1 Comentarios sobre las clases de MA–360

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

1.1. Clase 1

Esta es la clase virtual del 5 de abril del 2021, al inicio del primer ciclo lectivo. La temática es la Introducción al curso, el inicio del Capítulo 1: *Vectores en* $\mathbb{R}^n$, y la sección § 1.1: *El producto escalar de dos vectores en* $\mathbb{R}^n$.

Introducción

El álgebra lineal es una disciplina cuyo origen viene de la resolución de sistemas de primer grado en varias variables. Estos sistemas dan lugar a dos tipos de cantidad aritmética: los *vectores* y las *matrices*, junto con sus reglas de combinación. Su estudio comprende aspectos *estructurales* y *algorítmicos* y es necesario dominar estos dos aspectos.

Al principio, se pasará revista a lo que antes se llamaba la *geometría analítica* del plano cartesiano y del espacio tridimensional. Los conceptos vistos allí informarán los esquemas más generales que forman la materia propiamente del curso.

En la bibliografía del curso, se sugieren varios libros que pueden ayudar con la temática del curso, en la medida en que la biblioteca permite su acceso. De todos modos, se ofrecerá un jugo completo de apuntes en archivos pdf. Este guión comentará los apuntes, tratando de señalar los aspectos meas importantes.

Vectores en $\mathbb{R}^n$

El origen histórico de los vectores no está del todo claro. Viene de los tiempos del Renacimiento, que produjo estudios sobre equilibrios de fuerzas o combinaciones de fuerzas en problemas de mecánica. Un concepto mecánico importante de esa época fue el *paralelogramo de fuerzas*. Una fuerza se representa por una flecha con una determinada *magnitud* actuando en una determinada *dirección*. Para combinar dos fuerzas distintas, se considera la diagonal del paralelogramo cuyos lados están formados por copias de las dos fuerzas originales (Figura 1.1).

En el plano cartesiano, un vector está especificado por las coordenadas (x, y) de una flecha anclada en el origen. La *suma* o resultante de dos vectores corresponde a

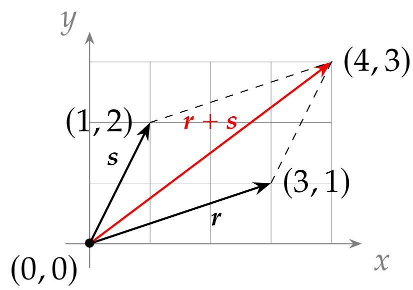

Figura 1.1: El paralelogramo de fuerzas en el plano $\mathbb{R}^2$

la operación de sumar sus coordenadas individuales:

$$\boldsymbol{r} + \boldsymbol{s} = (x, y) + (u, v) := (x + u, y + v).$$

Esta receta también es válida para vectores con tres (o más) coordenadas. Se denota por

$$\mathbb{R}^n := \{\, \boldsymbol{x} = (x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R} \,\}$$

la totalidad de “vectores” con n coordenadas cartesianas.

Otra operación, más sencilla pero no menos importante, es la de magnificar o dilatar un vector por un cambio de escala (conservando su dirección). El factor de escala $c \in \mathbb{R}$ se llamará un *escalar*. En resumen, los elementos de $\mathbb{R}^n$ admiten dos operaciones:

$$\begin{aligned} \underline{\boldsymbol{x} + \boldsymbol{y}} &= (x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n), \\ \underline{c\,\boldsymbol{x}} &= c(x_1, x_2, \dots, x_n) := (cx_1, cx_2, \dots, cx_n). \end{aligned} \tag{1.1}$$

El producto escalar en $\mathbb{R}^n$

El *producto escalar* o *producto punto* de los vectores $\boldsymbol{x} = (x_1, \dots, x_n)$, $\boldsymbol{y} = (y_1, \dots, y_n)$ en $\mathbb{R}^n$ es, como su nombre indica, un escalar:

$$\underline{\boldsymbol{x} \cdot \boldsymbol{y}} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{1.2}$$

Este producto escalar es obviamente *conmutativo*, es decir, $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$. La Prop. 1.3 de los apuntes enumera otras reglas aritméticas que el producto punto satisface. La más importante es la “positividad” $\boldsymbol{x} \cdot \boldsymbol{x} \geqslant 0$, que viene de

$$\boldsymbol{x} \cdot \boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2 \geqslant 0.$$

Nótese que

$$x \cdot x = 0 \implies \text{cada } x_j^2 = 0 \implies \text{cada } x_j = 0 \implies x = \mathbf{0}.$$

Entonces $x \cdot x$ tiene una raíz cuadrada no negativa, llamada su *norma* o *longitud*:

$$\|x\| := \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{1.3}$$

con $\|x\| = 0$ en $\mathbb{R}$ si y solo si $x = \mathbf{0}$ en $\mathbb{R}^n$.

► La propiedad más importante del producto escalar es la **desigualdad de Cauchy** (o de Schwarz, o de Bunyakovskiy):

$$|x \cdot y| \leqslant \|x\| \, \|y\| \tag{1.4}$$

con igualdad *si y solo si* x, y son proporcionales (es decir, paralelos).

Esta es la Prop. 1.5, cuya demostración usa la fórmula cuadrática vista en la enseñanza media. En efecto, se ve que

$$0 \leqslant \|x + ty\|^2 = (x + ty) \cdot (x + ty) = (x \cdot x) + 2t(x \cdot y) + t^2(y \cdot y),$$

una expresión de la forma $f(t) = \underline{at^2 + bt + c}$, cuyo grafo es una parábola vertical. Si la parábola cortara el eje horizontal dos veces – esto es, si la ecuación $at^2 + bt + c = 0$ tuviera dos soluciones distintas, entonces $f(t)$ tendría algunos valores negativos, cosa que no sucede. El criterio conocido para que haya a lo sumo una solución de $f(t) = 0$ dice que $b^2 - 4ac \leqslant 0$, lo cual es equivalente a la fórmula (1.4).

► Entre las consecuencias de la desigualdad de Cauchy es que (si $x \neq \mathbf{0}$, $y \neq \mathbf{0}$):

$$-1 \leqslant \frac{x \cdot y}{\|x\| \, \|y\|} \leqslant 1 \implies x \cdot y = \|x\| \, \|y\| \cos\theta \tag{1.5}$$

para algún ángulo θ (único, si se toma $0 \leqslant \theta \leqslant \pi$): *el ángulo entre los vectores* x, y.

Otra consecuencia es la *desigualdad triangular* para la norma, ver la Prop. 1.6:

$$\|x + y\| \leqslant \|x\| + \|y\|.$$

Para verla, tómese el cuadrado del lado izquierdo:

$$\|x + y\|^2 = (x + y) \cdot (x + y) = \|x\|^2 + 2\, x \cdot y + \|y\|^2$$

y la desigualdad de Cauchy entonces muestra que

$$\|x + y\|^2 \leqslant \|x\|^2 + 2\|x\| \, \|y\| + \|y\|^2 = \big(\|x\| + \|y\|\big)^2.$$

► Si fuera $\boldsymbol{x} \cdot \boldsymbol{y} = 0$, el cálculo anterior daría $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$, la *fórmula de Pitágoras*. Por eso se declara que dos vectores $\boldsymbol{x}, \boldsymbol{y}$ son *ortogonales* si $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. (Esta es la Defn. 1.7).

Más adelante, se trabajará muchas veces con vectores de tipo más general que no poseen un producto escalar (o al menos, cuyo producto escalar no está explícitamente declarado). En la ausencia de un producto punto, no se debe hablar de ortogonalidad. Esto significa que la ortogonalidad es un concepto secundario en el álgebra lineal. (Pero se retomará ese concepto en el Capítulo 5.)

1.2. Clase 2

Esta es la clase virtual del 8 de abril del 2021. La temática comprende las secciones § 1.2: *El producto vectorial en* $\mathbb{R}^3$ y § 1.3: *Rectas y planos en* $\mathbb{R}^2$ *y* $\mathbb{R}^3$.

El producto vectorial en $\mathbb{R}^3$

En la clase anterior, se introdujeron los vectores en $\mathbb{R}^n$, con sus sumas y múltiplos escalares. Se obtuvo además un *producto* $\boldsymbol{x} \cdot \boldsymbol{y}$ de dos vectores $\boldsymbol{x}, \boldsymbol{y}$, pero este no es otro vector; es un *escalar*, es decir, un número sin direccionalidad. Es natural preguntar por qué un producto de dos vectores no ofrece un tercer *vector*.

Resulta que sí es posible definir un producto vectorial, pero únicamente en 3 dimensiones: no hay tal cosa en $\mathbb{R}^2$, ni en $\mathbb{R}^4$, $\mathbb{R}^5$, etc. Además, el producto vectorial en $\mathbb{R}^3$, curiosamente, no es conmutativa sino *anti*conmutativo.

Para aliviar un poco la notación, es recomendable designar las coordenadas de ciertos 3-vectores con letras individuales:

$$\begin{aligned}
\boldsymbol{x} &= (x_1, x_2, x_3) \equiv (a, b, c), \\
\boldsymbol{y} &= (y_1, y_2, y_3) \equiv (p, q, r), \\
\boldsymbol{z} &= (z_1, z_2, z_3) \equiv (u, v, w).
\end{aligned}$$

Se define el **producto vectorial** o **producto cruz** de $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ por esta fórmula:

$$\begin{aligned}
\underline{\boldsymbol{x} \times \boldsymbol{y}} &:= (x_2 y_3 - x_3 y_2, x_3 y_1 - x_1 y_3, x_1 y_2 - x_2 y_1) \\
&= (br - cq, cp - ar, aq - bp).
\end{aligned} \tag{1.6}$$

De ahí se ve que $\boldsymbol{x} \times \boldsymbol{y} = -\boldsymbol{y} \times \boldsymbol{x}$.

Además, el producto cruz combina bien con sumas y múltiplos (Lema 1.9); por ejemplo, vale $\boldsymbol{x} \times (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} \times \boldsymbol{y}) + (\boldsymbol{x} \times \boldsymbol{z})$.

En $\mathbb{R}^3$ se dispone de este producto vectorial y del producto escalar ya visto. Es posible combinar estas dos operaciones en un **producto triple** *de tres vectores*:

$$\underline{[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]} := \boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}) = (\boldsymbol{x} \times \boldsymbol{y}) \cdot \boldsymbol{z}. \tag{1.8}$$

El Lema 1.10 chequea que las dos fórmulas a la derecha coinciden; se obtiene:

$$\begin{aligned} [\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}] &= aqw + bru + cpv - arv - bpw - cqu \\ &= x_1y_2z_3 + x_2y_3z_1 + x_3y_1z_2 - x_1y_3z_2 - x_2y_1z_3 - x_3y_2z_1. \end{aligned}$$

Y ahora una mala noticia: el producto cruz *no es asociativo*:

$$\begin{aligned} \boldsymbol{x} \times (\boldsymbol{y} \times \boldsymbol{z}) &= (\boldsymbol{x} \cdot \boldsymbol{z})\boldsymbol{y} - (\boldsymbol{x} \cdot \boldsymbol{y})\boldsymbol{z}, \\ (\boldsymbol{x} \times \boldsymbol{y}) \times \boldsymbol{z} &= (\boldsymbol{x} \cdot \boldsymbol{z})\boldsymbol{y} - (\boldsymbol{y} \cdot \boldsymbol{z})\boldsymbol{x}, \end{aligned} \tag{1.7}$$

Esto dice que no se puede omitir las paréntesis: la expresión "$\boldsymbol{x} \times \boldsymbol{y} \times \boldsymbol{z}$" es *ambigua*.

Sin embargo, el producto cruz en $\mathbb{R}^3$ sí resulta ser útil, por sus propiedades geométricas. Consideramos, en primer lugar, la *longitud* del vector $\boldsymbol{x} \times \boldsymbol{y}$:

$$\begin{aligned} \|\boldsymbol{x} \times \boldsymbol{y}\|^2 &= (\boldsymbol{x} \times \boldsymbol{y}) \cdot (\boldsymbol{x} \times \boldsymbol{y}) = \boldsymbol{x} \cdot \big(\boldsymbol{y} \times (\boldsymbol{x} \times \boldsymbol{y})\big) \\ &= \boldsymbol{x} \cdot \big((\boldsymbol{y} \cdot \boldsymbol{y})\boldsymbol{x} - (\boldsymbol{y} \cdot \boldsymbol{x})\boldsymbol{y}\big) \\ &= (\boldsymbol{x} \cdot \boldsymbol{x})(\boldsymbol{y} \cdot \boldsymbol{y}) - (\boldsymbol{x} \cdot \boldsymbol{y})(\boldsymbol{y} \cdot \boldsymbol{x}) = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - (\boldsymbol{x} \cdot \boldsymbol{y})^2 \\ &= \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \cos^2\theta = \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \operatorname{sen}^2\theta. \end{aligned}$$

Al tomar la raíz cuadrada (no negativa) de ambos lados, se llega a la fórmula:

$$\|\boldsymbol{x} \times \boldsymbol{y}\| = \|\boldsymbol{x}\|\, \|\boldsymbol{y}\|\, |\operatorname{sen}\theta|. \tag{1.9}$$

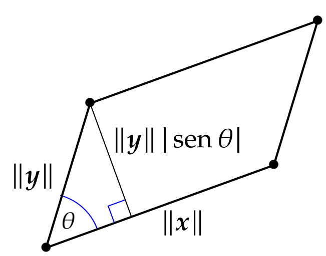

Figura 1.2: Área de un paralelogramo: $\|\boldsymbol{x} \times \boldsymbol{y}\|$

El lado derecho expresa el **área del paralelogramo** con lados adyacentes $\boldsymbol{x}, \boldsymbol{y}$ (Figura 1.2). Y si este paralelogramo es la base de un *paralelepípedo* con un tercer lado adyacente $\boldsymbol{z}$, el *volumen* de ese paralelepípedo es $\underline{\pm[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]}$.

Rectas en $\mathbb{R}^2$

Se supone conocido las fórmulas y cálculos para rectas en el plano cartesiano $\mathbb{R}^2$: lo que comúnmente se llama la *geometría analítica* del plano. Aun así, antes de pasar a $\mathbb{R}^3$, conviene hacer un breve repaso.

Una **recta** en $\mathbb{R}^2$ es – casi por definición – el conjunto de soluciones de una *ecuación de primer grado*:

$$\underline{ax + by + c = 0} \qquad \text{(donde } a^2 + b^2 > 0\text{).}$$

Dos fórmulas muy conocidas son: (a) la recta que pasa por el punto (x_0, y_0) con pendiente m,

$$y - y_0 = m(x - x_0);$$

y (b) la *recta que pasa por dos puntos* distintos (x_1, y_1) y (x_2, y_2):

$$(y_2 - y_1)(x - x_1) - (x_2 - x_1)(y - y_1) = 0. \tag{1.10}$$

Haciendo caso omiso de los problemas de división por 0, se puede reordenar esa ecuación en el formato de una igualdad de dos fracciones:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1} =: t.$$

La cantidad numérica t, el valor común de las dos fracciones, obviamente depende de las coordenadas del punto (x, y). Entonces t es una “variable extra” o *parámetro* para la recta. Es fácil despejar las variables x, y en términos de t:

$$\begin{aligned} x &= x_1 + (x_2 - x_1)t, \\ y &= y_1 + (y_2 - y_1)t, \end{aligned} \qquad \text{o bien:} \qquad \begin{aligned} x &= x_1 + at, \\ y &= y_1 + bt. \end{aligned}$$

Esta es la llamada **forma paramétrica** de las ecuaciones de una recta en $\mathbb{R}^2$. Su formato es de *dos* ecuaciones de primer grado con una variable extra (también en primer grado).

Si se parte de la forma paramétrica, se puede correr el proceso al revés. Primero se despeja t en términos de x o de y, obteniendo el despliegue anterior. Luego *se elimina el parámetro t* (por olvido) y se recupera la “forma simétrica” de la sola ecuación (1.10).

Sobre la primera lista de ejercicios: “Tarea 1”

Ejercicio 1.4 El punto medio de un segmento es el punto que lo divide en dos trozos de igual longitud. Es cuestión de calcular las normas de ciertos vectores.

Ejercicio 1.6 El dibujo sirve para explicar el nombre "ley del paralelogramo". La igualdad se muestra por un cálculo directo.

Ejercicio 1.7 La formula sigue por un cálculo directo. No es necesario deducirla del ejercicio anterior, aunque que eso "resulta posible".

Ejercicio 1.9 Nuevamente, el dibujo está ahí para explicar el nombre de la fórmula.

1.3. Clase 3

Esta es la clase virtual del lunes 12 de abril del 2021. La temática comprende la sección § 1.3: *Rectas y planos en* $\mathbb{R}^2$ *y* $\mathbb{R}^3$.

Rectas en $\mathbb{R}^3$

En la clase anterior, se vio que una recta en $\mathbb{R}^2$ puede ser expresada por ecuaciones de primer grado en las coordenadas cartesianas (x, y) de un punto y posiblemente un parámetro $t \in \mathbb{R}$, de dos modos. Se usa una sola ecuación (1.10) en (x, y) – la "forma simétrica" – o dos ecuaciones que involucran la variable extra t – la "forma paramétrica".

En $\mathbb{R}^3$, conviene empezar con la forma paramétrica. Las tres coordenadas de un punto constituyen un *vector* $\boldsymbol{r} = (x, y, z)$. Un punto específico será $\boldsymbol{r}_0 = (x_0, y_0, z_0)$; una *dirección* fija será otro vector no nulo, $\boldsymbol{v} = (l, m, n)$. Entonces la ecuación vectorial

$$\boldsymbol{r} - \boldsymbol{r}_0 = t\boldsymbol{v} \qquad (\text{donde } t \in \mathbb{R})$$

dice que el segmento de recta entre $\boldsymbol{r}_0$ y $\boldsymbol{r}$ es paralelo (es decir, proporcional) al vector $\boldsymbol{v}$. Al escribirla en las tres coordenadas:

$$\begin{aligned} x &= x_0 + lt, \\ y &= y_0 + mt, \\ z &= z_0 + nt, \end{aligned} \tag{1.12}$$

se obtiene la *forma paramétrica* de la recta: esta vez, *tres* ecuaciones de primer grado en las cuarto variables x, y, z, t.

Se procede a eliminar el parámetro t para obtener la *forma simétrica* de la recta:

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n} \quad (= t) \tag{1.13}$$

dada por *dos* ecuaciones de primer grado entre las variables originales x, y, z.

► En la práctica, rara vez se conoce la dirección de una recta de modo explícito. Si $\boldsymbol{r}_0$ y $\boldsymbol{r}_1$ son dos puntos (distintos) en la recta, se puede tomar $\boldsymbol{v} = \boldsymbol{r}_1 - \boldsymbol{r}_0$. Esto da la ecuación vectorial de la recta que pasa por estos dos puntos:

$$\boldsymbol{r} = (1-t)\,\boldsymbol{r}_0 + t\,\boldsymbol{r}_1. \tag{1.14}$$

El punto variable $\boldsymbol{r}$ divide el segmento $\overline{\boldsymbol{r}_0\,\boldsymbol{r}_1}$ en la proporción $t : (1-t)$ – es negativo si ese punto queda fiera del segmento (Ejercicio 1.22).

Planos en $\mathbb{R}^3$

Un *plano* en el espacio $\mathbb{R}^3$ podría estar dado por uno de sus puntos $\boldsymbol{r}_0$ y por si *inclinación* – la cual se obtiene al dar un *vector normal* $\boldsymbol{n} = (a, b, c)$, perpendicular al plano. La ecuación del plano es entonces

$$\boldsymbol{n}\cdot(\boldsymbol{r}-\boldsymbol{r}_0) = 0.$$

En términos de coordenadas cartesianas, esto es lo mismo que

$$ax + by + cz + d = 0. \tag{1.16}$$

Moraleja: en $\mathbb{R}^3$, *una sola* ecuación de primer grado representa un *plano*, no una recta. De hecho, como una recta requiere dos ecuaciones, se deduce que (siempre en $\mathbb{R}^3$) *una recta es la intersección de dos planos*.

En el espacio $\mathbb{R}^3$, un plano también está determinado por *tres puntos* $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, que no son colineales. Esta condición dice que los vectores $\boldsymbol{r}_1 - \boldsymbol{r}_0$ y $\boldsymbol{r}_2 - \boldsymbol{r}_0$ no son paralelos; en cuyo caso, el vector

$$\boldsymbol{n} = (\boldsymbol{r}_1 - \boldsymbol{r}_0)\times(\boldsymbol{r}_2 - \boldsymbol{r}_0)$$

no es nulo, y es perpendicular al plano. El producto escalar $\boldsymbol{n}\cdot(\boldsymbol{r}-\boldsymbol{r}_0) = (\boldsymbol{r}-\boldsymbol{r}_0)\cdot\boldsymbol{n}$ entonces es cero:

$$\big[\boldsymbol{r}-\boldsymbol{r}_0,\ \boldsymbol{r}_1-\boldsymbol{r}_0,\ \boldsymbol{r}_2-\boldsymbol{r}_0\big] = 0.$$

Esta última fórmula (un poco complicada) expresa la ecuación del plano que pasa por los tres puntos dados. También se puede entenderla como sigue: los tres vectores $\boldsymbol{r}-\boldsymbol{r}_0$, $\boldsymbol{r}_1-\boldsymbol{r}_0$, $\boldsymbol{r}_2-\boldsymbol{r}_0$ son coplanarios si y solo si el paralelepípedo con estos lados adyacentes colapsa en una figura geométrica de volumen cero.

¿Cuál sería el análogo de la fórmula (1.14) para planos: una *forma paramétrica* de exhibir el plano que pasa por tres puntos? Para verlo, primero se introduce en (1.14) un segundo parámetro $s := (1-t)$. Entonces la recta (1.14) se escribe así:

$$\boldsymbol{r} = s\,\boldsymbol{r}_0 + t\,\boldsymbol{r}_1 \quad \text{con} \quad s + t = 1. \tag{1.17a}$$

En el caso del plano, se puede emplear tres parámetros s, t, u, ligados ahora por la relación $s + t + u = 1$.

Entonces la ecuación paramétrica del plano que pasa por tres puntos (no colineales) $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ se expresa fácilmente con tres parámetros ligados:

$$\boldsymbol{r} = s\,\boldsymbol{r}_0 + t\,\boldsymbol{r}_1 + u\,\boldsymbol{r}_2 \quad \text{con} \quad s + t + u = 1. \tag{1.17b}$$

Examinemos esta fórmula: la ecuación vectorial es de primer grado en las variables (x, y, z) y en los parámetros s, t, u; la condición lateral es otra ecuación de primer grado en s, t, u. Por lo tanto, (1.17b) representa un plano.

¿Cuál plano? Si ahora ponemos $(s, t, u) = (1, 0, 0)$, la ecuación se reduce a $\boldsymbol{r} = \boldsymbol{r}_0$. Entonces $\boldsymbol{r}_0$ es un punto del plano, con estos valores de los parámetros. Si tomamos $(s, t, u) = (0, 1, 0)$ o $(s, t, u) = (0, 0, 1)$ respectivamente, se ve que el plano también pasa por $\boldsymbol{r}_1$ y $\boldsymbol{r}_2$.

► Las fórmulas (1.17a) y (1.17b) admiten una generalización a $\mathbb{R}^n$.

Al tomar vectores $\boldsymbol{r}_0, \boldsymbol{r}_1, \dots, \boldsymbol{r}_k$ en $\mathbb{R}^n$ y escalares $t_0, t_1, \dots, t_k \in \mathbb{R}$ que suman 1:

$$\boldsymbol{r} = t_0\,\boldsymbol{r}_0 + t_1\,\boldsymbol{r}_1 + \cdots + t_k\,\boldsymbol{r}_k \quad \text{con} \quad t_0 + t_1 + \cdots + t_k = 1 \tag{1.19}$$

es una **combinación afín** de los vectores dados.

Nótese que algunos de los t_i pueden ser negativos. En el caso especial de que $t_i \geqslant 0$ para cada i se dice que esta es una **combinación convexa**. ⟦ En el caso $k = 2$, las combinaciones convexas representan los puntos del interior y también del borde del triángulo con vértices $\boldsymbol{r}_0$, $\boldsymbol{r}_1$, $\boldsymbol{r}_2$. Las combinaciones con algún $t_j < 0$ corresponden a puntos del plano que son externos al triángulo. ⟧

► Volvamos brevemente a la fórmula paramétrica (1.17b) del plano. Al expandir la ecuación vectorial, se obtiene un juego de cuatro ecuaciones de primer grado:

$$\begin{aligned} x_0\,s + x_1\,t + x_2\,u &= x, \\ y_0\,s + y_1\,t + y_2\,u &= y, \\ z_0\,s + z_1\,t + z_2\,u &= z, \\ s + t + u &= 1. \end{aligned} \tag{1.18}$$

Esto significa dos cosas. En primer lugar, dado un punto (x, y, z) del plano, es posible despejar los valores correspondientes de los parámetros, al resolver este sistema de ecuaciones para las variables s, t, u.

En segundo lugar, como hay 4 ecuaciones para las 3 variables s, t, u, las cantidades al lado derecho no pueden ser arbitrarias: deben estar ligadas por alguna relación algebraica. Esta ligadura es precisamente la ecuación (que resulta ser de primer grado) del plano para las variables x, y, z.

Tales *sistemas de ecuaciones lineales* forman el tema principal del resto de este curso.

Sobre la segunda lista de ejercicios: "Tarea 2"

Ejercicio 1.10 Estos son tres fórmulas clásicas (muy útiles) para el producto cruz.

Ejercicio 1.12 Se trata de obtener la parametrización de un *segmento* de recta, a partir de su definición.

Ejercicio 1.16 Dos rectas no paralelas en $\mathbb{R}^3$ pueden tener un solo punto de intersección, o ninguno: en el segundo caso se llaman "rectas sesgadas".

