{2m}, por la Prop. 4.37: hay una matriz R tal que $A = R^{\mathrm{t}} J_{2m} R$. Como $\det J_{2m} = 1$ y $\det R^{\mathrm{t}} = \det R$, se ve que $\det A = (\det R)^2$. Si el cuerpo de escalares es $\mathbb{R}$, esto implica que $\det A > 0$.

Pero hay un resultado mucho más potente: para una matriz antisimétrica invertible A, hay una expresión $\underline{\operatorname{Pf} A}$ tal que $\det A = (\operatorname{Pf} A)^2$, donde $\operatorname{Pf} A$ es *un polinomio de grado $\underline{m}$ en las entradas de A*. (De paso, nótese que el rango de una matriz antisimétrica siempre es par, por la misma Prop. 4.37, así que A tiene lado $2m$ simplemente por ser invertible.)

El origen de $\operatorname{Pf} A$ viene de la teoría de polinomios en varias variables. En primer lugar, considérese esta matriz antisimétrica:

$$T := \begin{bmatrix} 0 & t_{12} & t_{13} & \dots & t_{1n} \\ -t_{12} & 0 & t_{23} & \dots & t_{2n} \\ -t_{13} & -t_{23} & 0 & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -t_{1n} & -t_{2n} & -t_{3n} & \dots & 0 \end{bmatrix}.$$

Las entradas del triángulo superior de T son "incógnitas" t_{ij} para $i < j$. Escribimos $\boldsymbol{t} := (t_{12}, t_{13}, \dots, t_{n-1,n})$. ¿En cuál cuerpo viven estas entradas? Este cuerpo es $\mathbb{Q}(\boldsymbol{t})$, cuyos elementos son cocientes de polinomios $q(\boldsymbol{t})/r(\boldsymbol{t})$, con $r(\boldsymbol{t}) \neq 0$; los coeficientes de estos polinomios son números racionales m/n con $m, n \in \mathbb{Z}$ y $n \neq 0$. Al multiplicar el numerador $q(\boldsymbol{t})$ y el denominador $r(\boldsymbol{t})$ por un número entero grande, se puede suponer que los coeficientes de ambos son *números enteros*. Para resumir: un elemento de $\mathbb{Q}(\boldsymbol{t})$ es un cociente $q(\boldsymbol{t})/r(\boldsymbol{t})$, donde $\underline{q(\boldsymbol{t}), r(\boldsymbol{t}) \in \mathbb{Z}[\boldsymbol{t}]}$ son polinomios *sin factor común*. (Es de notar que las entradas de T son monomios de primer grado en $\mathbb{Z}[\boldsymbol{t}]$.)

Obsérvese que si $n = 3$, entonces $\det T = -t_{12}t_{23}t_{13} + t_{13}t_{12}t_{23} = 0$; como ya se ha dicho, una matriz antisimétrica de lado impar no puede ser invertible. Continuamos, entonces, suponiendo que n es par.

Ahora el resultado $\det A = (\det R)^2$, válido en cualquier cuerpo $\mathbb{F}$, nos informa que debe haber una matriz S tal que $\det T = (\det S)^2$. Las entradas de S, así como la cantidad $(\det S)$, son elementos del cuerpo $\mathbb{Q}(\boldsymbol{t})$. Entonces hay tres polinomios $q(\boldsymbol{t}), r(\boldsymbol{t}), s(\boldsymbol{t}) \in \mathbb{Z}[\boldsymbol{t}]$ tales que

$$\det T = (\det S)^2 \implies s(\boldsymbol{t}) = \frac{q(\boldsymbol{t})^2}{r(\boldsymbol{t})^2}\,.$$

El polinomio $s(\boldsymbol{t}) := \det T$ viene de la fórmula de Leibniz (1.19) para un determinante; sus coeficientes son ± 1.

La fórmula anterior se puede reacomodar así:

$$s(\boldsymbol{t})r(\boldsymbol{t})^2 = q(\boldsymbol{t})^2. \tag{4.19}$$

Esto implica que cualquier factor irreducible de $r(\boldsymbol{t})$ es también un factor de $q(\boldsymbol{t})$; pero $q(\boldsymbol{t})$ y $r(\boldsymbol{t})$ no tienen factor común!