Ejercicio 1.19 La *distancia* de un punto externo a un plano es la longitud del segmento de la recta perpendicular al plano, desde el punto dado al punto donde la recta perpendicular corta el plano.

Ejercicio 1.21 Un tetraedro regular tiene 4 facetas congruentes (son triángulos equiláteros) y 6 aristas de igual longitud.

Ejercicios 1.22 y 1.23 El *centroide* de un triángulo, o bien de un tetraedro, es el punto en donde todas sus medianas se cortan. Esa *concurrencia* de las medianas, en ambos casos, se puede demostrar con métodos clásicos de la geometría euclidiana; pero es más fácil comprobarla con vectores. La clave del éxito es la división de segmentos de recta en la proporción $t : (1 - t)$. En los dos casos, el centroide resulta ser la media aritmética de los vértices.

2 Comentarios sobre las clases de MA–360, parte 2

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

2.1. Clase 4

Esta es la clase virtual del 15 de abril del 2021. La temática comprende la sección § 2.1: *El concepto de espacio vectorial*.

El cuerpo de escalares

Al inicio, se habló de *vectores* (cantidades con magnitud y dirección) y *escalares* de modo informal, concretando un ejemplo de escalares (números reales) y vectores (elementos de $\mathbb{R}^n$), dotados con ciertas operaciones algebraicas.

Ahora conviene empezar de nuevo, con un punto de vista más abstracta: los escalares y los vectores se toman como objetos cualesquiera que obedecen las reglas de combinación algebraicas que detallaremos a continuación.

En lugar de números racionales o reales (o complejos), se toma un **cuerpo** $\mathbb{F}$ de escalares, con elementos $a, b, c, \dots$ y dos operaciones binarias, *suma* y *producto*, que cumplen las nueve reglas aritméticas siguientes:

$$
\begin{aligned}
a + (b + c) &= (a + b) + c, & a(bc) &= (ab)c, \\
a + b &= b + a, & ab &= ba, \\
\exists\, 0 \in \mathbb{F} \text{ con } 0 + a &= a, & \exists\, 1 \in \mathbb{F} \text{ con } 1a &= a, \\
\forall\, a\ \exists\, (-a) \text{ con } a + (-a) &= 0, & \forall\, a \neq 0\ \exists\, a^{-1} \text{ con } aa^{-1} &= 1,
\end{aligned}
$$
$$
a(b + c) = ab + ac.
$$

Fíjese que $0a = (0 + 0)a = 0a + 0a \implies 0a = 0$ para $a \in \mathbb{F}$. Se exige que $1 \neq 0$.

El conjunto $\mathbb{N} := \{0, 1, 2, 3, \dots\}$ no es un cuerpo, porque $2 \in \mathbb{N}$ pero $(-2) \notin \mathbb{N}$.

El conjunto $\mathbb{Z} := \{\dots, -2, -1, 0, 1, 2, \dots\}$ tampoco, porque $2 \in \mathbb{Z}$ pero $\frac{1}{2} \notin \mathbb{Z}$.

Ejemplos de cuerpos:

(a) Los números racionales $\mathbb{Q} = \{\, p/q : p, q \in \mathbb{Z}, q \neq 0 \,\}$.

(b) Los números reales $\mathbb{R}$ forman un cuerpo. Nótese que $\mathbb{Q} \subset \mathbb{R}$.

(c) El *cuerpo binario* $\underline{\mathbb{F}_2 := \{0, 1\}}$, en donde $\underline{1 + 1 = 0}$.

► Otro ejemplo es el cuerpo de los números complejos

$$\mathbb{C} := \{\, a + bi : a, b \in \mathbb{R} \,\}, \tag{2.1}$$

con la regla $\underline{i^2 = -1}$. Las reglas de suma y producto son:

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i, \\ (a + bi)(c + di) &= (ac - bd) + (ad + bc)i, \end{aligned}$$

pero la división requiere un poco de manejo de fracciones:

$$\frac{1}{a + bi} = \frac{a - bi}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}\, i.$$

Vale la pena señalar (o recordar) dos particularidades del cuerpo $\mathbb{C}$:

- ⋄ El *conjugado complejo* de $\underline{z = a + bi}$ es $\underline{\bar{z} := a - bi}$. Nótese que $z\bar{z} = a^2 + b^2$.
- ⋄ El *valor absoluto* de z es $|z| := \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$ (real, con $|z| \geqslant 0$).

Espacio vectorial: definición

Un **espacio vectorial** sobre un cuerpo $\mathbb{F}$ es un conjunto V (no vacío) con elementos $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \dots$, los **vectores**. Este conjunto V admite dos operaciones: *suma* $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} + \boldsymbol{y}$; y *multiplicación escalar* $(c, \boldsymbol{x}) \mapsto c\boldsymbol{x} \in V$, donde $c \in \mathbb{F}$. Se debe cumplir las ocho reglas aritméticas siguientes:

$$\begin{aligned} \boldsymbol{x} + \boldsymbol{y} &= \boldsymbol{y} + \boldsymbol{x}, & a(\boldsymbol{x} + \boldsymbol{y}) &= a\boldsymbol{x} + a\boldsymbol{y}, \\ \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) &= (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}, & (a + b)\boldsymbol{x} &= a\boldsymbol{x} + b\boldsymbol{x}, \\ \exists\, \mathbf{0} \in V \text{ con } \mathbf{0} + \boldsymbol{x} &= \boldsymbol{x}, & a(b\boldsymbol{x}) &= (ab)\boldsymbol{x}, \\ \forall\, \boldsymbol{x}\ \exists -\boldsymbol{x} \in V \text{ con } \boldsymbol{x} + (-\boldsymbol{x}) &= \mathbf{0}, & 1\boldsymbol{x} &= \boldsymbol{x}. \end{aligned}$$

Nuevamente, $0\boldsymbol{x} = (0 + 0)\boldsymbol{x} = 0\boldsymbol{x} + 0\boldsymbol{x} \implies 0\boldsymbol{x} = \mathbf{0}$ para $\boldsymbol{x} \in V$.

Se escribe $\boldsymbol{x} - \boldsymbol{y} := \boldsymbol{x} + (-\boldsymbol{y})$, naturalmente.

El ejemplo más obvio es $\underline{V = \mathbb{R}^n}$, con $\mathbb{F} = \mathbb{R}$.

El Ejemplo 2.8 generaliza esto a $\underline{V = \mathbb{F}^n}$ sobre *cualquier* cuerpo de escalares $\mathbb{F}$. Se hacen sumas y múltiplos de las "n-tuplas" $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ coordenada por coordenada, de igual modo que en $\mathbb{R}^n$.

En particular, con $\mathbb{F} = \mathbb{F}^2$ y $n = 8$, los vectores en $\mathbb{F}_2^8$ son *bytes*. Cada byte es una fila de ocho *bits*, como $(0\,1\,1\,1\,0\,0\,1\,0)$ – esta es la representación binaria del entero decimal 114. La suma "bitwise" en $\mathbb{F}^8$ no coincide con la suma análoga en $\mathbb{N}$.

Otros ejemplos de espacios vectoriales

Los *polinomios* con coeficientes en $\mathbb{R}$, en una *incógnita* t, son elementos de un espacio vectorial (sobre $\mathbb{R}$). La totalidad de tales polinomios es $\underline{\mathbb{R}[t]}$. Un típico polinomio no nulo es

$$\underline{p(t)} = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \quad \text{con cada} \quad a_k \in \mathbb{R},\ a_n \neq 0.$$

Este n es el *grado* de $p(t)$. Se agrega el polinomio nulo, denotado 0. La suma y la multiplicación escalar son "obvias". (Un polinomio de grado 0 es una "constante" $p(t) = a_0$.)

Muchas veces es necesario limitar el grado de un juego polinomios, imponiendo un grado máximo, n. Entonces se considera

$$\mathbb{R}_n[t] := \{\, p(t) \in \mathbb{R}[t] : \text{grado}\, p(t) \leqslant n \,\}.$$

Este es un "subespacio vectorial" de $\mathbb{R}[t]$.

De hecho, $\mathbb{R}_n[t]$ es simplemente $\mathbb{R}^{n+1}$, levemente disfrazado. Se puede identificar el polinomio $p(t)$ anterior con la $(n+1)$-tupla $(a_0, a_1, a_2, \dots, a_n)$; y se nota que las reglas de suma y múltiplos coinciden bajo esa identificación.

► Las *funciones continuas* $f : [a,b] \to \mathbb{R}$ son elementos de un espacio vectorial (sobre $\mathbb{R}$) llamado $\underline{C[a,b]}$. Las operaciones vectoriales sobre funciones se definen *punto por punto*:

$$\underline{f+g}(t) := f(t) + g(t), \qquad \underline{cf}(t) := cf(t),$$

Es fácil verificar que las nuevas funciones $f+g$ y cf son también continuas.

► No se debe olvidar *el espacio vectorial trivial* (sobre $\mathbb{F}$). Este es el singulete $\{\mathbf{0}\}$ con las reglas vectoriales $\mathbf{0} + \mathbf{0} = \mathbf{0}$; y $c\,\mathbf{0} = \mathbf{0}$ para $c \in \mathbb{F}$.

► Otro método para formar espacios vectoriales es el de tomar un **subespacio** W de un espacio vectorial V preexistente. Este es una *parte* $W \subseteq V$ conservado por las reglas vectoriales:

$$\boldsymbol{x} + \boldsymbol{y} \in W, \quad c\boldsymbol{x} \in W, \quad \text{toda vez que} \quad \boldsymbol{x}, \boldsymbol{y} \in W,\ \ c \in \mathbb{F}.$$

Por ejemplo, *una recta* en $\mathbb{R}^3$ *que pasa por el origen* $(0,0,0)$; o bien *un plano que contiene el origen*, son dos casos de subespacios de $\mathbb{R}^3$. (Para que $W \subseteq V$ sea un subespacio, siempre es necesario que $\mathbf{0} \in W$.)

Se escribe $W \leqslant V$ para denotar que W es un subespacio de V.

2.2. Clase 5

Esta es la clase virtual del 19 de abril del 2021. La temática comprende las secciones § 2.2: *Independencia lineal* y § 2.3: *Bases y dimensión de un espacio vectorial*.

Combinaciones lineales

En un espacio vectorial V (sobre un cuerpo $\mathbb{F}$) las operaciones permitidos son: (i) multiplicar un vector $\boldsymbol{x}$ por un escalar a, para formar un nuevo vector $a\boldsymbol{x}$; y (ii) sumar dos o más vectores. Combinando los dos operaciones, se puede formar las siguientes **combinaciones lineales**:

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m, \tag{2.2}$$

a partir de los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in V$ e igual cantidad de **coeficientes** $a_1, \dots, a_m \in \mathbb{F}$.

Una combinación lineal que parece particularmente inútil es la que se forma con todas los coeficientes iguales a 0:

$$0\boldsymbol{x}_1 + 0\boldsymbol{x}_2 + \cdots + 0\boldsymbol{x}_m = \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}.$$

El resultado es el vector nulo y esta combinación lineal se llama "trivial". Pero ahora considérese el problema inverso: si vale

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m = \mathbf{0}, \tag{2.3}$$

¿se puede concluir que $a_1 = a_2 = \cdots = a_m = 0$? La respuesta a esa pregunta depende del juego de vectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\}$.

► Se dice que los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ son **linealmente dependientes** si hay algunas coeficientes, *no todos iguales a* 0, que forman la combinación lineal nula (2.3).

El Lema 2.15 dice que este ocurre si y solo si *uno* de estos vectores *es una combinación lineal de los demás*. Si, por ejemplo, $a_1 \neq 0$, entonces se puede despejar $\boldsymbol{x}_1$ como una combinación lineal de $\boldsymbol{x}_2, \dots, \boldsymbol{x}_m$:

$$\boldsymbol{x}_1 = -\frac{a_2}{a_1}\boldsymbol{x}_2 - \frac{a_3}{a_1}\boldsymbol{x}_3 - \cdots - \frac{a_m}{a_1}\boldsymbol{x}_m$$

y si $a_2 \neq 0$ se puede despejar $\boldsymbol{x}_2$ en términos de los otros; etc.

Pero a veces eso no es posible. Si el único juego de coeficientes que hace cumplir (2.3) es $a_1 = a_2 = \cdots = a_m = 0$, se dice que el conjunto de vectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\}$ se llama **linealmente independiente**.

En el espacio $\mathbb{R}^3$, por ejemplo, tres vectores $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ son linealmente dependientes si y solo son coplanarios, en un plano que pasa por el origen. Si son coplanarios y dos de ellos determinan un plano, el tercero también queda en ese plano (por ser una combinación lineal de los primeros dos). En el caso contrario, los tres vectores independientes $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ determinan tres planos diferentes, que solo tienen el origen en común.

El caso de un singulete, con $m = 1$: el conjunto $\{\mathbf{0}\}$ es linealmente *de*pendiente, porque la ecuación $1\mathbf{0} = \mathbf{0}$ es una posible solución no trivial de (2.3). En cambio, si $\boldsymbol{x} \neq \mathbf{0}$, la ecuación $a\boldsymbol{x} = \mathbf{0}$ solo es posible cuando $a = 0$; en consecuencia, $\{\boldsymbol{x}\}$ es linealmente *inde*pendiente cuando $\boldsymbol{x} \neq \mathbf{0}$.

► Después de esas consideraciones generales, consideremos unos ejemplos. En el espacio $\mathbb{F}^n$ (Ejemplo 2.17), hay n "vectores básicos":

$$\boldsymbol{e}_1 := (1, 0, 0, \dots, 0), \quad \boldsymbol{e}_2 := (0, 1, 0, \dots, 0), \ \dots, \quad \boldsymbol{e}_n := (0, 0, \dots, 0, 1).$$

Es fácil ver que las combinaciones lineales de ellos son todos los elementos de $\mathbb{F}^n$:

$$a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + \cdots + a_n\boldsymbol{e}_n = (a_1, a_2, \dots, a_n).$$

El vector cero de $\mathbb{F}^n$ es $\mathbf{0} = (0, 0, \dots, 0)$. Aquí (2.3) sólo se cumple si cada $a_k = 0$: los vectores básicos $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ son linealmente independientes en $\mathbb{F}^n$.

Ejemplo 2.19: en el espacio de polinomios, un juego de *monomios* $\{1, t, t^2, \dots, t^m\}$ es linealmente independiente. (Se pone $t^0 := 1$.) Una combinación lineal de monomios es simplemente un *polinomio*:

$$p(t) = a_0 + a_1t + a_2t^2 + \cdots + a_mt^m$$

y este es el polinomio nulo ($p(t) = 0$) solo si cada $a_k = 0$.

► Dado un juego de vectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\} \subset V$, la totalidad de sus combinaciones lineales forman un *subespacio* W de V:

$$\begin{gathered}(a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m) + (b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_m\boldsymbol{x}_m) \\ = (a_1 + b_1)\boldsymbol{x}_1 + (a_2 + b_2)\boldsymbol{x}_2 + \cdots + (a_m + b_m)\boldsymbol{x}_m, \\ c(a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m) = ca_1\boldsymbol{x}_1 + ca_2\boldsymbol{x}_2 + \cdots + ca_m\boldsymbol{x}_m.\end{gathered}$$

Este W se llama el **subespacio generado** por los vectores dados:

$$W = \operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle. \tag{2.4}$$

En $\mathbb{R}^3$, por ejemplo, se ve que $\operatorname{lin}\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle$ es el subespacio con elementos $(a_1, a_2, 0)$. Este es el plano $z = 0$.

Bases de un espacio vectorial

Al combinar los dos conceptos anteriores, se llega al concepto de una **base** en un espacio vectorial V (sobre $\mathbb{F}$). Este es una parte cualquiera $X \subset V$ tal que:

(a) X **genera** V, esto es, $\operatorname{lin}\langle X\rangle = V$;

(b) X es **linealmente independiente**.

En la mayoría de los ejemplos en este curso, X será un conjunto finito.

Ejemplo: en $\mathbb{F}^n$, el juego de los vectores básicos $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ es una base. Esta es la **base estándar** de $\mathbb{F}^n$.

Ejemplo 2.24: los monomios $\{1, t, t^2, \dots, t^n\}$ definen una base de $\mathbb{R}_n[t]$: cualquier polinomio de grado $\leqslant n$ es – obviamente – una combinación lineal de ellos.

Ejemplo 2.25: para generar el espacio vectorial $\mathbb{R}[t]$ de *todos* los polinomios reales, se requiere usar *todos* los monomios $\{1, t, t^2, \dots, t^n, \dots\}$, un conjunto infinito. Pero una combinación lineal de entre ellos es – por definición – una *suma finita* de monomios, el cual es otro polinomio. Este es un ejemplo de un espacio vectorial con una base infinita (pero contable).

► La Prop. 2.26 señala un propiedad: una base de V es un conjunto linealmente independiente *maximal*. En efecto: si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V y si $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in V$ con $\underline{m > n}$, los vectores $\boldsymbol{y}_k$ deben ser linealmente *dependientes*.

La prueba trata (sin éxito, eventualmente) de formar una nueva base con los vectores $\boldsymbol{y}_k$, asumiendo – por ahora – que $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ es linealmente *independiente*.

El primer paso es fácil: como $\boldsymbol{y}_1 \neq \boldsymbol{0}$, se puede expresar como una combinación lineal no trivial de los $\boldsymbol{x}_j$:

$$\boldsymbol{y}_1 = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n.$$

Algún $a_j \neq 0$; por ejemplo, si $a_1 \neq 0$, se puede despejar

$$\boldsymbol{x}_1 = \frac{1}{a_1}\,\boldsymbol{y}_1 - \frac{a_2}{a_1}\,\boldsymbol{x}_2 - \cdots - \frac{a_n}{a_1}\,\boldsymbol{x}_n,$$

así que $\boldsymbol{x}_1 \in \operatorname{lin}\langle \boldsymbol{y}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n\rangle$; entonces $\operatorname{lin}\langle \boldsymbol{y}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n\rangle = V$.

En seguida, se expresa $\boldsymbol{y}_2$ en términos de estos vectores:

$$\boldsymbol{y}_2 = a'_1\boldsymbol{y}_1 + a'_2\boldsymbol{x}_2 + \cdots + a'_n\boldsymbol{x}_n.$$

y si $a'_2 \neq 0$, se puede despejar $\boldsymbol{x}_2$. Entonces se llega a que $\operatorname{lin}\langle \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_n \rangle = V$.

Se puede continuar así; después de n repeticiones del proceso se arriba a la igualdad $\operatorname{lin}\langle \boldsymbol{y}_1, \dots, \boldsymbol{y}_n \rangle = V$.

Pero entonces – desastre! Como $m > n$, se puede tomar un paso más:

$$\boldsymbol{y}_{n+1} = d_1\boldsymbol{y}_1 + d_2\boldsymbol{y}_2 + \cdots + d_n\boldsymbol{y}_n,$$

arruinando la supuesta independencia lineal de $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$. Entonces los $\boldsymbol{y}_k$ son dependientes, después de todo. □

Dimensión de un espacio vectorial

La Prop. 2.26 anterior muestra que, si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ son *dos bases diferentes* del mismo espacio vectorial V, entonces $m \leqslant n$. Y simétricamente, vale $n \leqslant m$. Conclusión: se cumple $\underline{m = n}$.

En otras palabras: *el número de vectores en una base* es el mismo para todas las bases de V. Este número se llama la **dimensión** de V, escrito $\underline{\dim V}$.

Ejemplos:

$$\dim \mathbb{F}^n = n, \qquad \dim \mathbb{R}_n[t] = n + 1, \qquad \dim \mathbb{R}[t] = \infty.$$

2.3. Clase 6

Esta es la clase virtual del 26 de abril del 2021. La temática comprende la sección §2.3: *Bases y dimensión de un espacio vectorial* y el inicio del Capítulo 3.

Compleción de una base parcial

Sea V un espacio vectorial (sobre un cuerpo $\mathbb{F}$), de dimensión finita n. Esto quiere decir que cualquier base de V tiene exactamente n vectores. A veces se dispone de una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ previamente dada; pero otras veces, habrá que buscar – o construir – alguna base para V.

Los vectores de una base son linealmente independientes (por la definición de base). Cualquier parte de una base es también un juego de vectores linealmente independientes. Inversamente, si se dispone de unos m vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ linealmente independientes, con $\underline{m < n}$, cabe preguntar si forman parte de una base?

La Prop. 2.29 *dice que sí*: siempre es posible hallar *otros vectores* $\boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n$ tales que, juntos, $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\}$ es una base de V.

La prueba es sencilla: como el subespacio $\mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle$ no es todo V, debe haber algún otro vector $\boldsymbol{x}_{m+1} \in V$ tal que $\boldsymbol{x}_{m+1} \notin \mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle$. Como este vector no es una combinación lineal de los vectores dados (en particular, no es $\mathbf{0}$), el conjunto $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m, \boldsymbol{x}_{m+1}\}$ es linealmente independiente.

Y ahora se repite el argumento para hallar $\boldsymbol{x}_{m+2} \notin \mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_{m+1} \rangle$; y luego $\boldsymbol{x}_{m+3}$, etcétera. Cuando se llega a $\boldsymbol{x}_n$, el subespacio $\mathrm{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_n \rangle$ tiene dimensión n y por eso es todo V. Por lo tanto, $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V.

Obviamente, hay muchas maneras de elegir los vectores nuevos: las base así obtenida *no es única*. Pero lo importante es que siempre hay al menos una manera de completar la "base parcial" dada.

Para ver cómo se usa esta Prop. 2.29, fíjese en su Corolario 2.30: si $W \leqslant V$ es un subespacio, entonces $\dim W \leqslant \dim V$. En efecto: si $\dim W = m$, tómese alguna base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ de W. Esto es un conjunto linealmente independiente en V; si no genera V, es cuestión de agregarle otros vectores para obtener una base de V, con $n = \dim V$ elementos.

Sobre la tercera lista de ejercicios: "Tarea 3"

Ejercicio 2.1 Los residuos bajo división por un entero $m \geqslant 2$ forma un conjunto finito $\mathbb{F}_m$ en el cual es posible definir una suma y un producto. Para algunos m, este $\mathbb{F}_m$ es un cuerpo; para otros, no.

Ejercicio 2.2 La colección $\mathcal{P}(X)$ de las partes de un conjunto X resulta ser un espacio vectorial sobre el cuerpo finito $\mathbb{F}_2$. La *suma* de A y B es la llamada "diferencia simétrica" $A \# B$. El cero de $\mathcal{P}(X)$ es el conjunto vacío $\emptyset$.

Ejercicio 2.7 Se sabe que un polinomio real de grado n tiene a lo sumo n raíces distintas. (Cabe recordar el "teorema del factor"). Esto sirve para demostrar un caso de independencia lineal.

Ejercicios 2.9 y 2.10 Ciertos subespacios de $\mathbb{R}^n$ se caracterizan por ser conjuntos de soluciones de ecuaciones. En cada caso, se debe demostrar que tales conjuntos forman un subespacio; o bien indicar por qué no son un subespacio, si eso fuera el caso.

Ejercicio 2.11 Un determinado subespacio puede ser generado por diversos conjuntos de vectores. En cada caso, hay que mostrar que los vectores de un conjunto son combinaciones lineales de los vectores del otro.

3 Comentarios sobre las clases de MA–360, parte 3

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

3.1. Clase 6, continuación

Esta es la segunda parte de la clase virtual del 26 de abril del 2021. La temática comprende la sección § 3.1: *El álgebra de matrices*.

Las matrices son arreglos rectangulares de escalares

Una **matriz** $m \times n$ es un rectángulo de números (*m filas, n columnas*):

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

cuyas entradas son escalares $a_{ij} \in \mathbb{F}$. Se escribe $\underline{A = [a_{ij}]}$ como abreviatura.

Se definen operaciones vectoriales "entrada por entrada", por analogía con $\mathbb{F}^n$:

$$[a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}], \qquad c[a_{ij}] = [ca_{ij}].$$

Ellas forman un *espacio vectorial* $\underline{M_{m\times n}(\mathbb{F})}$.

Si $m = n$, A es una **matriz cuadrada** $n \times n$ y se escribe $\underline{M_n(\mathbb{F})}$ en vez de $M_{n\times n}(\mathbb{F})$.

Al intercambiar filas y columnas (Defn 3.2) se obtiene la **transpuesta** A^{t}, la cual es una matriz $n \times m$:

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

Una matriz cuadrada $A \in M_n(\mathbb{F})$ es **simétrica**, $\underline{A^{\mathsf{t}} = A}$, si $a_{ji} = a_{ij}$ para todo i, j.

Una sola columna de A se considera un *vector de columna* en $\mathbb{F}^m$:

$$A = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_n], \quad \text{donde} \quad \boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}. \tag{3.1}$$

Una sola fila de A se considera un *vector de fila* en $\mathbb{F}^n$: $\boldsymbol{a}'_i = \begin{bmatrix} a_{i1}\ a_{i2}\ \cdots\ a_{in} \end{bmatrix}$.

El producto de dos matrices

Algunas pares de matrices A y B dan lugar a una *matriz producto* $\underline{C = AB}$. Cada entrada de C es el *producto punto* de una *fila de A* por una *columna de B*:

$$c_{ij} := \boldsymbol{a}'_i \cdot \boldsymbol{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj}. \tag{3.2}$$

Esto es factible si y solo si las filas de A y las columnas de B tiene igual número de entradas. Si A es $m \times n$ y B es $n \times r$, entonces $C = AB$ es una matriz $n \times r$. Por ejemplo:

$$\begin{bmatrix} 2 & 0 & -1 & 1 \\ 0 & 3 & 1 & -2 \\ 1 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 2 & 1 \\ 0 & -2 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 8 & -3 \\ -2 & -1 \end{bmatrix}.$$

La entrada $(2, 1)$ de este producto matricial es

$$c_{21} = \boldsymbol{a}'_2 \cdot \boldsymbol{b}_1 = \begin{bmatrix} 0 & 3 & 1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix} = 0(1) + 3(2) + 1(0) + (-2)(-1) = 0 + 6 + 0 + 2 = 8.$$

La Prop. 3.5 explica las reglas algebraicas que involucran el producto de matrices.