La última afirmación depende de un teorema cuya prueba no es obvia: en $\mathbb{Z}[\boldsymbol{t}]$, al igual que en $\mathbb{Z}$, *la factorización es única*, salvo signo. En $\mathbb{Z}$, por ejemplo, se sabe que $15 = 3 \cdot 5 = (-3)(-5)$ y no hay otra factorización de 15. El teorema mencionado dice que lo mismo ocurre en $\mathbb{Z}[\boldsymbol{t}]$. Esto implica que $r(\boldsymbol{t}) = \pm 1$ y que $r(\boldsymbol{t})^2 = 1$. Entonces:

$$\det T = s(\boldsymbol{t}) = q(\boldsymbol{t})^2 \quad \text{donde} \quad q(\boldsymbol{t}) = \pm \det S.$$

Se obtiene A de T por las sustituciones $t_{ij} \mapsto a_{ij}$; y el mismo proceso convierte $q(\boldsymbol{t})$ en $\operatorname{Pf} A$.

La ambigüedad del signo $\pm$ en la fórmula universal para $\operatorname{Pf} A$ se resuelve al declarar el signo en un solo caso. Se decide por $\operatorname{Pf}(J_{2m}) = +1$.

La fórmula general para un pfaffiano

Ya se ha visto dos casos particulares: si $2m = 2$, entonces $\operatorname{Pf} A = a_{12}$. También,

$$\text{si } 2m = 4: \qquad \operatorname{Pf} A = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}. \tag{4.21}$$

He aquí la fórmula general:

$$\operatorname{Pf} A := \frac{1}{2^m\, m!} \sum_{\sigma} (-1)^{\sigma}\, a_{\sigma(1)\sigma(2)}\, a_{\sigma(3)\sigma(4)} \,\dots\, a_{\sigma(2m-1)\sigma(2m)} \tag{4.22}$$

donde σ recorre las permutaciones de $(1, 2, \dots, 2m)$. Si σ es la transposición $(1 \leftrightarrow 2)$, por ser $a_{12} = -a_{21}$ hay una duplicación de términos; el factor $1/2^m$ elimina estas duplicaciones. El cambio $12, 34 \leftrightarrow 34, 12$ conserva el lado derecho; el factor $1/m!$ elimina tales redundancias. Cada término es un monomio con m factores; $\operatorname{Pf} A$ es un polinomio homogéneo de grado m en las entradas de A.

Sobre la décima lista de ejercicios, la "Tarea 10"

Ejercicios 4.14 y 4.15 Una manera de estudiar formas cuadráticas reales es considerarlas como funciones continuas sobre la esfera unitaria en $\mathbb{R}^n$. Como esa esfera es compacta, se sabe que una función continua *alcanza sus valores máximo y mínimo* en dos puntos de la esfera. Sin entrar en el terreno de análisis, se puede hallar esos valores extremos usando autovalores y autovectores.

Ejercicio 4.16 Es probable que la matriz R buscada tenga entradas en $\mathbb{Z}$.

Ejercicios 4.18 y 4.19 Históricamente, las matrices 2×2 fueron inventadas por Cayley al estudiar transformaciones $w := (az+b)/(cz+d)$ del plano complejo. El caso $w := (1+z)/(1-z)$ convierte el disco unitario $|z| < 1$ en el semiplano derecho $\mathfrak{R}w > 0$. Curiosamente, al aplicar esta receta a matrices, convierte matrices antisimétricas reales en matrices ortogonales *y viceversa*.

Ejercicio 4.21 Debe ser obvio que R depende de A o de A^{t} pero no de las dos.

Ejercicio 4.22 Se recomienda ensayar este Ejercicio con una matriz 6×6, para ver el patrón general. Es un caso particular de la fórmula (4.22), pero se busca una demostración directa que no emplea esa fórmula.

4.7. Clase 23

Esta es la clase virtual del 26 de junio *Álgebra Lineal II*, en el primer semestre . La temática es la sección § 5.4: *Formas bilineales alternadas* (tercera y última parte).

Estructuras complejas ortogonales

El producto escalar complejo define dos formas $\mathbb{R}$-bilineales: en la fórmula

$$\langle x, y\rangle =: d(x, y) + i\, s(x, y) \tag{4.24}$$

la *parte real* $d: V \times V \to \mathbb{R}$ es simétrica y la *parte imaginaria* $s: V \times V \to \mathbb{R}$ es alternada. Esto sucede porque la relación $\langle y, x\rangle = \overline{\langle x, y\rangle}$ es equivalente a

$$d(y, x) + i\, s(y, x) = d(x, y) - i\, s(x, y).$$

Esto da lugar a un "problema inverso" interesante: dado un espacio vectorial *real* V y una forma bilineal d o s, según las circunstancias, se quiere *reconstruir* un producto escalar (4.24). Esto exige obtener una s a partir de d, o una d a partir de s.

Conviene abordar los dos casos por separado. Sea V un espacio $\mathbb{R}$-vectorial, con $\dim V = n = 2m$ (la dimensión debe ser par) y sea $\underline{d}\colon V \times V \to \mathbb{R}$ una forma bilineal simétrica definida positiva (esto es, un producto escalar real). Para obtener la s, se requiere un operador $\mathbb{R}$-lineal $J\colon V \to V$ tal que

$$J^2 = -1_V\,; \qquad d(J\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y}) \quad \text{para } \boldsymbol{x}, \boldsymbol{y} \in V. \tag{4.25}$$

Una tal J es una **estructura compleja ortogonal**: la relación $\underline{J^2 = -1_V}$ sugiere que J es un sustituto para los números complejos $\pm i$.

Si se identifica $V \simeq \mathbb{R}^n$ (al tomar una base apropiado para V) se puede suponer que $d(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{t}}\boldsymbol{y}$. Entonces el operador J se ve como una matriz $J \in M_n(\mathbb{R})$ tal que

$$J^2 = -1_n \qquad \text{y} \qquad J^{\mathrm{t}} J = 1_n\,.$$

Entonces J es una *matriz ortogonal*: $\underline{J^{\mathrm{t}} = J^{-1} = -J}$ y además *antisimétrica* e invertible.

(Es por eso que $n = r(J) = 2m$ necesariamente.)

Ahora la fórmula para la nueva forma $\underline{s}$ es:

$$\underline{s(\boldsymbol{x}, \boldsymbol{y}) := d(J\boldsymbol{x}, \boldsymbol{y})}.$$

Este cálculo muestra que s es alternada:

$$s(\boldsymbol{y}, \boldsymbol{x}) = d(J\boldsymbol{y}, \boldsymbol{x}) = d(J^2\boldsymbol{y}, J\boldsymbol{x}) = -d(\boldsymbol{y}, J\boldsymbol{x}) = -d(J\boldsymbol{x}, \boldsymbol{y}) = -s(\boldsymbol{x}, \boldsymbol{y}). \tag{4.26}$$

Resulta que d es definida positiva $\implies$ s es no degenerada. Luego $s\colon V \times V \to \mathbb{R}$ es una forma *simpléctica* sobre $V \simeq \mathbb{R}^{2m}$.

Estructuras complejas positivas

El segundo caso toma también un espacio $\mathbb{R}$-vectorial V con $\dim V = n = 2m$, pero ahora con una forma bilineal simpléctica $\underline{s}\colon V \times V \to \mathbb{R}$ desde el inicio. Se quiere fabricar una d, simétrica y definida positiva. En este caso la positividad tiene que ser parte de la definición de J:

$$J^2 = -1_V\,; \qquad s(J\boldsymbol{x}, J\boldsymbol{y}) = s(\boldsymbol{x}, \boldsymbol{y}) \quad \text{si } \boldsymbol{x}, \boldsymbol{y} \in V; \qquad \underline{s(\boldsymbol{x}, J\boldsymbol{x}) > 0} \quad \text{si } \boldsymbol{x} \neq \boldsymbol{0}. \tag{4.27}$$

Tal J se llama una **estructura compleja positiva**.