(a) *Asociatividad*: $A(BC) = (AB)C$,

(b) *Distributividad*: $(A + D)B = AB + DB$ y $A(B + E) = AB + AE$,

(c) *La transpuesta revierte el orden del producto*: $(AB)^{\mathsf{t}} = B^{\mathsf{t}} A^{\mathsf{t}}$.

Sin embargo,

(d) *El producto* no es *conmutativo*: $AB \neq BA$ en general:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{pero} \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Afortunadamente, para cada tamaño n existe una **matriz identidad** $\underline{1_n \in M_n(\mathbb{F})}$ tal que $A1_n = A = 1_n A$ para todo $A \in M_n(\mathbb{F})$.

Esta matriz identidad tiene entradas 1 en la diagonal y 0 fuera de la diagonal:

$$1_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{3.4}$$

► $M_n(\mathbb{F})$ no es un cuerpo (excepto si $n = 1$) porque no toda matriz $A \in M_n(\mathbb{F})$ tiene una **matriz inversa** $B \in M_n(\mathbb{F})$ tal que $\underline{AB = 1_n = BA}$. Si una tal B existe, resulta ser única y se denota por A^{-1}:

$$A\,A^{-1} = 1_n = A^{-1}A.$$

En el caso de matrices 2×2, se puede dar una fórmula explícita para la matriz inversa (Ejemplo 3.7). Esto se basa en el siguiente cálculo:

$$AC := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix}.$$

Si $ad - bc = 0$, el lado derecho es la *matriz nula* $0_2 := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ y no puede haber un inverso para A. En cambio, si $ad - bc \neq 0$, se obtiene la *fórmula*:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{3.5}$$

Por ejemplo,

$$\begin{bmatrix} 2 & 1 \\ -2 & 3 \end{bmatrix}^{-1} = \frac{1}{8} \begin{bmatrix} 3 & -1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & -\frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

A las reglas algebraicas se puede añadir una más:

(e) *La inversión revierte el orden del producto*:

$$(AB)^{-1} = B^{-1}A^{-1} \qquad \text{cuando } A^{-1}, B^{-1} \text{ existen.} \tag{3.6}$$

Esto se demuestra con dos cálculos que usan la asociatividad del producto:

$$\begin{aligned} (B^{-1}A^{-1})AB &= B^{-1}(A^{-1}A)B = B^{-1}1_nB = B^{-1}B = 1_n, \\ AB(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = A1_nA^{-1} = AA^{-1} = 1_n. \end{aligned}$$

Moraleja: la conmutatividad de productos no es necesaria para obtener una teoría algebraica útil e interesante.

3.2. Clase 7

Esta es la clase virtual del 29 de abril del 2021. La temática comprende la sección § 3.2: *Sistemas de ecuaciones lineales*.

Sistemas de ecuaciones lineales

Las matrices tienen dos fuentes históricas. La segunda es una serie de trabajos de los años 1840s por el matemático inglés Arthur Cayley acerca de las "transformaciones homográficas" del plano complejo:

$$z \longmapsto \frac{az+b}{cz+d}\,, \qquad \text{toda vez que} \quad ad - bc \neq 0.$$

Dos de estas transformaciones se combinan mediante el producto de las $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ para cada una. (James Joseph Sylvester, un colega de Cayley, inventó la palabra *matriz*.)

La primera fuente, mucha más antigua, es la búsqueda de un método sistemático para resolver ecuaciones de primer grado en varias variables. (Hay un ensayo escrito sobre esto en la época de la dinastía Han en China, por ahí del el segundo siglo a.C.)

Se trata de reducir un juego de m ecuaciones lineales en n variables $x_1, x_2, \dots, x_n$, a unas $(m-1)$ ecuaciones en las $(n-1)$ variables $x_2, \dots, x_n$, *eliminando* la primera variable x_1.

Si $m = n$, es cuestión de intentar eliminar x_1, x_2, etc. en forma consecutiva, hasta llegar eventual mente a una sola ecuación para x_n:

$$a\,x_n = b \implies x_n = \frac{b}{a} \quad \text{(siempre que} \quad a \neq 0).$$

La penúltima ecuación permite despejar x_{n-1}:

$$c\,x_{n-1} + d\,x_n = e \implies c\,x_{n-1} + \frac{bd}{a} = e \implies x_{n-1} = \frac{ae - bd}{ac}$$

(siempre que $c \neq 0$ también). Y se puede seguir despejando $x_{n-2}, x_{n-3}, \dots, x_1$ por "sustitución regresiva".

► Ya se sabe manejar dos ecuaciones lineales para dos variables; por ejemplo,

$$\begin{aligned} x + 2y &= 5 \\ 4x + y &= 6 \end{aligned} \implies \begin{aligned} x + 2y &= 5 \\ -7y &= -14 \end{aligned} \implies \begin{aligned} x + 2y &= 5 \\ y &= 2 \end{aligned} \implies \begin{aligned} x &= 1 \\ y &= 2 \end{aligned}$$

(primero se elimina x, luego se despeja y, después se despeja x.)

► Para manejar tres o más variables, es preferible nombrar las variables con unos subíndices: $x_1, x_2, x_3, \dots$ El ejemplo de los apuntes es:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ 3x_1 + 4x_2 - 4x_3 &= 7 \\ 3x_1 + 6x_2 + 2x_3 &= -3 \end{aligned} \tag{3.7a}$$

Primer paso: eliminar x_1. Si se denotan estas ecuaciones por E_1, E_2, E_3, se trata de *restar de* E_2 *y* E_3 unos múltiplos apropiados de E_1.

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ -5x_2 - x_3 &= 4 \qquad E_2 - 3\,E_1 \\ -3x_2 + 5x_3 &= -6 \qquad E_3 - 3\,E_1 \end{aligned}$$

Segundo paso: eliminar x_2. Se trata de restar de E_3 un múltiplo apropiado de E_2,

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ -5x_2 - x_3 &= 4 \\ \tfrac{28}{5}x_3 &= -\tfrac{42}{5} \qquad E_3 - \tfrac{3}{5}\,E_2 \end{aligned} \tag{3.7b}$$

Último paso: despejar x_3, x_2, x_1 en ese orden: *sustitución regresiva*:

$$x_3 = -\tfrac{3}{2}, \qquad x_2 = -\tfrac{1}{2}, \qquad x_1 = 1.$$

► Ahora podemos reformular el sistema de ecuaciones (3.7a) en el lenguaje de matrices, así:

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix}.$$

A la izquierda, se ve el producto de una matriz 3×3 por un vector columna:

$$A\,\boldsymbol{x} = \boldsymbol{b} \tag{3.8}$$

donde

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix}.$$

Debe ser obvio que toda la información del sistema está contenida en las entradas numéricas de A y $\boldsymbol{b}$. La columna $\boldsymbol{x}$ de las variables es solamente nomenclatura.

► Sin embargo, las variables x_1, x_2, x_3 sí juegan un papel interesante. Otra manera de interpretar el sistema (3.7a) es la siguiente:

$$x_1 \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -3 \end{bmatrix}.$$

Al recordar que las *columnas* de A son $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, esto toma el formato:

$$x_1\, \boldsymbol{a}_1 + x_2\, \boldsymbol{a}_2 + x_3\, \boldsymbol{a}_3 = \boldsymbol{b}. \tag{3.9}$$

En otras palabras: el sistema de ecuaciones tiene una solución *si y solo si* la columna $\boldsymbol{b}$ pertenece al *subespacio generado por las columnas* $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ de A.

(Como además las columnas $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ son linealmente independientes, como se puede verificar, las *coeficientes* x_1, x_2, x_3 de la combinación lineal que produce $\boldsymbol{b}$ son *únicos* – por el Ejercicio 2.16. Esto garantiza que el sistema de ecuaciones (3.7a) tiene *solución única*.)

Sobre la cuarta lista de ejercicios: "Tarea 4"

Ejercicio 2.13 Un espacio vectorial dado puede tener muchas bases diferentes.

Ejercicio 2.16 Si un juego de vectores genera un espacio vectorial, cualquier vector es una combinación lineal de ellos. Si este juego de vectores forma una *base*, se pide comprobar que los coeficientes de esa combinación lineal están determinados *unívocamente*.

Ejercicio 2.18 Dos subespacios M y N de V dan lugar a otros dos subespacios, su intersección $M \cap N$ y su *suma* $M + N$. Se pide comprobar una fórmula que conecta las dimensiones de estos cuatro subespacios, por el método de compleción de bases.

Ejercicio 2.19 Dadas dos espacios vectoriales V y W, se puede fabricar una suma (abstracta) de ellos de tal manera que la intersección es nula: esto se llama la *suma directa* $V \oplus W$. Como alternativa, una suma $M + N$ de subespacios M y N de V que cumplen $M \cap N = \{0\}$ también se llama "suma directa" y se escribe $M \oplus N$ en vez de $M + N$. (El ejercicio siguiente indica que se trata de la misma cosa.)

Ejercicio 2.20 Si dos subespacios U y W de V tienen intersección nula $U \cap W = \{0\}$ y suma total $U + W = V$, de manera que se puede identificar la suma directa $U \oplus W$ con V, se dice que W es un *suplemento* de V (y viceversa). Un subespacio U puede tener muchos suplementos W diferentes.

3.3. Clase 8

Esta es la clase virtual del 6 de mayo del 2021. La temática comprende la sección § 3.3: *El algoritmo de eliminación gaussiana*.

Operaciones de fila sobre una matriz aumentada

Es oportuno recordar el sistema de ecuaciones lineales de la clase anterior:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1 \\ 3x_1 + 4x_2 - 4x_3 &= 7 \\ 3x_1 + 6x_2 + 2x_3 &= -3. \end{aligned} \tag{3.7a}$$

Si $\boldsymbol{x} = (x_1, x_2, x_3)$ es el vector de las incógnitas, escrito ahora como vector de columna, el sistema se puede resumir en una sola ecuación matricial:

$$A\,\boldsymbol{x} = \boldsymbol{b}. \tag{3.8}$$

Toda la información del sistema queda incorporada en la **matriz aumentada**:

$$[A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 3 & 4 & -4 & 7 \\ 3 & 6 & 2 & -3 \end{array}\right].$$

en donde cada *fila* de $[A \mid \boldsymbol{b}]$ representa una *ecuación* del sistema. Las operaciones legales sobre ecuaciones individuales se transmitan en **operaciones de fila**.

Estas operaciones se clasifican en tres tipos:

(a) $F_i \leftrightarrow F_k$: *intercambiar dos filas*;

(b) $F_i \mapsto c\,F_i$: *multiplicar una fila por un escalar* $c \neq 0$;

(c) $F_i \mapsto F_i - c\,F_k$: *sustraer de una fila un múltiplo de cualquier otra fila*.

Una cadena de estas operaciones puede transformar $[A \mid \boldsymbol{b}]$ en una nueva matriz aumentada $[V \mid \boldsymbol{c}]$ que posee las mismas soluciones: $V\boldsymbol{x} = \boldsymbol{c}$. Si se logra que V tiene forma triangular, el nuevo sistema $V\boldsymbol{x} = \boldsymbol{c}$ se puede resolver rápidamente por "sustitución regresiva".

► La ventaja teórica de trabajar con matrices aumentadas es que cada una de estas operaciones de fila (y las composiciones de varias operaciones) se efectúan con ciertos *productos de matrices* a la izquierda. Este análisis permite *factorizar* las matrices A y $[A \mid \boldsymbol{b}]$ en matrices más manejables.

Sea $[A \mid \boldsymbol{b}]$ una matriz aumentada $m\times(n+1)$. Las operaciones de fila corresponden a las siguientes **premultiplicaciones** por matrices $m \times m$:

(a) $F_i \leftrightarrow F_k$: $[A \mid \boldsymbol{b}] \mapsto P_{ik}\,[A \mid \boldsymbol{b}]$. Ejemplo con $m = 5$:

$$P_{23} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.12a}$$

(b) $F_i \mapsto c\,F_i$: $[A \mid \boldsymbol{b}] \mapsto M_i(c)\,[A \mid \boldsymbol{b}]$. Ejemplo con $m = 5$:

$$M_4(9) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.12b}$$

(c) $F_i \mapsto F_i - c\,F_k$: $[A \mid \boldsymbol{b}] \mapsto R_{ik}(c)\,[A \mid \boldsymbol{b}]$. Ejemplo con $m = 5$:

$$R_{42}(3) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.12c}$$

También es muy notable que cada una de estas operaciones de fila son *reversibles*: las matrices en las fórmulas (3.12) son invertibles, y de hecho sus inversos son matrices del mismo estilo, en cada caso:

$$P_{ik}^{-1} = P_{ik}, \qquad M_i(c)^{-1} = M_i(1/c), \qquad R_{ik}(c)^{-1} = R_{ik}(-c). \tag{3.13}$$

En resumen: las operaciones de fila se deshacen con otras operaciones de fila.

Las operaciones de fila empleadas para transformar la matriz aumentada del ejemplo anterior en forma "triangular" son todos de tipo (c). Ellas son $F_2 \mapsto F_2 - 3F_1$, luego $F_3 \mapsto F_3 - 3F_1$ y después $F_3 \mapsto F_3 - \frac{3}{5}F_2$. El efecto cumulativo:

$$[A \mid \boldsymbol{b}] \mapsto R_{32}(\tfrac{3}{5})\,R_{31}(3)\,R_{21}(3)\,[A \mid \boldsymbol{b}] =: [V \mid \boldsymbol{c}]$$

convierte

$$[A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 3 & 4 & -4 & 7 \\ 3 & 6 & 2 & -3 \end{array}\right] \quad \text{en} \quad [V \mid \boldsymbol{c}] = \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & 0 & \frac{28}{5} & -\frac{42}{5} \end{array}\right].$$

La citada *factorización* de matrices viene de examinar el proceso inverso:

$$
\begin{aligned}
[A \mid \boldsymbol{b}] &= \bigl(R_{32}(\tfrac{3}{5})\, R_{31}(3)\, R_{21}(3)\bigr)^{-1} [V \mid \boldsymbol{c}] \\
&= R_{21}(-3)\, R_{31}(-3)\, R_{32}(-\tfrac{3}{5})\, [V \mid \boldsymbol{c}] =: L\, [V \mid \boldsymbol{c}],
\end{aligned}
$$

donde esta L es la *matriz triangular inferior*:

$$
L = R_{21}(-3)\, R_{31}(-3)\, R_{32}(-\tfrac{3}{5}) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{5} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix}.
$$

Si se suprime la cuarta columna de esas matrices aumentadas, se obtiene una factorización $A = LV$ de la matriz de coeficientes original, en *dos matrices triangulares*:

$$
\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & -5 & -1 \\ 0 & 0 & \frac{28}{5} \end{bmatrix}.
$$

La matriz L, además de ser triangular inferior, es **unipotente**; quiere decir, solo hay entradas 1 en la diagonal. Para obtener una factorización más simétrica, nótese que ninguna entrada diagonal de V es 0. Al dividir las filas de V por esas entradas diagonales – operaciones de fila de tipo (b) – se obtiene la factorización $A = LDU$:

$$
\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & \frac{28}{5} \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & 1 \end{bmatrix},
$$

en donde el factor central D es *diagonal* y el tercer factor U es triangular superior y unipotente.

Eliminación gaussiana simple

El sistema de ecuaciones lineales más general tiene m ecuaciones para n variables:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\cdots\cdots\cdots\cdots &= \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
$$

donde en general $m \neq n$. Si $a_{11} \neq 0$, se puede eliminar la variable x_1 de las demás ecuaciones, con las operaciones $E_i \mapsto E_i - (a_{i1}/a_{11})\, E_1$ de tipo (c). Esta entrada no cero a_{11} se llama el primer **pivote** del sistema. Si la nueva segunda ecuación tiene coeficiente no cero para x_2, se procede a eliminar x_2; etcétera.

En el caso cuadrado $m = n$, si todo sale bien, se llegará luego a un *sistema triangular*:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{22}^{(2)}x_2 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ \cdots\cdots\cdots\cdots &= \vdots \\ a_{nn}^{(n)}x_n &= b_n^{(n)} \end{aligned}$$

tal que las *nuevas ecuaciones* $E_1, E_2^{(2)}, E_3^{(3)}, \dots, E_n^{(n)}$ tendrán coeficientes iniciales no ceros:

$$a_{11} \neq 0, \quad a_{22}^{(2)} \neq 0, \quad a_{33}^{(3)} \neq 0, \dots, \quad a_{nn}^{(n)} \neq 0.$$

Los **pivotes** $a_{kk}^{(k)}$, señalados en rojo, aparecen uno por uno en el proceso de eliminación. Si ninguno es 0 (inclusive el último) se puede despejar $x_n, \dots, x_1$ por sustitución regresiva. Este caso, en donde solamente hay que emplear operaciones de tipo (c) solamente, es el de **eliminación gaussiana simple**.

No es difícil sistematizar este proceso en un *algoritmo*. (En los apuntes, al final de la sección 3.3, se ofrece un bloque de "pseudocódigo" para ponerlo en práctica.) Un algoritmo es un proceso sistemático, que previene la posibilidad de *fallar*. Como se sabe, las cosas no siempre salen bien. En la próxima clase, se verá lo que se puede hacer (o intentar hacer) si aparece un pivote 0 durante el proceso.

Sobre la quinta lista de ejercicios: "Tarea 5"

Ejercicio 3.4 Esta matriz A es la "matriz compañera" del polinomio t^3+4t^2+6t+4. Al sustituir la matriz A en lugar de la incógnita t, se calcula una matriz de cierto interés.

Ejercicio 3.5 Para calcular una potencia A^4, se calcula $B = A^2$ y luego B^2.

Ejercicio 3.7 Este ejercicio requiere unas fórmulas trigonométricas bien conocidas.

Ejercicios 3.8 y 3.9 Para "hallar todas" las matrices que cumplen las condiciones dadas, se trata de dar fórmulas para esas matrices que pueden incluir entradas a, b, etc. que no se despejan.

Ejercicio 3.11 Estas son matrices de operaciones de fila para un sistema lineal 3×3.

Ejercicio 3.13 y 3.14 Las matrices dan lugar a ciertos espacios vectoriales. Para calcular la dimensión de un tal espacio vectorial de matrices, un método posible es el de exhibir una base explícitamente – y de verificar que en efecto es una base.

3.4. Clase 9

Esta es la clase virtual del 10 de mayo del 2021. La temática comprende la sección § 3.4: *Eliminación gaussiana con pivoteo parcial*.

Intercambio de filas y pivoteo parcial

La eliminación gaussiana simple, vista en la clase anterior, depende de la posibilidad de que todos los *pivotes* sean diferentes de 0:

$$a_{11} \neq 0, \quad a_{22}^{(2)} \neq 0, \quad a_{33}^{(3)} \neq 0, \dots, \quad a_{nn}^{(n)} \neq 0.$$

Se debe recordar que solo el primer pivote a_{11} es visible al inicio del trabajo: los otros aparecen, uno por uno, en el transcurso del cálculo. Así que es difícil pronosticar si un determinado sistema de ecuaciones lineales cumplirá este requisito.

Además, si sucede que $a_{11} = 0$, el primer paso del algoritmo queda frustrado. Consideremos un caso sencillo (Ejemplo 3.14):

$$\begin{aligned} x_2 + x_3 &= 1 \\ x_1 + x_3 &= 1 \\ x_1 + x_2 &= 1 \end{aligned} \qquad \text{con matriz aumentada} \quad [A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right].$$

El remedio es obvio: es cuestión de subir la segunda fila a la primera posición, con la operación $F_1 \leftrightarrow F_2$ de tipo (a) – intercambio de filas. La nueva primera fila tiene entrada $(1, 1)$ diferente de 0, y es posible seguir con operaciones de tipo (c):

$$\left[\begin{array}{ccc|c} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow{F_1 \leftrightarrow F_2} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array}\right] \xrightarrow[F_3 - F_1]{F_3 :} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{array}\right] \xrightarrow[F_3 - F_2]{F_3 :} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -1 \end{array}\right]$$

y se llega, por sustitución regresiva, a la solución:

$$x_3 = \tfrac{1}{2}, \quad x_2 = \tfrac{1}{2}, \quad x_1 = \tfrac{1}{2}.$$

En el desarrollo anterior, se indican los pivotes sucesivos con un color rojo. (La indicación explícita de las operaciones es opcional.)

► En el Ejemplo 3.15, aparece un pivote nulo más tarde, $a_{22}^{(2)} = 0$:

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 4 \\ 2x_1 + 6x_2 + 5x_3 &= 2 \\ 3x_1 + 9x_2 + 8x_3 &= 0 \end{aligned} \quad \text{con} \quad [A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 0 \end{array}\right]. \tag{3.14}$$

Después de una primera operación de tipo (c), aparece 0 en la segunda entrada diagonal:

$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 0 \end{array}\right] \xrightarrow[F_2 - 2F_1]{F_2:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 3 & 9 & 8 & 0 \end{array}\right] \xrightarrow[F_3 - 3F_1]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 2 & -12 \end{array}\right] \xrightarrow[F_3 - 2F_2]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Tras dos operaciones de fila de tipo (c), el 0 en la posición (2,2) tiene otro cero debajo de él, en la posición (3,2). Entonces se puede mover un paso a la derecha y pivotear en $a_{23}^{(2)} = 1$.

Al final queda una fila de ceros! Pero eso es inofensivo: es *una ecuación trivial*, $0 = 0$. Se ha obtenido:

$$x_1 + 3x_2 + 2x_3 = 4, \qquad x_3 = -6, \qquad 0 = 0,$$

o más simplemente:

$$x_3 = -6, \qquad x_1 + 3x_2 = 16. \tag{3.15}$$

En este caso, queda una variable "libre" (que podría ser x_2); en definitiva:

$$x_1 = 16 - 3x_2, \qquad x_2 \text{ es arbitrario}, \qquad x_3 = -6.$$

Conclusión: el sistema original (3.14) es *consistente* y *posee infinitas soluciones*, parametrizadas por $x_2 \in \mathbb{R}$.

► Una variante de ese sistema es el siguiente:

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 4 \\ 2x_1 + 6x_2 + 5x_3 &= 2 \\ 3x_1 + 9x_2 + 8x_3 &= 7 \end{aligned} \quad \text{con} \quad [A \mid \boldsymbol{b}] = \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 7 \end{array}\right].$$

Ahora las operaciones de fila conducen algo más problemático:

$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 2 & 6 & 5 & 2 \\ 3 & 9 & 8 & 7 \end{array}\right] \xrightarrow[F_2 - 2F_1]{F_2:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 3 & 9 & 8 & 7 \end{array}\right] \xrightarrow[F_3 - 3F_1]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 2 & -5 \end{array}\right] \xrightarrow[F_3 - 2F_2]{F_3:} \left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 7 \end{array}\right].$$

La última fila corresponde a una ecuación sin solución alguna:

$$\underline{0x_1 + 0x_2 + 0x_3 = 7}\,.$$

Se concluye que el nuevo sistema de ecuaciones lineales es *inconsistente*: se puede plantearlo, pero su conjunto de soluciones es vacío.

En los ejemplos, las matrices aumentadas tienen entradas en $\mathbb{Z}$, o mejor dicho en el cuerpo $\mathbb{Q}$ de los números racionales. Pero las operaciones aritméticas son aplicables a cualquier cuerpo $\mathbb{F}$. Se puede resumir la situación como sigue (Escolio 3.13).

Sea A una matriz cuadrada $n \times n$ sobre $\mathbb{F}$ y sea $b \in \mathbb{F}^n$. Para las soluciones del sistema de un sistema de $\underline{m}$ ecuaciones lineales en $\underline{n}$ variables con estos coeficientes, se admiten tres posibilidades:

(a) *hay una solución única $\boldsymbol{x} \in \mathbb{F}^n$; o bien*

(b) *no hay solución alguna; o bien*

(c) *hay infinitas soluciones en $\mathbb{F}^n$.*

El algoritmo de eliminación modificada, con la inclusión de operaciones de fila de los tipos (a) y (c), se llama *pivoteo parcial*. La operaciones de fila de tipo (b) – multiplicación de una fila por una constante no cero – quedan en reserva; no son necesarios para obtener el conjunto de soluciones.

3.5. Clase 10

Esta es la clase virtual del 13 de mayo del 2021. La temática comprende la sección § 3.5: *Inversión de matrices*.

Inversión mediante eliminación gaussiana

Algunas matrices cuadradas tienen inversos, otras no. Ya se ha visto que una matriz 2×2 puede ser invertida con una *fórmula*:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{3.5}$$

siempre y cuando $ad - bc \neq 0$. (Esta es la condición de invertibilidad de matrices 2×2.)

Para matrices 3×3 y de tamaños mayores, existen fórmulas (entre ellas la regla de Cramer, que se verá más adelante) pero no son tan sencillas. En vez de buscar una fórmula, se requiere un *procedimiento* que invierte la matriz, cuando sea factible. Ese procedimiento es, ni más ni menos, la eliminación gaussiana.

Calcular A^{-1} es equivalente a resolver la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ para obtener $\boldsymbol{x}$ en términos de $\boldsymbol{b}$:

$$A\boldsymbol{x} = \boldsymbol{b} \implies \boldsymbol{x} = A^{-1}\boldsymbol{b}. \tag{3.16}$$

De hecho, el algoritmo de eliminación encuentra los pivotes $a_{kk}^{(k)}$ manejando las entradas de A solamente: la columna $\boldsymbol{b}$ solo se usa para reconstruir la solución $\boldsymbol{x}$.

Ya se ha visto (Escolio 3.16) que $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única si y solo si se encuentran un total de *n pivotes no ceros* – sea por eliminación simple o por pivoteo parcial. El caso contrario (falta de solución única) se manifiesta con una fila de ceros en la última iteración de la matriz A.

► La inversión de A se logra con reemplazar $\boldsymbol{b}$ por las columnas $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ de la base estándar de $\mathbb{R}^n$. Nótese el cálculo

$$B\boldsymbol{e}_2 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} = \boldsymbol{b}_2 \,.$$

Más generalmente, $B\boldsymbol{e}_k = \boldsymbol{b}_k$ es la columna # k de la matriz B. Entonces vale la pena poner a la derecha de una matriz aumentada que acumula *todas* las columnas de la base estándar; y esa es la matriz identidad, 1_n. En fin, una eliminación:

$$[A \mid 1_n] \longmapsto [1_n \mid A^{-1}]$$

produce, en el cajón derecho de la matriz aumentada, una copia de la matriz A^{-1} cuando ésta existe.

En el ejemplo 3.17, se muestra el mecanismo. Es de notar que al llegar al último pivote $a_{33}^{(3)} = -4$, no se continua con sustitución regresiva. Más bien, *se sigue pivotando* con el fin de "limpiar" las columnas de A *en cima de la diagonal* con operaciones de tipo (c), hasta reducir A a una matriz diagonal (la D de $A = LDU$).

Finalmente, con operaciones de tipo (b) se divide cada fila por su entrada diagonal y queda 1_n en el cajón izquierdo de la matriz aumentada. El cajón derecho es entonces la matriz A^{-1} buscada.

¿Y si A no es invertible? Pues, el algoritmo se detendrá sin éxito en el momento en el que aparece una fila de ceros en el cajón izquierdo: no será posible obtener 1_n con más operaciones.

Sobre la sexta lista de ejercicios: "Tarea 6"

Ejercicio 3.17 Este problema de geometría plana (obtener la ecuación del círculo que pasa por tres puntos dados) se resuelve al mudarlo en un problema de resolver un sistema de ecuaciones lineales.

Ejercicio 3.21 Para factorizar la matriz A de un sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$ en la forma $A = LDU$, se aplica eliminación gaussiana, guardando la cuenta de las matrices $R_{ik}(-c)$ cuyo producto es la matriz triangular L, como en el Ejemplo 3.11.

Ejercicio 3.24 La matriz A de este problema se llama una *matriz de permutación*: la solución del problema explica el nombre.

Ejercicio 3.27 La matriz A de este problema se llama una *matriz nilpotente* porque una potencia de ella es nula: $A^n = 0$. 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.

Ejercicios 3.28 y 3.29 Dos matrices de bloques, con "bloques" rectangulares del mismo tamaño, se pueden multiplicar como si fueran matrices 2×2. El problema 3.28 solo pide verificar que las dos matrices dadas tiene producto identidad en ambos órdenes: que $M\,M^{-1} = M^{-1}M = 1_{m+n}$. Luego se aplica esta receta para simplificar el cálculo de la matriz inversa en el problema 3.29.

4 Comentarios sobre las clases de MA–360, parte 4

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

4.1. Clase 11

Esta es la segunda parte de la clase virtual del 17 de mayo del 2021. La temática comprende el inicio de la sección § 4.1: *Aplicaciones lineales*.

Aplicaciones lineales

Después de haber introducido matrices en el capítulo 3 de modo "pragmático" como arreglos rectangulares de escalares formas por coeficientes en sistemas de ecuaciones lineales, es oportuno ahora reconsiderarlas con una visión algo más teórica. (Chesterton ha dicho: "no hay nada más práctico que una buena teoría".)

Volvemos a considerar espacios vectoriales V, W, Z sobre un cuerpo $\mathbb{F}$ cualquiera.

Una función $T\colon V \to W$ que respeta las operaciones de suma y multiplicación escalar se llama una **aplicación lineal** (a veces, *transformación lineal*) entre V y W:

(a) $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$, para todo $\boldsymbol{x}, \boldsymbol{y} \in V$;

(b) $T(c\boldsymbol{x}) = c\,T(\boldsymbol{x})$, para todo $\boldsymbol{x} \in V$ y cada $c \in \mathbb{F}$.

Es importante notar que una aplicación lineal lleva el vector $\mathbf{0} \in V$ al vector $\mathbf{0} \in W$:

$$T(\mathbf{0}) = T(0\boldsymbol{x}) = 0\,T(\boldsymbol{x}) = \mathbf{0}.$$

El conjunto de todas las aplicaciones lineales $T\colon V \to W$ se denota por $\underline{\mathcal{L}(V, W)}$. (La letra $\mathcal{L}$ sugiere "lineal".) Este resulta ser un *espacio vectorial*, con las operaciones usuales de funciones (Lema 4.6):

$$\underline{T + S} : \boldsymbol{x} \mapsto T(\boldsymbol{x}) + S(\boldsymbol{x}), \qquad \underline{cT} : \boldsymbol{x} \mapsto c\,T(\boldsymbol{x}).$$

El cero de $\mathcal{L}(V, W)$ es la *aplicación lineal nula*, denotado 0, esto es $\underline{0(\boldsymbol{x}) := \mathbf{0}}$ para todo $\boldsymbol{x}$.

En el caso $V = W$, se dice que $T\colon V \to V$ es un **operador lineal** sobre V. Se escribe $\underline{\mathcal{L}(V)}$ en vez de $\mathcal{L}(V, V)$. Hay un *operador identidad* $1_V \in \mathcal{L}(V)$ dado por $\underline{1_V(\boldsymbol{x}) := \boldsymbol{x}}$ para todo $\boldsymbol{x} \in V$.

Los valores de T en una base determinan T

En el resto de este capítulo, consideremos solamente espacios vectoriales de dimensión finita. Usualmente tomamos $\underline{n = \dim V}$ y fijamos una *base* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V.

Un resultado importante (Prop. 4.7) es que los n vectores $T(\boldsymbol{x}_k) \in V$ determinan la aplicación lineal T completamente. En efecto, si $T(\boldsymbol{x}_k) = \boldsymbol{z}_k \in W$, entonces por linealidad,

$$\begin{aligned} T(a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_n\boldsymbol{x}_n) &:= a_1\,T(\boldsymbol{x}_1) + a_2\,T(\boldsymbol{x}_2) + \cdots + a_n\,T(\boldsymbol{x}_n) \\ &= a_1\boldsymbol{z}_1 + a_2\boldsymbol{z}_2 + \cdots + a_n\boldsymbol{z}_n. \end{aligned}$$

Como cualquier vector $\boldsymbol{x} \in V$ es de la forma $\boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n$ con *los coeficientes $a_1, \dots, a_n$ determinados unívocamente* (Ejercicio 2.16) ya se conoce todos los valores $T(\boldsymbol{x})$.

Hay más: si uno *prescribe* los vectores $\boldsymbol{z}_k := T(\boldsymbol{x}_k)$ en W de modo arbitrario – los $\boldsymbol{z}_k$ no tienen que ser linealmente independientes – esto determina la $T \in \mathcal{L}(V, W)$.

La aplicación lineal T_A de una matriz A

Si A es una matriz $m \times n$ con entradas en $\mathbb{F}$, la fórmula

$$\underline{T_A(\boldsymbol{x}) := A\boldsymbol{x}} \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{F}^n \tag{4.1}$$

determina una aplicación lineal $\underline{T_A \colon \mathbb{F}^n \to \mathbb{F}^m}$. En efecto:

$$\begin{aligned} T_A(\boldsymbol{x} + \boldsymbol{y}) &= A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = T_A(\boldsymbol{x}) + T_A(\boldsymbol{y}), \\ T_A(c\boldsymbol{x}) &= A(c\boldsymbol{x}) = c\,A\boldsymbol{x} = c\,T_A(\boldsymbol{x}), \end{aligned}$$

lo cual verifica que T_A es lineal.

La matriz A de una aplicación lineal T

Por otro lado, si $T \in \mathcal{L}(V, W)$ es una aplicación lineal cualquiera, se puede asociar una matriz A a T. Sin embargo, es necesario preparar el terreno con una identificación concreta de V con $\mathbb{F}^n$; y otra identificación de W con $\mathbb{F}^m$.

Esto se logra con *la elección de dos bases*: $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V y $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W.

Como T está determinada por los vectores $\boldsymbol{z}_j := T(\boldsymbol{x}_j) \in W$, es cuestión de expandir cada $\boldsymbol{z}_j$ como combinación lineal $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$. Para cada $j = 1, \dots, n$, se obtiene

$$T(\boldsymbol{x}_j) = \boldsymbol{z}_j = a_{1j}\,\boldsymbol{y}_1 + a_{2j}\,\boldsymbol{y}_2 + \cdots + a_{mj}\,\boldsymbol{y}_m.$$

Esto se puede resumir en la siguiente fórmula – la más importante del curso:

$$\boxed{T(\boldsymbol{x}_j) =: \sum_{i=1}^{m} a_{ij}\, \boldsymbol{y}_i\,.} \tag{4.2}$$

Fíjese que hay *una sutil diferencia* entre la expresión al lado derecho y el producto de una matriz por un vector columna. El producto $\boldsymbol{b} = A\boldsymbol{c}$ de una matriz A por una columna $\boldsymbol{c} \in \mathbb{F}^n$ es otra columna $\boldsymbol{b} \in \mathbb{F}^m$ con entradas

$$b_i = \sum_{j=1}^{n} a_{ij} c_j\,.$$

En el producto $A\boldsymbol{c}$, se suma sobre índices contiguos j, mientras en la expansión (4.2) se suma sobre índices i separadas por una j intermedio. En el segundo caso, a_{ij} y c_j son ambos escalares, pero en la fórmula (4.2) el escalar a_{ij} da un múltiplo del *vector* $\boldsymbol{y}_i$.

No obstante, las dos fórmulas sí están relacionadas. Podemos expandir un vector $\boldsymbol{x} \in V$ y su imagen $T(\boldsymbol{x}) \in W$ en las bases respectivas:

$$\begin{aligned} \boldsymbol{x} &= c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n \\ T(\boldsymbol{x}) &= b_1\boldsymbol{y}_1 + \cdots + b_m\boldsymbol{y}_m \end{aligned}$$

y con la linealidad de T, se deduce que:

$$\sum_{i=1}^{m} b_i\boldsymbol{y}_i = T(\boldsymbol{x}) = T\Bigg(\sum_{j=1}^{n} c_j\boldsymbol{x}_j\Bigg) = \sum_{j=1}^{n} c_j\, T(\boldsymbol{x}_j) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij}\, \boldsymbol{y}_i = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}c_j\, \boldsymbol{y}_i\,.$$

Ahora bien: la unicidad de los coeficientes en una combinación lineal con respecto a la *base* $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W (usando el Ejercicio 2.16 de nuevo) muestra que, para cada $i \in \{1, \dots, m\}$: implica

$$b_i = \sum_{j=1}^{n} a_{ij}\, c_j \qquad \text{y esto dice que} \qquad \boldsymbol{b} = A\,\boldsymbol{c}.$$

Se dice que A **es la matriz de** T con respecto a las dos bases dadas.

La *dimensión* del espacio vectorial $\mathcal{L}(V, W)$ es la dimensión del espacio de matrices $m \times n$ (Corolario 4.10):

$$\dim \mathcal{L}(V, W) = mn = (\dim W)(\dim V).$$

4.2. Clase 12

Esta es la clase virtual del 20 de mayo del 2021. La temática comprende el resto de la sección § 4.1: *Aplicaciones lineales*.

Composición de aplicaciones lineales, producto de matrices

En la clase anterior, vimos que a una aplicación lineal $T\colon V \to W$ se le puede asociar una matriz A, de la siguiente manera. Primero se escoge una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ para V y otra base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ para W. Los valores de T en la primera base se expanden en la segunda base, dando lugar a una matriz de coeficientes $A = [a_{ij}]$:

$$T(\boldsymbol{x}_j) =: \sum_{i=1}^{m} a_{ij}\, \boldsymbol{y}_i\,. \tag{4.2}$$

Así se establece una correspondencia uno a uno $T \mapsto A$ entre los espacios vectoriales $\mathcal{L}(V, W)$ y $M_{m\times n}(\mathbb{F})$.

Esta correspondencia $T \mapsto A$ es útil, porque:

- Es *lineal*: $T \mapsto A$ y $T' \mapsto A' \implies T + T' \mapsto A + A'$ y $c\,T \mapsto c\,A$.
- Es *biunívoca*, porque es uno a uno (los coeficientes a_{ij} determinan T) y es sobreyectiva: dada una matriz A, la fórmula (4.2) define una aplicación lineal T.
- Lleva una *composición* de aplicaciones lineales en el *producto* de sus matrices (a continuación).
- Lleva la *transpuesta* de T en la transpuesta de A (más abajo).

► Dadas tres espacios vectoriales V, W y Z, y dos aplicaciones lineales $T\colon V \to W$ y $S\colon W \to Z$, se define una **aplicación lineal compuesta** $\underline{ST\colon V \to Z}$ por una receta muy simple:

$$ST(\boldsymbol{x}) := S(T(\boldsymbol{x})) \quad \text{para} \quad \boldsymbol{x} \in V.$$

Este es el chequeo de que ST respeta sumas:

$$ST(\boldsymbol{x} + \boldsymbol{y}) = S(T(\boldsymbol{x} + \boldsymbol{y})) = S(T(\boldsymbol{x}) + T(\boldsymbol{y})) = S(T(\boldsymbol{x})) + S(T(\boldsymbol{y})) = ST(\boldsymbol{x}) + ST(\boldsymbol{y}).$$

Por ejemplo, si $A \in M_{r\times m}(\mathbb{F})$ y $B \in M_{m\times n}(\mathbb{F})$, las aplicaciones T_A y T_B se componen así:

$$T_A\, T_B(\boldsymbol{x}) = T_A(T_B(\boldsymbol{x})) = T_A(B\boldsymbol{x}) = A(B\boldsymbol{x}) = AB\boldsymbol{x} = T_{AB}(\boldsymbol{x}) \in \mathbb{F}^r$$

tomando en cuenta que AB es una matriz $r \times n$, así que $T_{AB}\colon \mathbb{F}^n \to \mathbb{F}^r$.

Se deduce que $T_A\, T_B = T_{AB}$.

Se puede ilustrar esta composiciones con un diagrama de flechas:

$$
\begin{tikzcd}
V \arrow[r, "T"] \arrow[rr, bend right, "ST"'] & W \arrow[r, "S"] & Z
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbb{F}^n \arrow[r, "T_B"] \arrow[rr, bend right, "T_A T_B = T_{AB}"'] & \mathbb{F}^m \arrow[r, "T_A"] & \mathbb{F}^r
\end{tikzcd}
$$

La Prop. 4.11 dice que si $S \mapsto A$ y $T \mapsto B$ son las matrices asociadas a S y T – hay que fijar una base $\{z_1, \dots, z_r\}$ para Z también – entonces resulta que $\underline{ST \mapsto AB}$.

Esto se comprueba aplicando la fórmula clave (4.2) un par de veces. Por conveniencia, llámese $C = AB$ a la matriz producto:

$$
\begin{aligned}
ST(\boldsymbol{x}_j) = S(T(\boldsymbol{x}_j)) = S\biggl(\sum_{k=1}^{m} b_{kj}\,\boldsymbol{y}_k\biggr) &= \sum_{k=1}^{m} b_{kj}\,S(\boldsymbol{y}_k) \\
= \sum_{k=1}^{m} b_{kj} \sum_{i=1}^{r} a_{ik}\,\boldsymbol{z}_i &= \sum_{i=1}^{r}\sum_{k=1}^{m} a_{ik}b_{kj}\,\boldsymbol{z}_i = \sum_{i=1}^{r} c_{ij}\,\boldsymbol{z}_i \,.
\end{aligned}
$$

Espacios duales y aplicaciones transpuestas

Para verificar que la transpuesta de una aplicación lineal $R \colon V \to W$ corresponde a la transpuesta de su matriz, primero hay que *definir* la transpuesta de R. Resulta que esta es una aplicación lineal $\underline{R^{\mathsf{t}} \colon W^* \to V^*}$ entre dos espacios vectoriales nuevos.

La Defn. 4.12 introduce el llamado **espacio dual** de V, que es

$$
\underline{V^* := \mathcal{L}(V, \mathbb{F})}\,.
$$

Un elemento $f \in V^*$ es simplemente una función lineal $\underline{f \colon V \to \mathbb{F}}$. Por costumbre, se dice que f es una **forma lineal** sobre V.

Una manera de obtener ejemplos de formas lineales es elegir una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, expandir un vector $\boldsymbol{x} \in V$ en esa base, y tomar uno de los coeficientes de la expansión:

$$
\begin{gathered}
\boldsymbol{x} = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n \,, \\
f_k(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n) := c_k \,.
\end{gathered}
$$

De esta manera, se obtiene una conjunto $\{f_1, \dots, f_n\}$ de formas lineales sobre V.

Cada f_k queda determinada (en vista de su linealidad) por sus valores en la base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V:

$$
f_k(\boldsymbol{x}_j) := \begin{cases} 1, & \text{si } k = j, \\ 0, & \text{si } k \neq j. \end{cases} \tag{4.3}
$$

Se puede abreviar eso en la fórmula: $f_k(\boldsymbol{x}_j) = [\![k = j]\!]$.

El Lema 4.12 dice que $\{f_1, \dots, f_n\}$ es una *base* de V^*. Se debe demostrar dos cosas:

- *Los f_k son linealmente independientes*: Si $a_1 f_1 + \cdots + a_n f_n = 0$ en V^*, entonces al evaluar en el vector $\boldsymbol{x}_j$:
$$0 = (a_1 f_1 + \cdots + a_n f_n)(\boldsymbol{x}_j) = a_1\, f_1(\boldsymbol{x}_j) + \cdots + a_n\, f_n(\boldsymbol{x}_j) = a_j\, f_j(\boldsymbol{x}_j) = a_j.$$

- *Los f_k generan V^**: para cada $g \in V^*$, se puede definir $b_k := g(\boldsymbol{x}_k)$ para cada k. Entonces
$$(b_1 f_1 + \cdots + b_n f_n)(\boldsymbol{x}_j) = b_1\, f_1(\boldsymbol{x}_j) + \cdots + b_n\, f_n(\boldsymbol{x}_j) = b_j\, f_j(\boldsymbol{x}_j) = b_j = g(\boldsymbol{x}_j).$$
Esto vale para cada $j = 1, \dots, n$. Se deduce que $g = b_1 f_1 + \cdots +_n f_n$.

A partir de la fórmula (4.3), se dice que $\{f_1, \dots, f_n\}$ es la **base dual** de V^* – esto es, dual a la *base original* $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V.

► Ahora sí podemos definir la transpuesta de $R\colon V \to W$. Esto es la aplicación $\underline{R^{\mathsf{t}}\colon W^* \to V^*}$ dada por
$$\underline{R^{\mathsf{t}}(g) := g \circ R}; \quad \text{esto es,} \quad \underline{R^{\mathsf{t}}(g)(\boldsymbol{x}) := g(R(\boldsymbol{x}))}. \tag{4.4}$$

La base $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W también determina una *base dual* $\{g_1, \dots, g_m\}$ de W^*. Si las matrices de R y R^{t} están dadas por $R \mapsto C$ y $R^{\mathsf{t}} \mapsto D$ respecto de estas bases, se escribe dos copias de (4.2):
$$R(\boldsymbol{x}_j) = \sum_{r=1}^{m} c_{rj}\, \boldsymbol{y}_r \qquad \text{y} \qquad R^{\mathsf{t}}(g_i) =: \sum_{k=1}^{n} d_{ki} f_k$$
El siguiente cálculo muestra que $D = C^{\mathsf{t}}$:
$$\begin{aligned} d_{ji} &= \sum_{j=1}^{n} d_{ki}\, [\![j = k]\!] = \sum_{k=1}^{n} d_{ki}\, f_k(\boldsymbol{x}_j) \\ &= R^{\mathsf{t}}(g_i)(\boldsymbol{x}_j) = g_i(R(\boldsymbol{x}_j)) = g_i\Biggl(\sum_{r=1}^{m} c_{rj}\, \boldsymbol{y}_r\Biggr) \\ &= \sum_{r=1}^{m} c_{rj}\, g_i(\boldsymbol{y}_r) = \sum_{r=1}^{m} c_{rj}\, [\![i = r]\!] = c_{ij}\,. \end{aligned}$$

Si $S\colon W \to Z$, entonces $S^{\mathsf{t}}\colon Z^* \to W^*$. Al tomar $h \in Z^*$, se obtiene de (4.4):
$$\underline{R^{\mathsf{t}} S^{\mathsf{t}}}(h) = R^{\mathsf{t}}(S^{\mathsf{t}}(h)) = R^{\mathsf{t}}(h \circ S) = h \circ S \circ R = h \circ SR = \underline{(SR)^{\mathsf{t}}}(h).$$

Sobre la sétima lista de ejercicios: "Tarea 7"

Ejercicio 4.2 La matriz asociada a una aplicación lineal depende de las bases usadas. Se ejemplifica esto para la derivada de polinomios cúbicos.

Ejercicio 4.6 Las matrices 2×2 (con entradas en un cuerpo $\mathbb{F}$) forman un espacio vectorial de dimensión 4. Sobre ese espacio hay varios operadores lineales de interés. Con respecto a la base estándar de $M_2(\mathbb{F})$, esos operadores tienen matrices 4×4.

Ejercicio 4.7 En el análisis numérico, se encuentra el problema de obtener una función $f : [a, b] \to \mathbb{R}$ conociendo solamente un juego finito de sus valores $b_k = f(c_k)$ para algunos $c_0, c_1, \dots, c_n \in [a, b]$. Es posible pasar el grafo de un *polinomio $p(t)$ de grado $\leqslant n$* por los $(n + 1)$ puntos $(c_k, b_k) \in \mathbb{R}^2$. Ese polinomio es una combinación lineal $p(t) = b_0\pi_0(t) + \cdots + b_n\pi_n(t)$ de los *polinomios interpolativos de Lagrange* de este ejercicio. Para que eso sucede, es necesario (y suficiente) que los $\pi_k(t)$ formen una *base* para $\mathbb{R}_n[t]$.

4.3. Clase 13

Esta es la clase virtual del 24 de mayo del 2021. La temática comprende la sección §4.2: *Núcleo e imagen de una aplicación lineal*.

Núcleo e imagen, nulidad y rango

Una manera de describir una aplicación lineal $S : V \to W$, como se ha visto, es la de producir una matriz A – una vez que se haya fijado un par de bases, una para V y la otra para W. Un enfoque menos concreto, pero que tiene la ventaja de no depender de la selección de bases, es asociarla a S *cuatro subespacios* de importancia.

Los primeros dos de estos subespacios son su **núcleo** $\underline{\ker S}$, un subespacio del dominio V; y si **imagen** $\underline{\operatorname{im} S}$, un subespacio del codominio W. Ellos son:

$$\underline{\ker S} := \{\, \boldsymbol{x} \in V : S(\boldsymbol{x}) = \boldsymbol{0} \,\} \leqslant V; \tag{4.5}$$

$$\underline{\operatorname{im} S} := S(V) = \{\, S(\boldsymbol{x}) : \boldsymbol{x} \in V \,\} \leqslant W. \tag{4.6}$$

El núcleo (llamado *kernel*, en inglés), también conocido como *espacio nulo* de S, es el juego de vectores que S lleva al origen $\boldsymbol{0} \in W$. Si $\boldsymbol{x}, \boldsymbol{y} \in \ker S$ y si $c \in \mathbb{F}$,

$$S(\boldsymbol{x} + \boldsymbol{y}) = S(\boldsymbol{x}) + S(\boldsymbol{y}) = \boldsymbol{0} + \boldsymbol{0} = \boldsymbol{0}; \qquad S(c\boldsymbol{x}) = c\, S(\boldsymbol{x}) = c\,\boldsymbol{0} = \boldsymbol{0};$$

así que $\ker S$ sí es un subespacio de V: se escribe $\underline{\ker S \leqslant V}$.

También es evidente que $\operatorname{im} S = S(V)$ es un subespacio de W, escrito $\underline{\operatorname{im} S \leqslant W}$.

Las *dimensiones* de estos subespacios son dos *números* asociados con S: Ellos son su **nulidad** $\underline{n(S)}$ y su **rango** $\underline{r(S)}$:

$$\underline{n(S)} := \dim(\ker S), \qquad \underline{r(S)} := \dim(\operatorname{im} S).$$

Es obvio que $n(S) \leqslant \dim V$; y que $r(S) \leqslant \dim W$.

La Prop. 4.18 expresa las posibilidad de S sea inyectiva o sobreyectiva en términos de estos subespacios y números. Es obvio que S es *sobreyectivo* si y solo si $\underline{\operatorname{im} S = W}$, si y solo si $\underline{r(S) = \dim W}$.

A su vez, S es *uno-a-uno* si y solo si $\underline{\ker S = \{\mathbf{0}\}}$, si y solo si $\underline{n(S) = 0}$. Esto merece una prueba detallada:

$$\begin{aligned} S \text{ es uno-a-uno} \quad &\text{si y solo si} \quad S(\boldsymbol{x}) = S(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y} \\ &\text{si y solo si} \quad S(\boldsymbol{x}) - S(\boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{x} - \boldsymbol{y} = \mathbf{0} \\ &\text{si y solo si} \quad S(\boldsymbol{x} - \boldsymbol{y}) = \mathbf{0} \implies \boldsymbol{x} - \boldsymbol{y} = \mathbf{0} \\ &\text{si y solo si} \quad S(\boldsymbol{z}) = \mathbf{0} \implies \boldsymbol{z} = \mathbf{0} \\ &\text{si y solo si} \quad \boldsymbol{z} \in \ker S \implies \boldsymbol{z} = \mathbf{0} \\ &\text{si y solo si} \quad \ker S = \{\mathbf{0}\}. \end{aligned}$$

► Los otros dos de los "cuatro subespacios" son el núcleo $\underline{\ker S^{\mathsf{t}}}$ y la imagen $\underline{\operatorname{im} S^{\mathsf{t}}}$ *de la aplicación lineal transpuesta* $S^{\mathsf{t}} \colon W^* \to V^*$.