La receta para la nueva forma $\underline{d}$ es:

$$\underline{d(\boldsymbol{x}, \boldsymbol{y}) := s(\boldsymbol{x}, J\boldsymbol{y})}.$$

Este cálculo muestra que d es simétrica:

$$d(\boldsymbol{y}, \boldsymbol{x}) = s(\boldsymbol{y}, J\boldsymbol{x}) = s(J\boldsymbol{y}, J^2\boldsymbol{x}) = -s(J\boldsymbol{y}, \boldsymbol{x}) = s(\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y}). \tag{4.28}$$

De real a complejo: espacios de Hilbert

De nuevo, sea V un espacio $\mathbb{R}$-vectorial con $\underline{\dim_{\mathbb{R}} V = 2m}$, dotado de una d, una s y una J a partir de cualquiera de las dos casos anteriores. La relación $J^2 = -1_V$ hace de V un espacio vectorial *complejo* mediante la receta:

$$(a + ib)\boldsymbol{x} := a\,\boldsymbol{x} + b\,J\boldsymbol{x}. \tag{4.29a}$$

En seguida, se define $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := d(\boldsymbol{x}, \boldsymbol{y}) + i\,s(\boldsymbol{x}, \boldsymbol{y})$. Ahora V es un **espacio de Hilbert** de dimensión compleja m, esto es, $\underline{\dim_{\mathbb{C}} V = m}$. (Esta es la Prop. 4.50.)

Hay *otra manera* de hacer de V un espacio de Hilbert: se define su **complexificación**:

$$\underline{V_{\mathbb{C}} = V \oplus iV} := \{\, \boldsymbol{x} + i\,\boldsymbol{y} : \boldsymbol{x}, \boldsymbol{y} \in V \,\}.$$

Este espacio tiene mayor dimensión: $\underline{\dim_{\mathbb{C}} V_{\mathbb{C}} = 2m}$.

La forma $\mathbb{R}$-bilineal $d \colon V \times V \to \mathbb{R}$ se puede *ampliar* en una forma $\mathbb{C}$-bilineal $d \colon V_{\mathbb{C}} \times V_{\mathbb{C}} \to \mathbb{C}$ de manera "obvia":

$$d(\boldsymbol{x} + i\boldsymbol{y}, \boldsymbol{x}' + i\boldsymbol{y}') := d(\boldsymbol{x}, \boldsymbol{x}') + i\,d(\boldsymbol{x}, \boldsymbol{y}') + i\,d(\boldsymbol{y}, \boldsymbol{x}') - d(\boldsymbol{y}, \boldsymbol{y}').$$

Advertencia: aunque la d original es definida positiva, la d ampliada es *indefinida*. Fíjese que $d(i\boldsymbol{y}, i\boldsymbol{y}') = -d(\boldsymbol{y}, \boldsymbol{y}')$. ⟦ Una forma bilineal *compleja* no tiene signatura. ⟧

Sobre la undécima lista de ejercicios, la "Tarea 11"

Ejercicios 4.23 La forma general para el pfaffiano se puede deducir de una regla más simple (aunque no demostrada) de expansión en filas y columnas.

Ejercicio 4.24 Las matrices ortogonales R permutan las posibles estructuras complejas ortogonales, $J \mapsto RJR^{-1}$: no hay unicidad de la J.

Ejercicios 4.25 y 4.26 De hecho, estos ejercicios exhiben *todas* las posibles estructuras complejas ortogonales sobre $\mathbb{R}^4$.

Ejercicio 4.27 Uno de los usos de estructuras complejas es una descripción de matrices ortogonales mediante relaciones cuadráticas.

Ejercicio 4.28 Dada una estructura compleja ampliada $J \colon V_{\mathbb{C}} \to V_{\mathbb{C}}$, el operador $P := \frac{1}{2}(1 - iJ)$ sobre $V_{\mathbb{C}}$ es un proyector ortogonal.

Ejercicio 4.29 Dentro de la complexificación $V_{\mathbb{C}}$ se pueden identificar copias isomorfos de V, las *polarizaciones* W, que vienen en pares. Hay un caso $W_0 \oplus \overline{W_0} = V_{\mathbb{C}}$ que da la descomposición según los autovalores $\{i, -i\}$ de la $J \colon V_{\mathbb{C}} \to V_{\mathbb{C}}$ (ampliada).