Ellos son subespacios de los espacios duales: $\underline{\ker S^{\mathsf{t}} \leqslant W^*}$ y $\underline{\operatorname{im} S^{\mathsf{t}} \leqslant V^*}$.

Anuladores

Para poder calcular las dimensiones de estos últimos dos subespacios, hace falta un concepto extra, en la Defn. 4.19. Si $M \leqslant V$, se define su **anulador** $\underline{M^\perp \leqslant V^*}$ por:

$$\underline{M^\perp} := \{\, f \in V^* : f(\boldsymbol{x}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}. \tag{4.7a}$$

Fíjese que $V^\perp = \{0\}$ y que $\{\mathbf{0}\}^\perp = V^*$. (En general, por más grande que sea $M \leqslant V$, más pequeña es $M^\perp \leqslant V^*$.)

En la dirección contraria, si $N \leqslant V^*$, su anulador es un subespacio del V original, denotado por:

$$\underline{{}^\perp N} := \{\, \boldsymbol{x} \in V : f(\boldsymbol{x}) = 0 \text{ para todo } f \in N \,\}. \tag{4.7b}$$

Nótese que $N \leqslant V^*$ implica $\underline{{}^\perp N \leqslant V}$.

¡Basta de definiciones! Si $n = \dim V$, la Prop. 4.20 calcula las dimensiones:

$$\dim(M^\perp) = \underline{n - \dim M} \qquad \text{y} \qquad \dim({}^\perp N) = \underline{n - \dim N}.$$

Vamos a comprobar la primera; la segunda quedará como ejercicio.

Sea $n := \dim V$ y $k := \dim M$, con $M \leqslant V$ (lo que implica $k \leqslant n$). *Hace falta mostrar* que $\underline{\dim(M^\perp) = n - k}$.

Tómese una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ de M. Entonces se puede rellenar una base de V al agregar otros vectores $\boldsymbol{x}_{k+1}, \dots, \boldsymbol{x}_n$ – eso es la Prop. 2.29. Como ahora tenemos una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, esto da lugar a una *base dual* $\{f_1, \dots, f_n\}$ de V^*, tal que

$$f_i(\boldsymbol{x}_j) = 0 \quad \text{si} \quad i \neq j, \qquad f_j(\boldsymbol{x}_j) = 1.$$

Entonces

$$\begin{aligned} i > k \text{ pero } j \leqslant k &\implies f_i(\boldsymbol{x}_j) = 0, \\ i > k, \quad \boldsymbol{x} \in M &\implies f_i(\boldsymbol{x}) = 0, \\ i > k &\implies f_i \in M^\perp. \end{aligned}$$

Como resultado, se obtiene $\operatorname{lin}\langle f_{k+1}, \dots, f_n \rangle \leqslant M^\perp$.

Por otro lado, si $f \in M^\perp$, entonces $f \in V^*$, luego $\underline{f = b_1 f_1 + \cdots + b_n f_n}$ para ciertos b_i. Ahora $j \leqslant k$, se obtiene

$$j \leqslant k \implies b_j = \sum_{i=1}^n b_i \, [\![i = j]\!] = \sum_{i=1}^n b_i \, f_i(\boldsymbol{x}_j) = f(\boldsymbol{x}_j) = 0$$

porque $\boldsymbol{x}_j \in M$, $f \in M^\perp$. Entonces $\underline{f = b_{k+1} f_{k+1} + \cdots + b_n f_n}$.

Por lo tanto, $\{f_{k+1}, \dots, f_n\}$ es una *base* de $M^\perp$; y luego $\dim(M^\perp) = n - k$.

► En términos de subespacios, el "doble anulador" es el original (Cor. 4.21):

$$\underline{{}^\perp(M^\perp) = M} \leqslant V, \qquad \underline{({}^\perp N)^\perp = N} \leqslant V^*.$$

Veamos la primera: como $x \in M$, $f \in M^\perp$ implica $f(\boldsymbol{x}) = 0$, se ve que $\boldsymbol{x} \in {}^\perp(M^\perp)$. Esto dice que $M \leqslant {}^\perp(M^\perp)$. Ahora calculemos las dimensiones:

$$\dim({}^\perp(M^\perp)) = n - \dim(M^\perp) = n - (n - k) = k = \dim M$$

y con eso, hay igualdad: $M = {}^\perp(M^\perp)$. El otro caso es similar.

Ahora, los cuatro subespacios especiales:

$$\ker S \leqslant V, \quad \operatorname{im} S \leqslant W, \quad \ker S^{\mathsf{t}} \leqslant W^*, \quad \operatorname{im} S^{\mathsf{t}} \leqslant V^*$$

al parecer, dan lugar a otros cuatro:

$$(\ker S)^\perp \leqslant V^*, \quad (\operatorname{im} S)^\perp \leqslant W^*, \quad {}^\perp(\ker S^{\mathsf{t}}) \leqslant W, \quad {}^\perp(\operatorname{im} S^{\mathsf{t}}) \leqslant V.$$

Afortunadamente, los nuevos subespacios duplican los originales, como sigue.

La Prop. 4.22 dice que si $S\colon V \to W$ es lineal, entonces

$$\begin{aligned} \underline{(\operatorname{im} S)^\perp = \ker S^{\mathsf{t}}} \quad &\text{y} \quad \underline{(\ker S)^\perp = \operatorname{im} S^{\mathsf{t}}}, \\ \underline{{}^\perp(\operatorname{im} S^{\mathsf{t}}) = \ker S} \quad &\text{y} \quad \underline{{}^\perp(\ker S^{\mathsf{t}}) = \operatorname{im} S}. \end{aligned} \tag{4.8}$$

Por aquello de los dobles anuladores, solo hace falta comprobar dos de estos cuatro. Para la primera, tómese $g \in W^*$. Resulta que:

$$\begin{aligned} g \in (\operatorname{im} S)^\perp \quad &\text{si y solo si} \quad g(S(\boldsymbol{x})) = 0 \text{ para todo } \boldsymbol{x} \in V \\ &\text{si y solo si} \quad S^{\mathsf{t}}(g) = g \circ S = 0 \text{ en } V^* \\ &\text{si y solo si} \quad g \in \ker S^{\mathsf{t}}. \end{aligned}$$

Para la tercera, tómese $\boldsymbol{x} \in V$. Entonces:

$$\begin{aligned} \boldsymbol{x} \in {}^\perp(\operatorname{im} S^{\mathsf{t}}) \quad &\text{si y solo si} \quad S^{\mathsf{t}}(g)(\boldsymbol{x}) = 0 \text{ para todo } g \in W^* \\ &\text{si y solo si} \quad g(S(\boldsymbol{x})) = 0 \text{ para todo } g \in W^* \\ &\text{si y solo si} \quad S(\boldsymbol{x}) = \mathbf{0} \text{ en } W \\ &\text{si y solo si} \quad \boldsymbol{x} \in \ker S. \end{aligned}$$

En seguida, al calcular las dimensiones para los cuatro espacios en las fórmulas (4.8), se obtiene información sobre la nulidad y el rango de una aplicación lineal transpuesta. Este es el Corolario 4.23:

$$\underline{n(S^{\mathsf{t}})} = \dim W - r(S) \quad \text{y} \quad \underline{r(S^{\mathsf{t}})} = \dim V - n(S).$$

Por ejemplo, la primera igualdad sigue de:

$$n(S^{\mathsf{t}}) = \dim(\ker S^{\mathsf{t}}) = \dim((\operatorname{im} S)^\perp) = \dim W - \dim(\operatorname{im} S) = \dim W - r(S),$$

y la segunda es muy similar:

$$r(S^{\mathsf{t}}) = \dim(\operatorname{im} S^{\mathsf{t}}) = \dim((\ker S)^\perp) = \dim V - \dim(\ker S) = \dim V - n(S).$$

► Estas últimas fórmulas para $n(S^{\mathsf{t}})$ y $r(S^{\mathsf{t}})$ no son muy conocidas, porque en breve – en la próxima clase – estarán superadas por otras más potentes. El *teorema de rango y nulidad* va a establecer dos relaciones mucho más útiles:

$$\underline{r(S) = r(S^{\mathsf{t}})} \qquad \text{y} \qquad \underline{r(S) + n(S) = \dim V}.$$

4.4. Clase 14

Esta es la clase virtual del 27 de mayo del 2021. La temática comprende las secciones § 4.2: *Núcleo e imagen de una aplicación lineal* y § 4.3: *Forma escalonada de una matriz*

El teorema de rango y nulidad

Sea $S \colon V \to W$ una aplicación lineal; su transpuesta es $S^{\mathsf{t}} \colon W^* \to V^*$. En la clase anterior se observó que $r(S^{\mathsf{t}}) = \dim V - n(S)$. El teorema referido comprende dos fórmulas equivalentes:

$$\underline{r(S) = r(S^{\mathsf{t}})} \qquad \text{y} \qquad \underline{r(S) + n(S) = \dim V}.$$

Se puede demostrar cualquiera de las dos; optamos aquí por la segunda, que da el nombre al teorema.

La prueba usa el método de la *compleción de bases*. Si $\dim V = n$ y $n(S) = k$, se toma una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ de $\ker S$ y con otros vectores se llega a una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V. Si $\boldsymbol{x} \in V$ se expande como $\boldsymbol{x} = a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n$ respecto de esta base, se ve que

$$\begin{aligned} S(\boldsymbol{x}) &= S(a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n) = a_1\, S(\boldsymbol{x}_1) + \cdots + a_n\, S(\boldsymbol{x}_n) \\ &= a_{k+1}\, S(\boldsymbol{x}_{k+1}) + \cdots + a_n\, S(\boldsymbol{x}_n). \end{aligned}$$

Por lo tanto, $\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ *genera* $S(V) = \underline{\operatorname{im} S} \leqslant W$.

Afirmación: $\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ es linealmente independiente:

$$\begin{aligned} & c_{k+1}\, S(\boldsymbol{x}_{k+1}) + \cdots + c_n\, S(\boldsymbol{x}_n) = \mathbf{0} \\ \implies & S(c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n) = \mathbf{0} \\ \implies & (c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n) \in \ker S \\ \implies & c_{k+1}\, \boldsymbol{x}_{k+1} + \cdots + c_n\, \boldsymbol{x}_n = b_1\boldsymbol{x}_1 + \cdots + b_k\boldsymbol{x}_k \end{aligned}$$

para algunos $b_1, \dots, b_k \in \mathbb{F}$. Como $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es linealmente independiente, cada $c_j = 0$ [y cada $b_i = 0$ también].

Conclusión:$\{S(\boldsymbol{x}_{k+1}), \dots, S(\boldsymbol{x}_n)\}$ es una *base* de $\operatorname{im} S$; y $r(S) = \dim(\operatorname{im} S) = \underline{n - k}$.

Rango y nulidad de una matriz $m \times n$

Si $A \in M_{m\times n}(\mathbb{F})$, se sabe que $\underline{\boldsymbol{x} \mapsto A\boldsymbol{x}}$ es una aplicación lineal $\underline{T_A \colon \mathbb{F}^n \to \mathbb{F}^m}$. La *imagen* de T_A es

$$\operatorname{im} T_A = \{\, A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = \operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \rangle. \tag{4.9b}$$

La segunda igualdad es válida porque cada $A\boldsymbol{x} \in \mathbb{F}^m$ es una combinación lineal de las *columnas* de A:

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n\,. \tag{4.9a}$$

El rango y la nulidad de A son los de la aplicación lineal T_A (Defn 4.25):

$$\underline{r(A)} := r(T_A) = \dim\{\, A\boldsymbol{x} \in \mathbb{F}^m : \boldsymbol{x} \in \mathbb{F}^n \,\},$$
$$\underline{n(A)} := n(T_A) = \dim\{\, \boldsymbol{x} \in \mathbb{F}^n : A\boldsymbol{x} = \mathbf{0} \,\}.$$

Esto, junto con la relación $\underline{r(A^t) = r(A)}$ del teorema anterior, ofrece dos maneras de visualizar el rango de A:

- $r(A)$ es *el número máximo de columnas linealmente independientes* en A.
- $r(A)$ es *el número máximo de filas linealmente independientes* en A.

(Las filas de A forman las columnas de A^t.) Como consecuencia, se ve que

$$r(A) \leqslant \min\{m, n\}.$$

Sistemas de ecuaciones lineales homogéneas

Un sistema de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{b}$ es **homogéneo** si $\boldsymbol{b} = \mathbf{0}$.

Un sistema homogénea $A\boldsymbol{x} = \mathbf{0}$ siempre *tiene al menos una solución*: el vector nulo $\boldsymbol{x} = \mathbf{0}$ (en $\mathbb{F}^n$) es la llamada “solución trivial”.

Ahora $\underline{A\boldsymbol{x} = \mathbf{0}} \iff \underline{\boldsymbol{x} \in \ker T_A}$. Entonces $\ker T_A$ es el *espacio de soluciones* de esta ecuación. Su dimensión es $n(A)$.

Prop. 4.27: el sistema $A\boldsymbol{x} = \mathbf{0}$ tiene solución única ($\boldsymbol{x} = \mathbf{0}$) si y solo si $\underline{n(A) = 0}$, si y solo si $\underline{r(A) = n}$.

Forma escalonada de una matriz A

Una matriz rectangular (o cuadrada) A tiene **forma escalonada** (*reducida por filas*) si cumple tres requisitos (Defn. 4.28):

(a) Ciertas columnas son **columnas básicas**: $\boldsymbol{a}_{j_1} = \boldsymbol{e}_1$, $\boldsymbol{a}_{j_2} = \boldsymbol{e}_2, \dots, \boldsymbol{a}_{j_k} = \boldsymbol{e}_k$ en $\mathbb{F}^m$.

(b) Estas columnas básicas aparecen *en el orden usual*: $j_1 < j_2 < \cdots < j_k$.

(c) Si $j < j_1$, es $\boldsymbol{a}_j = \mathbf{0}$; si $j_r \leqslant j < j_{r+1}$, los últimos $(m - r)$ elementos de $\boldsymbol{a}_j$ son ceros; si $j \geqslant j_k$, los últimos $(m - k)$ elementos de $\boldsymbol{a}_j$ son ceros.

Por ejemplo, una matriz en forma escalonada es la siguiente:

$$A = \begin{bmatrix} 0 & 1 & 3 & 0 & 8 & 4 & 0 & 7 & 4 \\ 0 & 0 & 0 & 1 & 6 & 7 & 0 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Aquí $m = 5$, $n = 9$, $k = 3$.

Las columnas básicos son $\boldsymbol{a}_2 = \boldsymbol{e}_1$, $\boldsymbol{a}_4 = \boldsymbol{e}_2$ y $\boldsymbol{a}_7 = \boldsymbol{e}_3$.

Las últimas dos filas son filas de ceros.

Es evidente que, en general, *las últimas* $(m - k)$ *filas son filas de ceros*, si $k < m$.

Prop. 4.30: en una matriz A en forma escalonada con k columnas básicas, resulta:

$$\operatorname{im} T_A = \operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \rangle = \operatorname{lin}\langle \boldsymbol{a}_{j_1}, \dots, \boldsymbol{a}_{j_k} \rangle = \operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_k \rangle$$

y en consecuencia, $\underline{r(A) = k}$.

► Una manera de *calcular* el rango de una matriz rectangular cualquiera es *reducirla a una forma escalonada mediante operaciones de fila*. (Esto es esencialmente el algoritmo de Gauss y Jordan.)

Es necesario estar seguro de que las operaciones de fila (que son reversibles) no cambian el rango (Prop. 4.31).

De hecho, si A y B son dos matrices $m \times n$ tal que B se transforma en A mediante operaciones de fila elementales:

$$A = L_p \cdots L_2 L_1 B,$$

entonces $\underline{A = LB}$ donde $\underline{L := L_p \cdots L_2 L_1}$ en $M_m(\mathbb{F})$, y a su vez $B = L^{-1}A$ usando las operaciones inversas, $L^{-1} = L_1^{-1} L_2^{-1} \cdots L_p^{-1}$. Entonces

$$\operatorname{im} T_A = \{\, A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = \{\, LB\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\} = T_L\big(\{\, B\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^n \,\}\big) = T_L(\operatorname{im} T_B).$$

y también $\operatorname{im} T_B = T_{L^{-1}}(\operatorname{im} T_A)$.

Aquí $T_L : \boldsymbol{z} \mapsto L\boldsymbol{z}$ y $T_{L^{-1}} : \boldsymbol{y} \mapsto L^{-1}\boldsymbol{y}$ son operadores lineales sobre $\mathbb{F}^m$, mutuamente inversos. Esto dice que T_L es una *biyección lineal* o (sinónimo) un *isomorfismo lineal*; así que $\underline{\operatorname{im} T_A}$ y $\underline{\operatorname{im} T_B}$ son espacios vectoriales isomorfos, con la misma dimensión: esto es, $\underline{r(A) = r(B)}$.

Sobre la octava lista de ejercicios: “Tarea 8”

Ejercicio 4.8 Para completar la prueba de la Proposición 4.20, se debe asociar a cada base del espacio vectorial dual V^* una “base dual” del espacio vectorial V original.

Ejercicio 4.11 La parte (a) dice que, si $S: V \to W$ es lineal, la *preimagen* de un conjunto linealmente independiente en W es un conjunto linealmente independiente en V. En la dirección contraria (imagen directa de un conjunto linealmente independiente en V), se requiere que el núcleo de S sea cero: eso es la parte (b).

Ejercicio 4.13 Una consecuencia importante del teorema de rango y nulidad es que una aplicación lineal entre dos espacios vectoriales *de la misma dimensión* es uno-a-uno si y solo si es sobreyectiva. En particular, un operador lineal sobre V es inyectivo si y solo si es sobreyectivo.

Ejercicio 4.14 En cambio, un operador lineal sobre una espacio vectorial infinitodimensional puede ser inyectivo sin ser sobreyectivo (y viceversa).

Ejercicio 4.15 La composición de dos aplicaciones lineales tiene menor rango que cualquiera de ellas. Una de estas desigualdades es fácil; la otra requiere el uso de las transpuestas.

4.5. Clase 15

Esta es la clase virtual del 31 de mayo del 2021. La temática comprende las secciones § 4.3: *Forma escalonada de una matriz* y § 4.4: *Cambios de base*.

Solubilidad de sistema de ecuaciones lineales

En las clase anterior, se definió la nulidad y rango de A, una matriz rectangular $m \times n$, a través de la aplicación lineal $T_A : \boldsymbol{x} \mapsto A\boldsymbol{x}$, como:

$$\begin{aligned} \underline{n(A)} &:= n(T_A) = \dim\{\, \boldsymbol{x} \in \mathbb{F}^n : A\boldsymbol{x} = \mathbf{0} \,\}, \\ \underline{r(A)} &:= r(T_A) = \dim\{\, A\boldsymbol{x} \in \mathbb{F}^m : \boldsymbol{x} \in \mathbb{F}^n \,\}. \end{aligned}$$

Los $\boldsymbol{x} \in \mathbb{F}^n$ tales que $A\boldsymbol{x} = \mathbf{0}$ constituyen el **espacio de soluciones** de esta ecuación lineal *homogénea*. Por otro lado, la totalidad de los $A\boldsymbol{x} \in \mathbb{F}^m$ forman el **espacio de columnas** de A: cada vector $A\boldsymbol{x}$ es una combinación lineal de las columnas de A.

También se vio que cualquier matriz $m \times n$, A, puede reducirse a una matriz con un formato especial, A', llamada su *forma escalonada*, mediante operaciones de fila. Estas operaciones no cambian el espacio de columnas, y por ende no cambian el rango: $\underline{r(A) = r(A')}$.

Como $r(A')$ es simplemente el número de columnas básicas en A', esto permite *calcular* $r(A)$, como efecto secundario de la reducción a forma escalonada.

► Ahora consideremos un sistema de ecuaciones lineales *inhomogéneo*: $\underline{A\boldsymbol{x} = \boldsymbol{b}}$.

Este sistema posee una *matriz aumentada*: $\underline{[A \mid \boldsymbol{b}]}$, la cual es una matriz $m \times (n+1)$.

Con operaciones de fila, ésta también se reduce a su forma escalonada: $\underline{[A' \mid \boldsymbol{b}']}$. (El cajón izquierdo de ésta, A', es la forma escalonada de A.)

Los rangos de estas matrices dan información sobre las posibilidades de resolver el sistema $A\boldsymbol{x} = b$. Sean:

$$\underline{k = r(A)} = r(A'), \qquad \underline{l = r\big([A \mid \boldsymbol{b}]\big)} = r\big([A' \mid \boldsymbol{b}']\big).$$

Hay tres posibilidades por considerar:

(1) $\underline{l = k + 1}$: esto sucede si la última columna de $[A' \mid \boldsymbol{b}']$ es básica, $\boldsymbol{b}' = \boldsymbol{e}_{k+1}$. Por ejemplo:

$$[A' \mid \boldsymbol{b}'] = \left[\begin{array}{ccccccccc|c} 0 & 1 & 3 & 0 & 8 & 4 & 0 & 7 & 4 & 0 \\ 0 & 0 & 0 & 1 & 6 & 7 & 0 & 2 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

En tal caso, la fila #$(k+1)$ es $[\boldsymbol{0} \mid 1]$, que representa *la ecuación inconsistente*: $\boldsymbol{0} \cdot \boldsymbol{x} = 1$; por ende, en este caso el sistema $A\boldsymbol{x} = \boldsymbol{b}$ **no tiene solución alguna**.

(2) $\underline{l = k = n}$: aquí la última columna no es básica, pero como $k = r(A) = n$ y a la vez $r(A) \leqslant m$, podría haber más ecuaciones que variables, $m \geqslant n$. Si $\underline{m > n}$, entonces las últimas $(m-n)$ filas de A' y $[A' \mid \boldsymbol{b}']$ son nulas; y el cajón A' tendrá n columnas básicas. De este modo se puede ver que

$$[A' \mid \boldsymbol{b}'] = \left[\begin{array}{c|c} 1_n & \boldsymbol{c} \\ 0 & \boldsymbol{0} \end{array}\right].$$

Las últimas filas corresponden a ecuaciones triviales $\boldsymbol{0} \cdot \boldsymbol{x} = 0$ y las primeras filas dan las *ecuaciones desacopladas* $\underline{x_i = c_i}$, para $i = 1, 2, \dots, n$.

En resumen: este es el caso donde el procedimiento de Gauss y Jordan llega a una **solución única** al sistema $A\boldsymbol{x} = \boldsymbol{b}$, descartando unas $(m-n)$ ecuaciones redundantes. (Desde luego, si $m = n$, hay solución única sin redundancias.)

(3) $l = k < n$: En este tercer caso, en el cajón A' de $[A' \mid \boldsymbol{b}']$, hay una k columnas básicas y otras $(n-k)$ columnas no básicas. Entonces se puede hacer una maniobra extra: la de *permutar las columnas* para llevar todas las columnas básicas a la izquierda (manteniendo su orden relativo). Por ejemplo:

$$\left[\begin{array}{ccccccc|c} 0 & 1 & 2 & 0 & 1 & 0 & 7 & 4 \\ 0 & 0 & 0 & 1 & 5 & 0 & 2 & 8 \\ 0 & 0 & 0 & 0 & 5 & 1 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccccccc|c} 1 & 0 & 0 & 0 & 2 & 1 & 7 & 4 \\ 0 & 1 & 0 & 0 & 0 & 5 & 2 & 8 \\ 0 & 0 & 1 & 0 & 0 & 5 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Esta "operación de columna" simplemente *reordena las variables*:

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) \mapsto (x_2, x_4, x_6, x_1, x_3, x_5, x_7),$$

que no afecta la resolubilidad del sistema de ecuaciones. En general:

$$[A' \mid \boldsymbol{b}'] \longmapsto \left[\begin{array}{cc|c} 1_k & B & \boldsymbol{c} \\ 0 & 0 & \boldsymbol{0} \end{array}\right],$$

donde el bloque B es de tamaño $k \times (n-k)$. Se descartan las $(m-k)$ ecuaciones triviales $\boldsymbol{0} \cdot \boldsymbol{x} = 0$ de las últimas filas. Al poner las "variables no básicas" igual a 0 (en el ejemplo $x_1 = x_3 = x_5 = x_7 = 0$, se obtiene una **solución particular** $\boldsymbol{x}_0 := [c_1 \ \cdots \ c_k \ 0 \ \cdots \ 0]^{\mathsf{t}}$. En el ejemplo, esto es

$$x_1 = 0, \quad x_2 = 4, \quad x_3 = 0, \quad x_4 = 8, \quad x_5 = 0, \quad x_6 = 6, \quad x_7 = 0.$$

Para la solución general, se toman "libres" las variables no básicas y se despejan las "variables básicas" en términos de ellas:

$$\begin{aligned} x_2 &= 4 - 2x_3 - x_5 - 7x_7\,, \\ x_4 &= 8 - 5x_5 - 2x_7\,, \\ x_6 &= 6 - 5x_5 - 3x_7\,. \end{aligned}$$

Si $\boldsymbol{z} := \boldsymbol{x} - \boldsymbol{x}_0$, entonces $A\boldsymbol{z} = A\boldsymbol{x} - A\boldsymbol{x}_0 = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$, de modo que $\boldsymbol{z} \in \ker T_A$. La nulidad de A es $n(A) = n - r(A) = n - k$: esta es la cantidad de columnas (o variables) no básicas en la forma escalonada A'.

Cambios de base

Sea $T\colon V \to W$ una aplicación lineal con matriz A (de tamaño $m \times n$) con respecto a un par de bases $\mathcal{B} := \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V y $\mathcal{C} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ de W.

¿Qué pasa con la matriz de T si se usan otras bases de V yW?

Sean $\underline{\mathcal{B}'} := \{x'_1, \dots, x'_n\}$ y $\underline{\mathcal{C}'} := \{y'_1, \dots, y'_m\}$ otras bases de V y W, respectivamente.

Podemos expandir cada vector $x'_s \in V$ en términos de la base $\mathcal{B}$; y cada vector $y_i \in W$ en términos de la base $\mathcal{C}'$:

$$x'_s = 1_V(x'_s) =: \sum_{j=1}^{n} p_{js}\, x_j \quad \text{mientras} \quad y_i = 1_W(y_i) =: \sum_{r=1}^{m} q_{ri}\, y'_r\,. \tag{4.10}$$

Las coeficientes en estas fórmulas forman dos matrices cuadradas: P es $n\times n$, mientras Q es $m \times m$. Estas son las **matrices de cambio de base**.

Aquí P es la matriz de la identidad $1_V : V \to V$ referida a las bases $\mathcal{B}$ y $\mathcal{B}'$ de V. También Q es la matriz de la identidad $1_W : W \to W$ referida a las bases $\mathcal{C}'$ y $\mathcal{C}$ de W.

► El Lema 4.33 dice que si B es la matriz de T con respecto a las nuevas bases $\mathcal{B}'$ y $\mathcal{C}'$, entonces

$$\underline{B = QAP}\,.$$

Por definición, la matriz $B = [b_{rs}]$ viene de la fórmula:

$$T(x'_s) = \sum_{r=1}^{m} b_{rs}\, y'_r.$$

De las fórmulas (4.10) y la linealidad de T, se obtiene el cálculo siguiente:

$$\begin{aligned} T(x'_s) &= T\Biggl(\sum_{j=1}^{n} p_{js}\, x_j\Biggr) = \sum_{j=1}^{n} p_{js}\, T(x_j) = \sum_{j=1}^{n} p_{js} \sum_{i=1}^{m} a_{ij}\, y_i \\ &= \sum_{j=1}^{n} p_{js} \sum_{i=1}^{m} a_{ij} \sum_{r=1}^{m} q_{ri}\, y'_r = \sum_{r=1}^{m} \Biggl(\sum_{i=1}^{m} \sum_{j=1}^{n} q_{ri}\, a_{ij}\, p_{js}\Biggr) y'_r \end{aligned}$$

y la expresión entre paréntesis a la derecha es la entrada (r, s) de la matriz QAP.

Matrices equivalentes

Defn. 4.34: dos matrices A y B del mismo tamaño $m \times n$ se llaman **equivalentes** si hay un par de matrices *invertibles* $P \in M_n(\mathbb{F})$ y $Q \in M_m(\mathbb{F})$ tales que $\underline{B = QAP}$.

Esta relación es:

- ⋄ *reflexiva*: $A = 1_m\, A\, 1_n$;
- ⋄ *simétrica*: $B = QAP \implies A = Q^{-1}BP^{-1}$;
- ⋄ *transitiva*: $B = QAP,\ C = Q'BP' \implies C = (Q'Q)A(PP')$.

Fíjese que $(Q'Q)^{-1} = Q^{-1}Q'^{-1}$ y también $(PP')^{-1} = P'^{-1}P^{-1}$.

4.6. Clase 16

Esta es la clase virtual del 3 de junio del 2021. La temática comprende las secciones § 4.4: *Cambios de base* y § 4.5: *Matrices semejantes*.

Equivalencia de matrices

En la clase anterior, se introdujo una relación de equivalencia entre las matrices de un tamaño fijo $m \times n$. Dos matrices $A, B \in M_{m\times n}(\mathbb{F})$ son **equivalentes** si cumplen

$$B = QAP,$$

donde Q y P son ciertas matrices cuadradas *invertibles*. Es necesario y obvio que la matriz Q es $m \times m$ mientras P es $n \times n$.

Se ha observado que esta es una *relación de equivalencia*, como el nombre indica.

En cada clase de equivalencia, hay una matriz especial. La Prop. 4.35 dice que si $r(A) = k$, se puede encontrar Q y P (invertibles) tales que

$$QAP = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.11}$$

Cabe recordar que $r(A) \leqslant \min\{m, n\}$; entonces es factible colocar 1_k en una matriz $m \times n$ como un *bloque* superior izquierdo. Los demás entradas '0' al lado derecho de (4.11) son bloques rectangulares de ceros, de tamaños $k \times (n - k)$, $(m - k) \times k$, y $(m - k) \times (n - k)$, respectivamente.

Se construyen estas matrices Q y P en *cuatro pasos*. El primer paso es cambiar A a su forma escalonada A' por operaciones de fila (reversibles). Esto conlleva una *premultiplicación* de A por una matriz Q que es $m \times m$ e invertible: $\underline{A' = QA}$.

El segundo paso es mover todas las columnas básicas de A' a la izquierda para formar el bloque 1_k. Recordemos que la operación de cambiar dos *filas* $F_i \leftrightarrow F_k$ de A se hace por $\underline{A \mapsto P_{ik}A}$, con una cierta matriz P_{ik}. Si se multiplica por este tipo de matriz *a la derecha*, $\underline{A \mapsto AP_{ik}}$, el efecto es el de cambiar dos *columnas*, $C_i \leftrightarrow C_k$, de la matriz A. (La P_{ik} a $\overline{\text{la derecha}}$ debe ser de tamaño $n \times n$.)

Para obtener el bloque deseado 1_k a la izquierda, habrá que hacer varias intercambios de columnas: en total, hay una matriz invertible $n \times n$ que llamaremos P_1 que postmultiplica A':

$$A \longmapsto QAP_1 = A'P_1 = \begin{bmatrix} 1_k & F \\ 0 & 0 \end{bmatrix},$$

donde F es un bloque $k \times (n - k)$.

En el tercer paso, tomamos la transpuesta de la nueva matriz:

$$\begin{bmatrix} 1_k & F \\ 0 & 0 \end{bmatrix} \longmapsto \begin{bmatrix} 1_k & 0 \\ F^t & 0 \end{bmatrix} = (QAP_1)^t$$

y en seguida, con eliminación gaussiana *simple*, se "limpia" debajo del bloque 1_k con nuevas operaciones de fila:

$$(QAP_1)^t = \begin{bmatrix} 1_k & 0 \\ F^t & 0 \end{bmatrix} \longmapsto \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix} =: P_2(QAP_1)^t$$

Aquí P_2 es $\underline{n \times n}$ e invertible – estas últimas matrices son de tamaño $\underline{n \times m}$.

El cuarto y último paso es volver a matrices $\underline{m \times n}$ al tomar transpuestas de nuevo. Si se pone $P := P_1 P_2^t$ (otra matriz invertible $n \times n$), se termina con la fórmula (4.11):

$$QAP = QAP_1P_2^t = \big(P_2(QAP_1)^t\big)^t = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}.$$

► Al parecer, solo hay un número finito de clases de equivalencia, distinguidas por el rango $r(A) = k$. Y así es, en efecto: el Cor. 4.36 hace notar que dos matrices A y B son equivalentes si y solo si tienen el mismo rango.

Para chequear esto, se debe comprobar que $\underline{r(QAP) = r(A)}$ toda vez que Q y P son invertibles. Ahora bien, el Ejercicio 4.15 muestra que

$$r(QAP) \leqslant r(AP) \quad \text{y} \quad r(AP) \leqslant r(A), \quad \text{así que} \quad r(QAP) \leqslant r(A).$$

Para la desigualdad inversa, se escribe $B := QAP$ y se nota que

$$r(A) = r(Q^{-1}BP^{-1}) \leqslant r(B) = r(QAP).$$

Matrices semejantes

Consideremos el caso de cambio de base para un operador lineal $T : V \to V$. Como ahora $V = W$, solo se requiere dos bases en vez de cuatro: se toma $\mathcal{C} = \mathcal{B}$ y $\mathcal{C}' = \mathcal{B}'$. En otras palabras, $\boldsymbol{y}_j = \boldsymbol{x}_j$ y $\boldsymbol{y}'_r = \boldsymbol{x}'_r$, con $m = n$. Las fórmulas (4.10) se simplifican en:

$$\boldsymbol{x}'_s =: \sum_{j=1}^{n} p_{js}\, \boldsymbol{x}_j \quad \text{mientras} \quad \boldsymbol{x}_i =: \sum_{r=1}^{n} q_{ri}\, \boldsymbol{x}'_r\,. \tag{4.10$'$}$$

Es casi inmediato que esto implica que $\underline{Q = P^{-1}}$.

► Esto permite definir una nueva relación de equivalencia entre *matrices cuadradas* de un tamaño fijo $n \times n$, llamado *semejanza*.

Defn. 4.37: dos matrices $A, B \in M_n(\mathbb{F})$ son **semejantes** si hay una matriz invertible $P \in M_n(\mathbb{F})$ tal que

$$\underline{B = P^{-1}AP}.$$

Esta relación es:

◇ *reflexiva*: $A = (1_n)^{-1} A \, 1_n$;

◇ *simétrica*: $B = P^{-1}AP \implies A = PBP^{-1}$;

◇ *transitiva*: $B = P^{-1}AP,\ C = R^{-1}BR \implies C = (PR)^{-1}A(PR)$.

Nótese que $(PR)^{-1} = R^{-1}P^{-1}$ en este caso.

Sobre la novena lista de ejercicios: "Tarea 9"

Ejercicios 4.17 a 4.19 En cada caso, se obtiene el rango al convertir la matriz a su forma escalonada.

Ejercicio 4.25 Se puede aplicar "operaciones de columna" directamente; o bien, se puede tomar la transpuesta, aplicar operaciones de fila, y volver a transponer.

Ejercicio 4.27 Una propiedad importante y útil de matrices (rectángulos o cuadrados) es poco conocido: una matriz de rango k siempre puede expresarse como una suma de k matrices de rango 1.

Sobre la décima lista de ejercicios: "Tarea 10"

Ejercicio 4.29 Las fórmulas (4.2) para la matriz A de T y (4.10) para las matrices de cambio de base dependen del orden de la lista de vectores en la base: esto es, se refieren a *bases ordenadas*.

Ejercicio 4.30 Para ver si dos matrices A y B son semejantes o no, se plantea la igualdad $B = P^{-1}AP$, o más simplemente: $PB = AP$, para ver si haya alguna matriz invertible P que la satisfaga.

Ejercicio 4.32 Un matriz cuadrada A se llama *diagonalizable* si es semejante a una matriz diagonal. Esto requiere encontrar un cambio de base apropiada.

5 Comentarios sobre las clases de MA–360, parte 5

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

5.1. Clase 17

Esta es la clase virtual del 7 de junio del 2021. La temática comprende la sección § 5.1: *Productos escalares reales*.

Productos escalares reales

En el capítulo 1 se estudió el producto escalar (o *producto punto*) de dos vectores en $\mathbb{R}^n$, pero hasta ahora no se ha vuelta a mencionar la posibilidad de formar productos de vectores en un contexto más general. En parte, esto se debe al uso de cuerpos de escalares $\mathbb{F}$ cualesquiera – hasta ahora. En este capítulo los escalares serán *los números reales*, $\mathbb{F} = \mathbb{R}$, o bien *los números complejos*, $\mathbb{F} = \mathbb{C}$.

Por ahora, usaremos solamente *escalares reales*: en esta sección, V y W serán espacios vectoriales sobre $\mathbb{R}$.

Se debe advertir un cambio de notación con respecto al capítulo 1: el producto escalar de dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ se denotará ahora con corchetes angulares: $\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}}$.

La Defn 5.1 exhibe la lista de propiedades de $\langle -, - \rangle$. Para $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ y $c \in \mathbb{R}$:

(a) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$,

(b) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$,

(c) $\langle \boldsymbol{x}, c\boldsymbol{y} \rangle = \langle c\boldsymbol{x}, \boldsymbol{y} \rangle = c\langle \boldsymbol{x}, \boldsymbol{y} \rangle$,

(d) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{x} = \mathbf{0}$ en V.

Al combinar las propiedades (a) y (b), también se ve que:

(b′) $\langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{z}, \boldsymbol{y} \rangle$.

De (b), (b′) y (c) se ve que la expresión $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es *lineal en* $\boldsymbol{x}$ y también *lineal en* $\boldsymbol{y}$: se dice que $\langle -, - \rangle$ es una función **bilineal** de $V \times V$ en $\mathbb{R}$.

La propiedad (d) hace uso del orden de $\mathbb{R}$ (números no negativos), que no está disponible en muchos otros cuerpos $\mathbb{F}$.

Un espacio vectorial real V, dotado de un producto escalar $\langle -, - \rangle$, se llamará un **espacio vectorial euclidiano**, o simplemente *espacio euclidiano*. (El concepto de ángulo recto viene de los *Elementos* de Euclides, un tratado escrito cerca de 300 a.C.)

► El ejemplo ya conocido de un producto escalar real es el *producto punto* de $V = \mathbb{R}^n$:

$$\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x} \cdot \boldsymbol{y}} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n. \tag{5.1}$$

Otro ejemplo es este producto escalar de funciones continuas $f, g \colon [a, b] \to \mathbb{R}$:

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

También se puede definir un producto escalar en $V = M_{m \times n}(\mathbb{R})$ – matrices rectangulares reales. Para escribirlo, se aprovecha la **traza** de una matriz cuadrada $n \times n$:

$$\underline{\operatorname{tr}(C)} := c_{11} + c_{22} + \cdots + c_{nn} = \sum_{k=1}^{n} c_{kk}\,. \tag{5.2}$$

Entonces (Ejemplo 5.4) se define:

$$\underline{\langle A, B \rangle := \operatorname{tr}(A^{\mathsf{t}} B)} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}\,.$$

Esta sumatoria hace obvio que $\langle A, B \rangle = \langle B, A \rangle$ y se puede chequear que esta receta es lineal en A y en B. Para la positividad (d), se nota que

$$\langle A, A \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \geqslant 0,$$

y esta suma de cuadrados *reales* es 0 si y solo si cada entrada es cero, $a_{ij} = 0$.

La norma y la desigualdad de Schwarz

En el capítulo 1 se estudió todos los aspectos del producto punto en $\mathbb{R}^n$ usando las propiedades (a), (b), (c) y (d) solamente, en lugar de hacer cálculos particulares. Por eso, las definiciones y las pruebas se transfieren sin cambio al caso general. Por ejemplo, la **norma** de un vector $\boldsymbol{x} \in V$ se define como

$$\underline{\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}} \geqslant 0,$$

y se generaliza la desigualdad de Cauchy en la siguiente **desigualdad de Schwarz** (Prop. 5.5):

$$\big|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\big| \leqslant \|\boldsymbol{x}\|\, \|\boldsymbol{y}\|. \tag{5.3}$$

con igualdad si y solo si $\boldsymbol{x}, \boldsymbol{y}$ son proporcionales.

(El caso de $\mathbb{R}^n$ fue notado por Cauchy en 1821. Fue mostrado para integrales por Viktor Bunyakovsky en 1859; y el caso general, aplicables a series infinitas, fue escrito por Hermann Schwarz en 1888.)

Se demuestra la desigualdad de Schwarz por el argumento del capítulo 1; se define una función cuadrática por:

$$\begin{aligned} f(t) &= \langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2t\langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2 \langle \boldsymbol{y}, \boldsymbol{y} \rangle \quad \text{porque} \quad \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}, \\ &= \|\boldsymbol{y}\|^2\, t^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle\, t + \|\boldsymbol{x}\|^2 \\ &=: at^2 + bt + c. \end{aligned}$$

Si $\boldsymbol{x} \neq 0$, entonces $a > 0$ y la relación "$f(t) \geqslant 0$ para todo t" implica $\underline{b^2 - 4ac \leqslant 0}$, de donde se deduce (5.3).

► También se define la **distancia** entre dos vectores $\boldsymbol{x}, \boldsymbol{y}$ por

$$\underline{d(\boldsymbol{x}, \boldsymbol{y})} := \|\boldsymbol{x} - \boldsymbol{y}\|,$$

y se define el **ángulo** θ entre $\boldsymbol{x}, \boldsymbol{y}$ (no nulos) por

$$\cos\theta := \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\, \|\boldsymbol{y}\|}, \quad \text{tomando} \quad 0 \leqslant \theta \leqslant \pi.$$

Las propiedades de la norma de vectores en un espacio euclidiano (Escolio 5.8) también imitan las del capítulo 1:

(a) $\|\alpha\boldsymbol{x}\| = |\alpha|\, \|\boldsymbol{x}\|$,,

(b) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$,

(c) $\|\boldsymbol{x}\| \geqslant 0$, con igualdad si y solo si $\boldsymbol{x} = \mathbf{0}$.

Cabe recordar que (b), la desigualdad triangular, se obtuvo con el uso de la desigualdad de Cauchy (y ahora, con la de Schwarz).

5.2. Clase 18

Esta es la clase virtual del 10 de junio del 2021. La temática comprende el inicio de la sección § 5.2: *Bases ortonormales*.

Ortogonalidad y bases ortonormales

En un espacio euclidiano V, dos vectores $\boldsymbol{x}, \boldsymbol{y}$ se llaman **ortogonales** si $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. Esto puede ocurrir si $\boldsymbol{x} = \boldsymbol{0}$ o si $\boldsymbol{y} = \boldsymbol{0}$ o si el ángulo θ entre $\boldsymbol{x}, \boldsymbol{y}$ cumple $\cos\theta = 0$, es decir, si $\theta = \frac{\pi}{2}$. Si escribe $\boldsymbol{x} \perp \boldsymbol{y}$ cuando $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

Lema 5.10: si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es un conjunto de vectores no nulos ortogonales entre sí, este conjunto es linealmente independiente. De hecho, si $c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m = \boldsymbol{0}$, entonces

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

mostrando que cada $c_j = 0$.

Si $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\} \subset V$ son vectores ortogonales entre sí y todos de longitud 1, esto es:

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = 0 \text{ si } j \neq k; \qquad \langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = \|\boldsymbol{e}_k\|^2 = 1 \text{ para todo } k,$$

se dice que $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ es una **familia ortonormal**. Si $n = \dim V$, entonces $m \leqslant n$.

En el caso $m = n$, esta es una *base* de V; y $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ se llama una **base ortonormal** (Defn. 5.12).

Dada una base ortonormal $\mathcal{E}$ de V, el producto escalar de dos vectores es el *producto punto* de sus componentes:

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= \left\langle \sum_{j=1}^n x_j \, \boldsymbol{e}_j, \sum_{k=1}^n y_k, \boldsymbol{e}_k \right\rangle = \sum_{j=1}^n \sum_{k=1}^n \langle x_j \, \boldsymbol{e}_j, y_k \, \boldsymbol{e}_k \rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n x_j y_k \, \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \sum_{k=1}^n x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \,. \end{aligned}$$

Los espacios original y dual

Si V es un espacio vectorial euclidiano con $\dim V = n$ finito, se sabe que el espacio vectorial dual V^* tiene la misma dimensión: $\dim V^* = n$. En el caso euclidiano, es posible *identificar V con V^**. Para ver eso, tome una base ortonormal de V, $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, y nótese que hay n formas lineales

$$f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle. \tag{5.5}$$

Hay una única aplicación lineal $J \colon V \to V^*$ determinado por $J\boldsymbol{e}_k := f_k$ para cada k. En detalle: si $\boldsymbol{y} \in V$, con $\boldsymbol{y} = \sum_{k=1}^n y_k \boldsymbol{e}_k$, entonces

$$J(\boldsymbol{y}) = \sum_{k=1}^n y_k f_k.$$

Al aplicar este elemento de V^* al vector $\boldsymbol{x} \in V$, se obtiene

$$J(\boldsymbol{y})(\boldsymbol{x}) = \sum_{k=1}^{n} y_k\, f_k(\boldsymbol{x}) = \sum_{k=1}^{n} y_k x_k = \langle \boldsymbol{y}, \boldsymbol{x} \rangle. \tag{5.6b}$$

Entonces la forma lineal $J(\boldsymbol{y})\colon \boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x} \rangle \in \mathbb{R}$ *no depende de la base ortonormal $\mathcal{E}$ usada para construirla*. O bien: $J\colon V \to V^*$ solo depende del producto escalar que V posee.

► Si $M \leqslant V$ es un subespacio, los vectores ortogonales a todo M forman otro subespacio:

$$\underline{M^\perp} := \{\, \boldsymbol{y} \in V : \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0, \text{ para todo } \boldsymbol{x} \in M \,\}. \tag{5.4}$$

Esto parece tener un conflicto de notación con el *anulador* $M^\perp$ del capítulo 4, por cuanto el anulador es un subespacio de V^*. Pero resulta que el anulador de $M \leqslant V$ no es otra cosa que la imagen bajo J del complemento ortogonal. En efecto:

$$J(\boldsymbol{y})(\boldsymbol{x}) = 0 \quad \text{para todo } \boldsymbol{x} \in M \qquad \Longleftrightarrow \quad \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \quad \text{para todo } \boldsymbol{x} \in M.$$

Cuando identificamos V con V^* mediante J, los dos usos del símbolo $M^\perp$ coinciden.

Transpuesta de una aplicación lineal

La identificación $J\colon V \to V^*$ también sirven para simplificar la **transpuesta** de una aplicación lineal $S\colon V \to W$ entre espacios euclidianos. Sea $J'\colon W \to W^*$ el mapa análogo para el caso de W.

Si $\boldsymbol{x}, \boldsymbol{z} \in V$ y si $\boldsymbol{y}, \boldsymbol{w} \in W$, se dispone de dos fórmulas:

$$J(\boldsymbol{z}) : \boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x} \rangle_V \quad \text{y} \quad J'(\boldsymbol{y}) : \boldsymbol{y} \mapsto \langle \boldsymbol{y}, \boldsymbol{w} \rangle_W$$

(etiquetando los dos productos escalares para distinguirlos).

Se quiere reemplazar la transpuesta $S^{\mathsf{t}}\colon W^* \to V^*$ con otra aplicación lineal de W en V. Esto tendría que ser $J^{-1}S^{\mathsf{t}}J'\colon W \to V$,

$$\begin{tikzcd} W \arrow[r, "??"] \arrow[d, "J'"'] & V \\ W^* \arrow[r, "S^{\mathsf{t}}"] & V^* \arrow[u, "J^{-1}"'] \end{tikzcd}$$

y el cálculo que sigue muestra su efecto:

$$\langle J^{-1}S^{\mathsf{t}}J'(\boldsymbol{y}), \boldsymbol{x} \rangle_V = S^{\mathsf{t}}J'(\boldsymbol{y})(\boldsymbol{x}) := J'(\boldsymbol{y})(S(\boldsymbol{x})) = \langle \boldsymbol{y}, S(\boldsymbol{x}) \rangle_W\,. \tag{5.7a}$$

con $\boldsymbol{z} := J^{-1}S^{\mathsf{t}}J'(\boldsymbol{y})$, $J(\boldsymbol{z}) = S^{\mathsf{t}}J'(\boldsymbol{y})$, $\boldsymbol{w} = S(\boldsymbol{x})$ en términos del despliegue anterior.

Ahora es cuestión de *renombrar* esta aplicación compuesta por S^{t} simplemente, obteniendo $\underline{S^{\mathsf{t}}\colon W \to V}$ como se quería. La fórmula anterior entonces se simplifica de esta manera:

$$\boxed{\langle S^{\mathsf{t}}(\boldsymbol{y}), \boldsymbol{x}\rangle = \langle \boldsymbol{y}, S(\boldsymbol{x})\rangle.} \tag{5.7b}$$

Esta es una nueva "fórmula mágica": en la presencia de productos escalares reales, *se obtiene la transpuesta* de una aplicación lineal nada más *con su movimiento de un lado al otro del producto escalar*.

Como el producto escalar real es simétrica en sus dos vectores, la fórmula anterior también se puede escribir así:

$$\langle S(\boldsymbol{x}), \boldsymbol{y}\rangle = \langle \boldsymbol{x}, S^{\mathsf{t}}(\boldsymbol{y})\rangle.$$

Esto evidencia que la transpuesta de S^{t} es la S original: $(S^{\mathsf{t}})^{\mathsf{t}} = S$.

► Si en vez de $S\colon V \to W$ consideramos la versión matricial $T_A\colon \mathbb{R}^n \to \mathbb{R}^m$ (o bien, si se toma A como la matriz de S con respecto a ciertas bases – ortonormales – de V y W), conviene escribir el producto punto en $\mathbb{R}^n$ como

$$\underline{\boldsymbol{x}^{\mathsf{t}}\boldsymbol{y}} = \boldsymbol{x}\cdot\boldsymbol{y} = \boldsymbol{y}\cdot\boldsymbol{x} = \boldsymbol{y}^{\mathsf{t}}\boldsymbol{x}.$$

⟦ Aquí $\boldsymbol{y}$ es un vector de columna, esto es, una matriz $n \times 1$; mientras $\boldsymbol{x}^{\mathsf{t}}$ es un *vector de fila*, o sea, una matriz $1 \times n$. Su producto es una matriz 1×1, esto es, un *escalar* $\boldsymbol{x}^{\mathsf{t}}\boldsymbol{y} \in \mathbb{R}$. ⟧

Se sabe que A^{t} es la matriz de S^{t}. Entonces la fórmula (5.7b) se reduce, es el presento caso, a un cálculo de matrices 1×1:

$$\boldsymbol{x}^{\mathsf{t}}A^{\mathsf{t}}\boldsymbol{y} = (A^{\mathsf{t}}\boldsymbol{y})^{\mathsf{t}}\boldsymbol{x} = \boldsymbol{y}^{\mathsf{t}}A\boldsymbol{x}. \tag{5.7c}$$

5.3. Clase 19

Esta es la clase virtual del 14 de junio del 2021. La temática comprende la segunda parte de la sección § 5.2: *Bases ortonormales*.

Construcción inductiva de una base ortonormal

En la clase anterior, se observó cómo se simplifica la teoría del espacio dual y las aplicaciones lineales transpuestas en un espacio vectorial euclidiano V, una vez que se dispone de una *base ortonormal* de V. Aquí V es un espacio vectorial sobre $\mathbb{R}$ dotado de un producto escalar real. Si $\dim V = n$ (finito), una base ortonormal es un juego de vectores $\underline{\mathcal{E} := \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}}$ que cumplen:

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = 0 \ \text{ si } j \neq k;$$
$$\|\boldsymbol{e}_k\|^2 = \langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = 1 \ \text{ para cada } k.$$

(Estas condiciones garantizan la independencia lineal del conjunto $\mathcal{E}$; como tiene exactamente $n = \dim V$ vectores, $\mathcal{E}$ es ciertamente una base de V.)

Queda una pregunta por contestar: ¿cómo se garantiza que V posee bases ortonormales? (Una de ellas sería suficiente para justificar lo que se ha hecho hasta ahora.) Resulta que hay una plétora de bases ortonormales de V; más aún, se puede *fabricar* o construir una base ortonormal a partir de una base cualquiera, por un procedimiento sistemático.

► Este proceso de construcción se llama el **algoritmo de Gram y Schmidt** (aunque la idea se remonta a Laplace). El *insumo* del algoritmo es una base cualquiera $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, no necesariamente ortonormal. Luego se calculan los n vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$ de *salida*, junto con $(n-1)$ vectores auxiliares $\boldsymbol{y}_2, \dots, \boldsymbol{y}_n$ que se descartan al final.

El algoritmo se resume en esta esquema:

$$\begin{aligned}
& & \boldsymbol{e}_1 &:= \boldsymbol{x}_1/\|\boldsymbol{x}_1\|; \\
\boldsymbol{y}_2 &:= \boldsymbol{x}_2 - \langle \boldsymbol{e}_1, \boldsymbol{x}_2 \rangle\, \boldsymbol{e}_1, & \boldsymbol{e}_2 &:= \boldsymbol{y}_2/\|\boldsymbol{y}_2\|; \\
\boldsymbol{y}_3 &:= \boldsymbol{x}_3 - \langle \boldsymbol{e}_1, \boldsymbol{x}_3 \rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_3 \rangle\, \boldsymbol{e}_2, & \boldsymbol{e}_3 &:= \boldsymbol{y}_3/\|\boldsymbol{y}_3\|; \\
&\ \vdots & &\ \vdots \\
\boldsymbol{y}_k &:= \boldsymbol{x}_k - \langle \boldsymbol{e}_1, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_{k-1}, & \boldsymbol{e}_k &:= \boldsymbol{y}_k/\|\boldsymbol{y}_k\|, \\
&\ \vdots & &\ \vdots \\
\boldsymbol{y}_n &:= \boldsymbol{x}_n - \langle \boldsymbol{e}_1, \boldsymbol{x}_n \rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_n \rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_n \rangle\, \boldsymbol{e}_{n-1}, & \boldsymbol{e}_n &:= \boldsymbol{y}_n/\|\boldsymbol{y}_n\|.
\end{aligned} \tag{5.8}$$

El primer paso es sencillo: como $\boldsymbol{x}_1 \neq \boldsymbol{0}$ y por ende $\|\boldsymbol{x}_1\| > 0$, se puede dividir $\boldsymbol{x}_1$ por su longitud, para obtener un vector $\boldsymbol{e}_1$ de longitud 1. En los pasos subsiguientes, se hace lo mismo con los vectores auxiliares $\boldsymbol{y}_2, \dots, \boldsymbol{y}_n$; esto garantiza que cada $\|\boldsymbol{e}_k\| = 1$.

Los vectores $\boldsymbol{y}_k$, por su parte, se obtienen de $\boldsymbol{x}_k$ al restarle ciertos múltiplos de los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$. Se trata de un proceso inductivo, o mejor dicho *iterativo*, en donde la producción de $\boldsymbol{y}_k$ y $\boldsymbol{e}_k$ depende de haber obtenido antes los vectores $\boldsymbol{e}_j$ para $j = 1, \dots, k$.

El algoritmo funciona por dos razones: (a) si $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}\}$ ya es una base ortonormal parcial, entonces $\boldsymbol{y}_k \perp \boldsymbol{e}_j$ para cada $j < k$; y (b) cada $\boldsymbol{y}_k \neq \boldsymbol{0}$, para poder dividir por $\|\boldsymbol{y}_k\|$.

De hecho, $\boldsymbol{y}_k \neq \boldsymbol{0}$ porque $\boldsymbol{x}_k$ no es una combinación lineal de los $\boldsymbol{e}_j$ con $j < k$. Y esa segunda afirmación es válida porque cada $\boldsymbol{e}_j$ sí es una combinación de los $\boldsymbol{x}_i$ para $i \leqslant j$. Como $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}, \boldsymbol{x}_k\}$ es linealmente independiente, eso dice que $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}, \boldsymbol{x}_k\}$ es linealmente independiente, y de ahí se ve que $\boldsymbol{y}_k \neq \boldsymbol{0}$.

La naturaleza iterativa del algoritmo garantiza otra cosa: para cada valor de k en $\{1, 2, 3, \dots, n\}$, el juego de vectores parcial $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ es una base ortonormal del subespacio $\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\rangle$ generado por un trozo inicial de la base original $\mathcal{B}$.

Un ejemplo de base ortonormal: polinomios de Legendre

Consideremos un espacio finitodimensional de polinomios reales, por ejemplo $\mathbb{R}_3[t]$. Para dotarlo de un producto escalar, se nota que cada polinomio $p(t)$ en $\mathbb{R}_3[t]$ puede ser considerado como una función continua $p\colon [-1, 1] \to \mathbb{R}$. Entonces se usa la recta del Ejemplo 5.3:

$$\langle p, q\rangle := \int_{-1}^{1} p(t)q(t)\,dt.$$

Los monomios $\{1, t, t^2, t^3\}$ forman una base de $\mathbb{R}_3[t]$, que no es ortonormal – el algoritmo de Gram y Schmidt producirá una base ortonormal $\{p_0, p_1, p_2, p_3\}$ de $\mathbb{R}_3[t]$.

Antes de efectuar los cálculos necesarios, conviene repasar las tres fórmulas más útiles sobre integrales en una variable real:[1]

- ◇ La integral $\int_{-a}^{a} g(t)\,dt = 0$ si $g(t)$ es una *función impar*.
- ◇ La integral $\int_{-a}^{a} h(t)\,dt = 2\int_{0}^{a} h(t)\,dt$ si $h(t)$ es una *función par*.
- ◇ Vale $\displaystyle\int_0^1 t^n\,dt = \frac{1}{n+1}$.

En el desarrollo del algoritmo, se denotará por q_1, q_2, q_3 los polinomios auxiliares.

Paso 0 El polinomio $p_0(t)$ es un múltiplo de la función constante 1:

$$\|1\|^2 = \langle 1, 1\rangle = \int_{-1}^{1} 1\,dt = 2; \qquad \|1\| = \sqrt{2}; \qquad \underline{p_0(t)} \equiv \frac{1}{\sqrt{2}}.$$

[1]La cuarta fórmula más útil, $\int_0^\pi \operatorname{sen}\theta\,d\theta = 2$, no es relevante para este ejemplo.

Paso 1 En el cálculo de $q_1(t)$, la integral $\langle p_0, t\rangle = 0$ al tener integrando impar:

$$\underline{q_1(t)} := t - \langle p_0, t\rangle p_0 = t - \frac{1}{\sqrt{2}} \int_{-1}^{1} \frac{t}{\sqrt{2}}\, dt = t - \frac{1}{2} \int_{-1}^{1} t\, dt = t - 0 = t.$$

$$\|q_1\|^2 = \int_{-1}^{1} t^2\, dt = 2 \int_{0}^{1} t^2\, dt = \frac{2}{3} \quad \text{así que} \quad \underline{p_1(t)} = \frac{t}{\sqrt{2/3}} = \sqrt{\frac{3}{2}}\, t.$$

Paso 2 Lo mismo sucede con $\langle p_1, t^2\rangle = 0$, pero $p_2(t)$ ya no es un monomio:

$$\begin{aligned} \underline{q_2(t)} := t^2 - \langle p_0, t^2\rangle p_0 - \langle p_1, t^2\rangle p_1 &= t^2 - \frac{1}{2} \int_{-1}^{1} t^2\, dt - \frac{3}{2}\, t \int_{-1}^{1} t^3\, dt \\ &= t^2 - \int_{0}^{1} t^2\, dt - 0 = t^2 - \frac{1}{3}\,. \end{aligned}$$

De ahí, se obtiene $p_2(t)$:

$$\|q_2\|^2 = \int_{-1}^{1} \bigl(t^2 - \tfrac{1}{3}\bigr)^2\, dt = 2 \int_{0}^{1} \bigl(t^4 - \tfrac{2}{3}t^2 + \tfrac{1}{9}\bigr)\, dt = 2\bigl(\tfrac{1}{5} - \tfrac{2}{9} + \tfrac{1}{9}\bigr) = \frac{8}{45},$$

$$\|q_2\| = \frac{2}{3}\sqrt{\frac{2}{5}}\,, \qquad \underline{p_2(t)} = \frac{3}{2}\sqrt{\frac{5}{2}}\, q_2(t) = \frac{1}{2}\sqrt{\frac{5}{2}}\, (3t^2 - 1).$$

Paso 3 El patrón se repite con el cálculo de $q_3(t)$ y $p_3(t)$:

$$\begin{aligned} \underline{q_3(t)} &:= t^3 - \langle p_0, t^3\rangle p_0 - \langle p_1, t^3\rangle p_1 - \langle p_2, t^3\rangle p_2 \\ &= t^3 - \frac{1}{2} \int_{-1}^{1} t^3\, dt - \frac{3}{2}\, t \int_{-1}^{1} t^4\, dt - \frac{5}{8}\,(3t^2 - 1) \int_{-1}^{1} (3t^5 - t^3)\, dt \\ &= t^3 - 0 - 3t \int_{0}^{1} t^4\, dt - 0 = t^3 - \frac{3}{5}\, t. \end{aligned}$$

Finalmente, se calcula $p_3(t)$ así:

$$\|q_3\|^2 = \int_{-1}^{1} \bigl(t^3 - \tfrac{3}{5}\, t\bigr)^2\, dt = 2 \int_{0}^{1} \bigl(t^6 - \tfrac{6}{5}t^4 + \tfrac{9}{25}t^2\bigr)\, dt = 2\bigl(\tfrac{1}{7} - \tfrac{6}{25} + \tfrac{3}{25}\bigr) = \frac{8}{175},$$

$$\|q_3\| = \frac{2}{5}\sqrt{\frac{2}{7}}\,, \qquad \underline{p_3(t)} = \frac{5}{2}\sqrt{\frac{7}{2}}\, q_2(t) = \frac{1}{2}\sqrt{\frac{7}{2}}\, (5t^3 - 3t).$$

Conviene suprimir los radicales, al introducir $\underline{P_k(t)} := \sqrt{2/(2k+1)}\, p_k(t)$. Entonces

$$P_0(t) = 1, \quad P_1(t) = t, \quad P_2(t) = \tfrac{1}{2}(3t^2 - 1), \quad P_3(t) = \tfrac{1}{2}(5t^3 - 3t).$$

Estos $P_k(t)$, y otros para $k = 4, 5, \dots$ se llaman **polinomios de Legendre**. Forman una familia de *polinomios ortogonales* sobre el intervalo $[-1, 1]$.

5.4. Clase 20

Esta es la clase virtual del 21 de junio del 2021. La temática comprende la primera parte de la sección § 5.3: *Productos escalares complejos*.

Números complejos

Hay dos tipos de productos escalares, dependiendo de si el cuerpo de escalares es $\mathbb{R}$ ó $\mathbb{C}$. Ya se ha vista el caso real. En esta clase V será un espacio vectorial sobre $\mathbb{C}$. Si $\boldsymbol{z}, \boldsymbol{w} \in V$, su producto escalar será un número complejo $\langle \boldsymbol{z}, \boldsymbol{w} \rangle \in \mathbb{C}$.

Un número complejo tiene la forma $\underline{\alpha = s + it}$, donde $s, t \in \mathbb{R}$, y por regla vale $i^2 = -1$. El **conjugado complejo** de α es

$$\bar{\alpha} := s - it.$$

(Si $t = 0$, entonces $\alpha = s = \bar{\alpha}$ es real; pero si $t \neq 0$, entonces $\alpha \neq \bar{\alpha}$ y $\alpha \notin \mathbb{R}$.)

Debido a que

$$\alpha\bar{\alpha} = \bar{\alpha}\alpha = (s - it)(s + it) = s^2 + t^2 \geqslant 0,$$

se puede definir el **valor absoluto** de α como

$$|\alpha| := \sqrt{\bar{\alpha}\alpha} = \sqrt{s^2 + t^2} \geqslant 0.$$

Es de notar que $|\alpha| = 0 \iff s = t = 0 \iff \alpha = 0$ en $\mathbb{C}$. También, es útil observar que $\underline{|\alpha\beta| = |\alpha|\,|\beta|}$ para $\alpha, \beta \in \mathbb{C}$.

Un número complejo tal que $\underline{|\alpha| = 1}$ cumple $s^2 + t^2 = 1$; por eso, hay un ángulo θ con $s = \cos\theta$, $t = \operatorname{sen}\theta$. En tal caso $\underline{\alpha = \cos\theta + i\operatorname{sen}\theta}$.

Es fácil calcular que

$$(\cos\theta + i\operatorname{sen}\theta)^2 = \cos^2\theta + 2i\operatorname{sen}\theta\cos\theta - \operatorname{sen}^2\theta = \cos(2\theta) + i\operatorname{sen}(2\theta),$$

y también

$$\alpha^{-1} = (\cos\theta + i\operatorname{sen}\theta)^{-1} = \cos\theta - i\operatorname{sen}\theta = \bar{\alpha}$$

porque $(\cos\theta + i\operatorname{sen}\theta)(\cos\theta - i\operatorname{sen}\theta) = \cos^2\theta + \operatorname{sen}^2\theta = 1$ es decir, $\alpha\bar{\alpha} = 1$. Hay una *fórmula de de Moivre*:

$$(\cos\theta + i\operatorname{sen}\theta)^n = \cos(n\theta) + i\operatorname{sen}(n\theta) \quad \text{para todo} \quad n \in \mathbb{Z},$$

que se puede comprobar por inducción sobre n. Eso justifica ampliar la "ley de exponentes" a exponentes imaginarios al definir

$$\underline{e^{i\theta} := \cos\theta + i\operatorname{sen}\theta}, \quad \text{así que} \quad (e^{i\theta})^n = e^{in\theta}.$$

En general, si $\alpha \neq 0$ en $\mathbb{C}$, se pone $\underline{|\alpha| = r} > 0$ y eso permite escribir

$$\alpha = r(\cos\theta + i \operatorname{sen}\theta) = r\, e^{i\theta}.$$

Esta es la llamada *forma polar* del número complejo α; el par (r, θ) son las coordenadas polares del punto $(s, t) \in \mathbb{R}^2$.

Productos escalares complejos

Se puede introducir un producto escalar en V (Defn. 5.18). Si $\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{v} \in V$ y $\alpha \in \mathbb{C}$:

(a) $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{z} \rangle}$,

(b) $\langle \boldsymbol{z}, \boldsymbol{w} + \boldsymbol{v} \rangle = \langle \boldsymbol{z}, \boldsymbol{w} \rangle + \langle \boldsymbol{z}, \boldsymbol{v} \rangle$,

(c) $\langle \boldsymbol{z}, \alpha\, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle$, pero $\langle \alpha\, \boldsymbol{z}, \boldsymbol{w} \rangle = \bar{\alpha} \langle \boldsymbol{z}, \boldsymbol{w} \rangle$,

(d) $\langle \boldsymbol{z}, \boldsymbol{z} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{z} = \boldsymbol{0}$ en V.

El espacio $\mathbb{C}$-vectorial V, dotado de $\langle -, - \rangle$, se llama un **espacio hilbertiano**.

Se puede notar dos diferencias con el producto escalar real: en el inciso (a), $\langle \boldsymbol{z}, \boldsymbol{w} \rangle$ *ya no es simétrico* en los dos vectores, el intercambio $\boldsymbol{z} \leftrightarrow \boldsymbol{w}$ pasa su valor a su conjugado complejo. En el inciso (c), el escalar α "sale" del segundo vector sin cambio, pero "sale" del primer factor como su conjugado $\bar{\alpha}$.

Combinando (b) con (c), se ve que $\langle \boldsymbol{z}, \boldsymbol{w} \rangle$ es *lineal en la segunda variable* $\boldsymbol{w}$, pero sólo es *semilineal en la primera variable* $\boldsymbol{z}$. Se dice que una aplicación $T \colon V \to W$ es **semilineal** (o *antilineal*) si

$$T(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \bar{\alpha}\, T(\boldsymbol{x}) + \bar{\beta}\, T(\boldsymbol{y})$$

para $\boldsymbol{x}, \boldsymbol{y} \in V$, $\alpha, \beta \in \mathbb{C}$.

Con ese vocabulario, también se dice que $\langle \boldsymbol{z}, \boldsymbol{w} \rangle$ es **sesquilineal** (esto es, $\frac{3}{2}$-lineal) en sus dos argumentos.

► El ejemplo principal es similar, pero no igual, al producto punto de $\mathbb{R}^n$. Si $V = \mathbb{C}^n$, se define

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle \equiv \bar{\boldsymbol{z}} \cdot \boldsymbol{w} \equiv \bar{\boldsymbol{z}}^{\mathsf{t}} \boldsymbol{w} := \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n. \tag{5.9}$$

Se ve que $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = |z_1|^2 + \cdots + |z_n|^2 = 0 \iff \text{cada } z_i = 0 \iff \boldsymbol{z} = \boldsymbol{0}$ en $\mathbb{C}^n$.

Nótese que en esta fórmula, $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \bar{\boldsymbol{z}}^{\mathsf{t}} \boldsymbol{w}$ es lineal en el vector de columna $\boldsymbol{w}$, pero semilineal en el vector de fila $\boldsymbol{z}^{\mathsf{t}}$, como debe ser.

(Algunos libros de matemáticas toman el producto escalar lineal en la primera variable y semilineal en la segunda variable: eso es un error de procedimiento.)

► Al igual que en el caso real, se puede definir un producto escalar de matrices en $M_{m\times n}(\mathbb{C})$. Esto requiere distinguir 4 matrices:

$$A = [\alpha_{ij}], \qquad A^{\mathsf{t}} = [\alpha_{ji}], \qquad \overline{A} := [\bar{\alpha}_{ij}], \qquad A^* := [\bar{\alpha}_{ji}].$$

Si A es $m \times n$, su **matriz conjugada** $\overline{A}$ es también $m \times n$.

En cambio, su *matriz transpuesta* A^{t} y su **matriz adjunta** A^* son matrices $n \times m$.

Nótese que las transformaciones $A \mapsto \overline{A}$ y $A \mapsto A^*$ son *semilineales*. La primera mantiene el orden de productos, la segunda lo cambia:

$$\overline{AB} = \overline{A}\,\overline{B}, \quad \text{mientras} \quad (AB)^* = B^*A^*.$$

Ahora (Ejemplo 5.22), para $A, B \in M_{m\times n}(\mathbb{C})$ se puede definir

$$\langle A, B\rangle := \underline{\operatorname{tr}(A^*B)} = \sum_{i=1}^{m}\sum_{j=1}^{n} \bar{\alpha}_{ij}\beta_{ij}\,.$$

Se nota que A^*B es una matriz compleja $n \times n$. Además,

$$\langle A, A\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} |\alpha_{ij}|^2 \geqslant 0,$$

con $\langle A, A\rangle = 0 \iff$ cada $\alpha_{ij} = 0 \iff A = 0$ en $M_{m\times n}(\mathbb{C})$.

La desigualdad de Schwarz, de nuevo

Se define la **norma** de $z \in V$ por $\|z\| := \sqrt{\langle z, z\rangle}$, al igual que en el caso real.

La desigualdad de Schwarz también se cumple:

$$\bigl|\langle z, w\rangle\bigr| \leqslant \|z\|\,\|w\| \quad \text{para todo} \quad z, w \in V. \tag{5.3}$$

Sin embargo, su demostración para escalares en $\mathbb{C}$ requiere una discusión extra.

Sean $z \neq \mathbf{0}$ y $w \neq \mathbf{0}$ en V. [Si uno de ellos es $\mathbf{0}$, (5.3) es obvio, con igualdad.]

Hay dos casos: (1) si $\langle z, w\rangle \in \mathbb{R}$, se procede como antes, poniendo para $t \in \mathbb{R}$,

$$f(t) := \|z + tw\|^2 = at^2 + bt + c$$

con $a > 0$; la condición $f(t) \geqslant 0$ para todo $t \in \mathbb{R}$ implica $b^2 - 4ac \leqslant 0$ y de ahí sigue la desigualdad (5.3).

Caso (2): si $\langle z, w\rangle \notin \mathbb{R}$, se escribe $\langle z, w\rangle =: re^{i\theta}$ en forma polar y se define $\underline{v := e^{-i\theta}w} \in V$. Entonces

$$\langle z, v\rangle = \langle z, e^{-i\theta}w\rangle = e^{-i\theta}\langle z, w\rangle = e^{-i\theta}(re^{i\theta}) = r \in \mathbb{R}.$$

Del caso (1), se obtiene $\bigl|\langle z, v\rangle\bigr| \leqslant \|z\|\,\|v\|$.

Ahora se nota que $\|v\| = \|w\|$, porque

$$\|v\|^2 = \langle v, v\rangle = \langle e^{-i\theta}w, e^{-i\theta}w\rangle = e^{+i\theta}e^{-i\theta}\langle w, w\rangle = \|w\|^2.$$

La desigualdad deseada sigue:

$$\bigl|\langle z, w\rangle\bigr| = \bigl|\langle z, e^{i\theta}v\rangle\bigr| = \bigl|e^{i\theta}\langle z, v\rangle\bigr| = \bigl|\langle z, v\rangle\bigr| \leqslant \|z\|\,\|v\| = \|z\|\,\|w\|.$$

► Resulta que la norma de un espacio hilbertiano determina su producto escalar, mediante la siguiente **fórmula de polarización** (Lema 5.24):

$$4\,\langle w, z\rangle = \|z + w\|^2 - \|z - w\|^2 + i\,\|z + i\,w\|^2 - i\,\|z - i\,w\|^2. \tag{5.10}$$

No es difícil verificar (5.10). Del cálculo

$$\|z \pm w\|^2 = \langle z \pm w, z \pm w\rangle = \|z\|^2 \pm \langle z, w\rangle \pm \langle w, z\rangle + \|w\|^2$$

se ve en seguida que

$$\|z + w\|^2 - \|z - w\|^2 = 2\langle z, w\rangle + 2\langle w, z\rangle. \tag{$*$}$$

Otro cálculo similar:

$$\|z \pm iw\|^2 = \langle z \pm iw, z \pm iw\rangle = \|z\|^2 \pm i\langle z, w\rangle \mp i\langle w, z\rangle + \|w\|^2,$$

produce un resultado parecido,

$$i\|z + iw\|^2 - i\|z - iw\|^2 = -2\langle z, w\rangle + 2\langle w, z\rangle. \tag{$**$}$$

Al sumar las dos expresiones ($*$) y ($**$), se obtiene la fórmula (5.10).

Sobre la undécima lista de ejercicios: "Tarea 11"

Ejercicio 5.1 La desigualdad de Schwarz (o de Cauchy) tiene muchas aplicaciones. Un resultado útil en cálculos de estadística dice que la media aritmética de varios números positivos es menor o igual que su media cuadrática – y se identifica el caso de igualdad.

Ejercicio 5.4 El producto escalar de funciones continuas sobre $[-1, 1]$, que dio lugar a los *polinomios de Legendre* (Ejemplo 5.16), no es la única posibilidad. Con otro producto escalar se obtiene la familia de *polinomios de Chebyshev*, esta vez sin la necesidad de aplicar el algoritmo de Gram y Schmidt.

Ejercicio 5.8 El algoritmo de Gram y Schmidt ofrece una manera (aunque no la más eficiente) de verificar independencia lineal: si un determinado juego de vectores es linealmente dependiente, el algoritmo se detendrá sin éxito. Si eso no sucede, los vectores dados son independientes.

Ejercicio 5.10 La norma asociada a un producto escalar obedece una generalización del teorema de Pitágoras, llamada la *igualdad de Parseval*: la longitud al cuadrado del vector es la suma de los cuadrados de (los valores absolutos de) sus componentes en una base ortonormal.

5.5. Clase 21

Esta es la clase virtual del 24 de junio del 2021. La temática comprende la segunda parte de la sección § 5.3: *Productos escalares complejos*.

Bases ortonormales en espacios hilbertianos

En un espacio vectorial V sobre $\mathbb{C}$ con un producto escalar complejo, se define una base ortonormal $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ como antes:

$$\begin{aligned} \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle &= 0 \ \text{ si } j \neq k; \\ \|\boldsymbol{e}_k\|^2 = \langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle &= 1 \ \text{ para cada } k. \end{aligned}$$

El algoritmo de Gram y Schmidt se aplica sin cambio alguno para convertir una base vectorial cualquiera $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V en una base ortonormal.

Sólo se debe notar un detalle: en el paso número k del algoritmo:

$$\boldsymbol{y}_k := \boldsymbol{x}_k - \langle \boldsymbol{e}_1, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k \rangle\, \boldsymbol{e}_{k-1}, \qquad \boldsymbol{e}_k := \boldsymbol{y}_k / \|\boldsymbol{y}_k\|, \tag{5.8'}$$

los coeficientes $\langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle$ no deben cambiarse en $\langle \boldsymbol{x}_k, \boldsymbol{e}_j \rangle$, porque éstos pueden ser diferentes, en general.

La *base estándar* de $\mathbb{C}^n$, donde $\boldsymbol{e}_k = (0, 0, \dots, 1, \dots, 0)^t$ con el 1 en la k-ésima posición, es una base ortonormal.

El espacio dual de un espacio hilbertiano

Hay una diferencia significativa con el caso real: la aplicación $\underline{J\colon V \to V^*}$ que hace corresponder una base ortonormal de V con la base dual $\mathcal{F} = \{f_1, \dots, f_n\}$ de V^*, determinado por $\underline{J(\boldsymbol{e}_k) := f_k}$ *ya no es lineal, sino semilineal*:

$$J(\boldsymbol{y}) \equiv J\left(\textstyle\sum_{k=1}^n y_k \boldsymbol{e}_k\right) := \textstyle\sum_{k=1}^n \bar{y}_k f_k. \tag{5.6a$'$}$$

Cada forma lineal f_k está dada por

$$f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle \quad \text{para} \quad k = 1, \dots, n. \tag{5.5$'$}$$

y la semilinealidad de J garantiza que

$$J(\boldsymbol{y})(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k\, f_k(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k x_k = \langle \boldsymbol{y}, \boldsymbol{x} \rangle. \tag{5.6b$'$}$$

En breve: cada $J(\boldsymbol{y}) \in V^*$ es $\underline{\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x} \rangle}$, así que J no depende de la base ortonormal usada.

La aplicación lineal adjunta

Otra diferencia con el caso real es que la fórmula que simplificaba la aplicación lineal transpuesta ahora define algo diferente. Dado $S\colon V \to W$ lineal, donde V y W son espacios hilbertianos, se dispone de dos correspondencias:

$$\begin{aligned} J &: V \to V^* \\ J(\boldsymbol{z}) &: \boldsymbol{x} \mapsto \langle \boldsymbol{z}, \boldsymbol{x} \rangle_V \end{aligned} \qquad \text{y} \qquad \begin{aligned} J' &: W \to W^* \\ J'(\boldsymbol{y}) &: \boldsymbol{y} \mapsto \langle \boldsymbol{y}, \boldsymbol{w} \rangle_W \end{aligned}$$

que son semilineales (e invertibles).

Ahora (Defn. 5.26) se define la aplicación lineal **adjunta** $\underline{S^*\colon W \to V}$ por

$$S^* := J^{-1} S^{\mathsf{t}} J', \qquad \text{o sea,} \qquad \begin{tikzcd} W \arrow[r, "S^*"] \arrow[d, "J'"'] & V \\ W^* \arrow[r, "S^{\mathsf{t}}"] & V^* \arrow[u, "J^{-1}"'] \end{tikzcd}$$

La fórmula es la misma que se usó para redefinir S^{t} en el caso real. Pero los productos escalares son otros! El mismo cálculo de antes,

$$\langle J^{-1} S^{\mathsf{t}} J'(\boldsymbol{y}), \boldsymbol{x} \rangle_V = S^{\mathsf{t}} J'(\boldsymbol{y})(\boldsymbol{x}) := J'(\boldsymbol{y})(S(\boldsymbol{x})) = \langle \boldsymbol{y}, S(\boldsymbol{x}) \rangle_W \tag{5.11a}$$

produce un resultado sutilmente diferente:

$$\boxed{\langle S^*(\boldsymbol{y}), \boldsymbol{x}\rangle = \langle \boldsymbol{y}, S(\boldsymbol{x})\rangle} \quad \text{para todo} \quad \boldsymbol{x} \in V,\ \boldsymbol{y} \in W. \tag{5.11b}$$

Este es una variante de la "fórmula mágica" (5.7b): en la presencia de productos escalares **complejos**, se obtiene la **adjunta** de una aplicación lineal nada más *con su movimiento de un lado al otro del producto escalar*.

Al tomar el conjugado complejo de ambos lados de (5.11b), se obtiene

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

Esto dice que la adjunta de S^* es la S original: $\underline{(S^*)^* = S}$.

La transformación $S \mapsto S^*$ revierte el orden de composición. Si $R\colon W \to Z$ es lineal (donde Z es otro espacio hilbertiano), se obtiene $\underline{(RS)^* = S^*R^*}\colon Z \to V$, del cálculo:

$$\langle (RS)^*(\boldsymbol{z}), \boldsymbol{x}\rangle = \langle \boldsymbol{z}, RS(\boldsymbol{x})\rangle = \langle R^*(\boldsymbol{z}), S(\boldsymbol{x})\rangle = \langle S^*(R^*(\boldsymbol{z})), \boldsymbol{x}\rangle.$$

5.6. Clase 22

Esta es la clase virtual del 28 de junio del 2021. La temática comprende la primera parte de la sección § 5.4: *Matrices ortogonales y positivas*.

Matrices ortogonales y unitarias

En esta sección, aprovechamos los productos escalares y las bases ortonormales para introducir y discutir algunas clases de matrices de especial interés. De nuevo, las matrices reales y matrices complejas aparecen en casos paralelos. En este primer curso de álgebra lineal, se dedicará más atención al caso real.

► Sea $A \in M_n(\mathbb{R})$ una matriz $n \times n$ con entradas reales. Sus columnas $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$ son vectores en $\mathbb{R}^n$. Decimos que A es una **matriz ortogonal** *si esas columnas forman una base ortonormal* de $\mathbb{R}^n$.

Esto significa que las columnas de A cumplen:

$$\begin{aligned} \boldsymbol{a}_i^{\mathsf{t}}\, \boldsymbol{a}_j &= \boldsymbol{a}_i \cdot \boldsymbol{a}_j = 0 \ \text{ si } i \neq j; \\ \|\boldsymbol{a}_j\|^2 &= \boldsymbol{a}_j \cdot \boldsymbol{a}_j = 1 \ \text{ para cada } j. \end{aligned}$$

Dicho de otro modo: las entradas $\boldsymbol{a}_i^{\mathsf{t}}\, \boldsymbol{a}_j$ de la matriz $A^{\mathsf{t}}A$ coinciden con las entradas de la matriz identidad 1_n. En breve: $\underline{A^{\mathsf{t}}A = 1_n}$.

En particular, las columnas de A son linealmente independientes; por eso, vale $r(A) = n$. La nulidad es $n(A) = 0$ porque $n(A) + r(A) = n$ (el teorema de rango y nulidad). Entonces T_A es biyectiva y *A es invertible*.

Como A^{-1} existe, la igualdad $A^{\mathsf{t}}A = 1_n$ dice que $A^{-1} = A^{\mathsf{t}}$. Esto, a su vez, implica que $AA^{\mathsf{t}} = 1_n$ (y viceversa). En síntesis:

$$A \text{ es ortogonal} \iff A^{\mathsf{t}}A = 1_n \iff A^{-1} = A^{\mathsf{t}} \iff AA^{\mathsf{t}} = 1_n\,. \tag{5.12}$$

► Las matrices ortogonales dan lugar a una nueva manera de factorizar matrices. Supóngase que X es una matriz real $n \times n$ invertible (para no complicarse demasiado). A sus columnas $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ – que son l.i., pues $r(X) = n$ al ser X invertible – dan lugar a una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de $\mathbb{R}^n$, mediante el algoritmo de Gram y Schmidt. Estas son las columnas de una *matriz ortogonal Q*.

El paso # k del algoritmo da la siguiente relación:

$$\boldsymbol{x}_k = \langle \boldsymbol{e}_1, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_1 + \langle \boldsymbol{e}_2, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_2 + \cdots + \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_{k-1} + \|\boldsymbol{y}_k\|\, \boldsymbol{e}_k. \tag{5.8″}$$

Se define una *matriz triangular superior R* al poner:

$$r_{kk} := \|\boldsymbol{y}_k\| > 0, \qquad r_{jk} := \langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle = \boldsymbol{e}_j \cdot \boldsymbol{x}_k \quad \text{para} \quad j < k.$$

Entonces es fácil chequear que $\underline{X = QR}$. Esta es la llamada “factorización QR” de la matriz invertible X.

► Hay conceptos análogos en el caso complejo. Las diferencias se deben a que el producto escalar en $\mathbb{C}^n$ es $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \bar{\boldsymbol{u}} \cdot \boldsymbol{v} = \bar{\boldsymbol{u}}^{\mathsf{t}}\, \boldsymbol{v}$: en vez de la matriz transpuesta A^{t} hay que usar la *matriz adjunta A^**.

Entonces $U \in M_n(\mathbb{C})$ se llama una **matriz unitaria** si sus columnas forman una base ortonormal $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ de $\mathbb{C}^n$. Esta U es también invertible y resulta que:

$$U \text{ es unitaria} \iff U^*U = 1_n \iff U^{-1} = U^* \iff UU^* = 1_n\,. \tag{5.13}$$

Matrices simétricas y positivas

Otra clase de matrices (reales, por ahora) es ligada a los valores de $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \boldsymbol{x}^{\mathsf{t}}A\boldsymbol{x} \in \mathbb{R}$ mientras $\boldsymbol{x}$ recorre el espacio vectorial $\mathbb{R}^n$.

Recuerde que una matriz cuadrada B es *simétrica* si $\underline{B^{\mathsf{t}} = B}$. Si $A \in M_{m\times n}(\mathbb{R})$ es una matriz rectangular, entonces $A^{\mathsf{t}}A$ es una matriz $n \times n$ simétrica, porque:

$$(A^{\mathsf{t}}A)^{\mathsf{t}} = A^{\mathsf{t}}A^{\mathsf{tt}} = A^{\mathsf{t}}A.$$

La matriz $A^{\mathsf{t}}A$ tiene una propiedad curiosa (Prop. 5.31): su rango es igual al rango de A; esto es, $\underline{r(A^{\mathsf{t}}A) = r(A)}$.

Para demostrar esta igualdad, se usa dos técnicas. Primero, se consideran los vectores $A\mathbf{x} \in \mathbb{R}^m$ y $A^{\mathsf{t}}A\mathbf{x} \in \mathbb{R}^n$, con $\mathbf{x} \in \mathbb{R}^n$.

Es obvio que $A\mathbf{x} = \mathbf{0} \implies A^{\mathsf{t}}A\mathbf{x} = \mathbf{0}$. En la dirección contraria:

$$\begin{aligned} A^{\mathsf{t}}A\mathbf{x} = \mathbf{0} &\implies \mathbf{x}^{\mathsf{t}}A^{\mathsf{t}}A\mathbf{x} = 0 \\ &\implies (A\mathbf{x})^{\mathsf{t}}A\mathbf{x} = 0 \implies \|A\mathbf{x}\|^2 = 0 \implies A\mathbf{x} = \mathbf{0}. \end{aligned}$$

Esto muestra que $A\mathbf{x} = \mathbf{0}$ (en $\mathbb{R}^m$) si y solo si $A^{\mathsf{t}}A\mathbf{x} = \mathbf{0}$ (en $\mathbb{R}^n$). Y de ahí, las dos matrices tiene la misma *nulidad*: $\underline{n(A) = n(A^{\mathsf{t}}A)}$.

El segundo paso es *apelar al teorema de rango y nulidad*, que comprueba la igualdad:

$$r(A^{\mathsf{t}}A) = n - n(A^{\mathsf{t}}A) = n - n(A) = r(A).$$

El teorema de rango y nulidad también dice que $\underline{r(A^{\mathsf{t}}) = r(A)}$. Al intercambiar $A \leftrightarrow A^{\mathsf{t}}$, se obtiene otra igualdad: $r(AA^{\mathsf{t}}) = r(A^{\mathsf{t}}) = r(A)$.

► La Defn. 5.33 dice que una matriz simétrica real $A = A^{\mathsf{t}}$ es una **matriz positiva** (o bien, una *matriz semidefinida positiva*) si satisface

$$\langle \mathbf{x}, A\mathbf{x} \rangle \geqslant 0 \quad \text{para todo} \quad \mathbf{x} \in \mathbb{R}^n. \tag{5.14}$$

O bien: $\underline{\mathbf{x}^{\mathsf{t}}A\mathbf{x} \geqslant 0}$ para $\mathbf{x} \in \mathbb{R}^n$.

A es **definida positiva** si $\langle \mathbf{x}, A\mathbf{x} \rangle > 0$ para $\mathbf{x} \neq \mathbf{0}$.

Las matrices de tipo $\underline{A = B^{\mathsf{t}}B}$ – con $B \in M_n(\mathbb{R})$ – son positivas, porque

$$\mathbf{x}^{\mathsf{t}}A\mathbf{x} = \mathbf{x}^{\mathsf{t}}B^{\mathsf{t}}B\mathbf{x} = (B\mathbf{x})^{\mathsf{t}}B\mathbf{x} = \|B\mathbf{x}\|^2 \geqslant 0. \tag{5.15}$$

(De hecho, aunque no es posible dar una prueba todavía, cualquier matriz positiva A tiene una "raíz cuadrada" simétrica $B = B^{\mathsf{t}}$ tal que $B^{\mathsf{t}}B = B^2 = A$.)

⟦ En el caso complejo, una matriz cuadrada $A \in M_n(\mathbb{C})$ se llama positiva (semidefinida) si (5.14) es válida para todo $\mathbf{x} \in \mathbb{C}^n$. La condición $A = A^{\mathsf{t}}$ se cambia por $A = A^*$ (se dice que A es *hermítica* cuando $A = A^*$). En contraste con el caso real, resulta que una matriz compleja que cumple (5.15) para $\mathbf{x} \in \mathbb{C}^n$ es automáticamente hermítica. ⟧

Dada una matriz simétrica real que es invertible, ¿cómo es posible ver si es definida positiva sin tener que examinar todos los posibles valores de $\mathbf{x}^{\mathsf{t}}A\mathbf{x}$? Resulta que hay dos *criterios* bien conocidos para detectar positividad: uno que emplea determinantes (en el capítulo siguiente) y otro que usa la eliminación gaussiana (próximamente).

5.7. Clase 23

Esta es la clase virtual del 1 de julio del 2021. La temática comprende la segunda parte de la sección § 5.4: *Matrices ortogonales y positivas*.

Matrices definidas positivas

Hoy, A es una matriz real simétrica: $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$. La matriz A es:

- **positiva** si $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \underline{\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} \geqslant 0}$ para todo $\boldsymbol{x} \in \mathbb{R}^n$,
- **definida positiva** si $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \underline{\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} > 0}$ para $\boldsymbol{x} \neq \mathbf{0}$.

Si A es *definida* positiva, entonces es positiva (obviamente) y también invertible:

$$A\boldsymbol{x} = \mathbf{0} \implies \boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} = 0 \implies \boldsymbol{x} = \mathbf{0},$$

y por eso su nulidad es $n(A) = 0$ y su rango es $r(A) = n$; por eso, A es invertible.

La Prop. 5.36 dice que una matriz positiva e invertible tiene que ser *definida* positiva. Para ver eso, se considera este apareamiento de vectores:

$$\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle := \boldsymbol{x}^{\mathsf{t}} A \boldsymbol{y}. \tag{5.16}$$

Eso es *casi* un producto escalar: es simétrico bajo $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$, es lineal en $\boldsymbol{y}$ y tiene valores no negativos. Por ejemplo,

$$\begin{aligned} \langle\!\langle \boldsymbol{y}, \boldsymbol{x} \rangle\!\rangle &= \boldsymbol{y}^{\mathsf{t}} A \boldsymbol{x} = \boldsymbol{x}^{\mathsf{t}} A^{\mathsf{t}} \boldsymbol{y} = \boldsymbol{x}^{\mathsf{t}} A \boldsymbol{y} = \langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle, \\ \langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle &= \boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} \geqslant 0. \end{aligned}$$

Lo que falta (todavía) es la propiedad $\langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle = 0 \implies \boldsymbol{x} = \mathbf{0}$.

Sin embargo, aun sin esa última propiedad, la desigualdad de Schwarz es válida:

$$\bigl|\langle\!\langle \boldsymbol{x}, \boldsymbol{y} \rangle\!\rangle\bigr|^2 \leqslant \langle\!\langle \boldsymbol{x}, \boldsymbol{x} \rangle\!\rangle \, \langle\!\langle \boldsymbol{y}, \boldsymbol{y} \rangle\!\rangle \quad \text{para} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

(Para verla, se repite la prueba usando la función $f(t) := \langle\!\langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle\!\rangle$. Solo hay que omitir el caso de igualdad.) Esto es lo mismo que

$$\bigl(\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{y}\bigr)^2 \leqslant (\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x})(\boldsymbol{y}^{\mathsf{t}} A \boldsymbol{y}) \quad \text{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

Ahora tómese algún vector $\boldsymbol{x}$ tal que $\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x} = 0$. La desigualdad anterior muestra que también vale $\underline{\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{y} = 0}$ para todo $\boldsymbol{y}$. Si, por ejemplo, se usa $\underline{\boldsymbol{y} := A^{-1}\boldsymbol{x}}$ – puesto que A es invertible – se obtiene $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^{\mathsf{t}} \boldsymbol{x} = 0$ y por ende $\boldsymbol{x} = \mathbf{0}$, como se quería. ⟦ Fíjese que ahora (5.16) sí define un verdadero producto escalar sobre $\mathbb{R}^n$. ⟧

Positividad definida por eliminación gaussiana

Ahora sea $A = A^{\mathsf{t}} \in M_n(\mathbb{R})$ una matriz real simétrica *invertible*. (No necesariamente positiva.)

Se puede ensayar eliminación gaussiana *simple* a esta matriz A (sin intercambiar filas). Si es exitosa, se obtendrá un juego de *pivotes no ceros* $a_{kk}^{(k)} \neq 0$ para $k = 1, \dots, n$; y eso da una factorización $A = LDU$ que es *única*. Pero $A = A^{\mathsf{t}} = U^{\mathsf{t}} D L^{\mathsf{t}}$ es una segunda factorización del mismo tipo. Se deduce que $U = L^{\mathsf{t}}$ (y $L = U^{\mathsf{t}}$), de modo que la "factorización LDU" es $\underline{A = LDL^{\mathsf{t}}}$.

► La Prop. 5.37 aborda el caso en donde A es definida positiva. Para $k = 1$, se ve que

$$a_{11}^{(1)} = a_{11} = \langle \boldsymbol{e}_1, A\,\boldsymbol{e}_1 \rangle > 0.$$

Después de varias operaciones de fila de "tipo (c)" solamente, se produce $A \mapsto L_k A$, donde las primeras $(k-1)$ filas tienen ceros debajo de la diagonal; y L_k es una matriz triangular inferior unipotente.

En seguida, se puede aplicar las *operaciones de columna* del mismo estilo, que sería equivalente a una *posmultiplicación* $L_k A \mapsto L_k A L_k^{\mathsf{t}}$. El resultado intermedio es la matriz *simétrica* $\underline{A^{(k)} := L_k A L_k^{\mathsf{t}}}$, de este tipo:

$$A^{(k)} = L_k A L_k^{\mathsf{t}} = \begin{bmatrix} a_{11} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{(2)} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk}^{(k)} & * & \dots & * \\ 0 & 0 & \dots & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & * & * & \dots & * \end{bmatrix}.$$

Ahora, con $\boldsymbol{x}_k := L_k^{\mathsf{t}}\,\boldsymbol{e}_k$, se calcula que

$$a_{kk}^{(k)} = \langle \boldsymbol{e}_k, A^{(k)}\boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, L_k A L_k^{\mathsf{t}}\,\boldsymbol{e}_k \rangle = \langle L_k^{\mathsf{t}}\,\boldsymbol{e}_k, A L_k^{\mathsf{t}}\,\boldsymbol{e}_k \rangle = \langle \boldsymbol{x}_k, A\boldsymbol{x}_k \rangle > 0, \tag{5.17}$$

porque $\boldsymbol{x}_k \neq \boldsymbol{0}$. (De hecho, vale $\langle \boldsymbol{x}_k, \boldsymbol{e}_k \rangle = \langle L_k^{\mathsf{t}}\,\boldsymbol{e}_k, \boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, L_k\,\boldsymbol{e}_k \rangle = 1$.)

En el último paso, con $k = n$, se llega a la matriz diagonal

$$\underline{D := L_n A L_n^{\mathsf{t}}} = \operatorname{diag}\bigl[a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}\bigr] \quad \text{con cada} \quad a_{jj}^{(j)} > 0.$$

A poner $L := L_n^{-1}$, se obtiene $\underline{A = LDL^{\mathsf{t}}}$, como se quería.

La Prop. 5.37 es un resultado "si y solo si": siendo A simétrica e invertible, si $A = LDL^{\mathsf{t}}$ como la D diagonal y cada $d_{jj} > 0$, se puede introducir una "raíz cuadrada positiva" de D,

$$\underline{D^{1/2}} := \operatorname{diag}[\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}].$$

En tal caso se puede expresar la matriz A como

$$A = LDL^{\mathsf{t}} = LD^{1/2}D^{1/2}L^{\mathsf{t}} = C\,C^{\mathsf{t}}, \quad \text{al colocar} \quad \underline{C := LD^{1/2}}.$$

Pero se sabe que la matriz $\underline{A = C\,C^{\mathsf{t}}}$ es positiva (Ejemplo 5.34). Como A^{-1} existe, este A es definida positiva.

► La **factorización de Cholesky** de A es esta fórmula $A = C\,C^{\mathsf{t}}$, donde A es definida positiva y C es triangular inferior (pero no unipotente, en general).

El cálculo algorítmico de C es una variante de eliminación gaussiana simple, donde se "limpian" las filas y columnas alternadamente. Esto es: dado un pivote intermedio $a_{kk}^{(k)}$, se obtiene ceros en la columna k debajo del pivote por operaciones de fila; y en seguida se obtiene ceros en la fila k a la derecha del pivote por operaciones de columna.

Sobre la duodécima lista de ejercicios: "Tarea 12"

Ejercicios 5.14 y 5.15 Una matriz ortogonal conserva las longitudes de vectores. En el caso 2×2, una matriz ortogonal corresponde con una *rotación* por un ángulo θ o bien una *reflexión* en una recta.

Ejercicio 5.16 Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, la expresión $\underline{\boldsymbol{x}\,\boldsymbol{y}^{\mathsf{t}}}$ es el producto de una columna $(n \times 1)$ por una fila $(1 \times n)$ y como tal, es una matriz $n \times n$.

Ejercicio 5.19 Si A es una matriz definida positiva, la factorización de Cholesky ofrece una manera de expresar la "forma cuadrática" $\boldsymbol{x}^{\mathsf{t}} A \boldsymbol{x}$ como una suma de cuadrados.

Ejercicio 5.21 Un polinomio cuadrática en n variables es la suma de una forma cuadrática, una forma lineal y una constante. Si la forma cuadrática es positiva, el polinomio alcanza un mínimo absoluto, el cual se puede calcular sin usar derivadas parciales ni puntos críticos.

6 Comentarios sobre las clases de MA–360, parte 6

Este es un guión informal que acompaña las clases "virtuales" de MA–360, *Álgebra Lineal I*, en el primer semestre del 2021, a causa de la suspensión de clases presenciales.

6.1. Clase 24

Esta es la clase virtual del 5 de julio del 2021. La temática comprende la sección § 6.1: *Traza y determinante de una matriz cuadrada*.

En este capítulo final se vuelve a usar escalares de cualquier cuerpo $\mathbb{F}$, no solo los números reales y complejos.

El tema el capítulo es un para de funciones de los entradas de una matriz cuadrada (con valores en $\mathbb{F}$: la *traza* y el *determinante*. Parte del interés de estas funciones es que toman iguales valores sobre dos matrices semejantes, como se verá. La primera es una función lineal; la segunda es un polinomio en las entradas.

Trazas

La **traza** de una matriz cuadrada $C \in M_n(\mathbb{F})$ ya fue mencionada en el capítulo 5. Su definición es

$$\underline{\operatorname{tr} C} := c_{11} + c_{22} + \cdots + c_{nn} = \sum_{k=1}^{n} c_{kk} \in \mathbb{F}. \tag{6.1}$$

Las expresiones: $\operatorname{tr}(A^{\mathsf{t}}B)$ en el caso $\mathbb{F} = \mathbb{R}$, y $\operatorname{tr}(A^*B)$ en el caso $\mathbb{F} = \mathbb{C}$, definen productos escalares sobre $M_n(\mathbb{R})$ y $M_n(\mathbb{C})$, respectivamente. Para otros cuerpos $\mathbb{F}$, interesan las trazas de productos, tales como $\operatorname{tr}(AB)$ y $\operatorname{tr}(ABC)$.

La propiedad esencial de la traza (Lema 6.2) es la relación:

$$\underline{\operatorname{tr}(AB) = \operatorname{tr}(BA)}. \tag{6.2a}$$

Las matrices AB y BA deben ser cuadradas, pero las propias A y B pueden ser rectangulares: por ejemplo, si A es $m \times n$ y B es $n \times m$. La prueba de (6.2a) es inmediata:

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \operatorname{tr}(BA).$$

Dadas tres matrices A, B, C de dimensiones apropiadas, se deduce otra fórmula:

$$\underline{\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)}, \tag{6.2b}$$

al aplicar (6.2a) a las parejas $\{A, BC\}$ y $\{B, CA\}$. Esta es la *ciclicidad* de la traza.

Supóngase que dos matrices $A, B \in M_n(\mathbb{F})$ son *semejantes*: esto es, $\underline{B = P^{-1}AP}$ para alguna matriz invertible P. Entonces, por la ciclicidad:

$$\operatorname{tr} B = \operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr} A.$$

Esto es, $\underline{\operatorname{tr} A = \operatorname{tr} B}$: *dos matrices semejantes tienen la misma traza*.

► Este resultado es especialmente importante cuando se considera P como *una matriz de cambio de base*.

Si $T : V \to V$ un operador lineal, con $\dim V = n$, tómese dos bases de V, tales como $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$. La fórmula (4.10) define P por

$$\boldsymbol{x}'_s = \sum_{j=1}^{n} p_{js}\, \boldsymbol{x}_j\,.$$

El operador lineal T tiene dos matrices, A y B, con respecto a estas dos bases, por la "fórmula mágica" (4.2):

$$T(\boldsymbol{x}_j) =: \sum_{i=1}^{n} a_{ij}\, \boldsymbol{x}_i\,, \qquad T(\boldsymbol{x}'_s) =: \sum_{r=1}^{n} b_{rs}\, \boldsymbol{x}'_r\,.$$

Sabemos (por la sección 4.5) que $B = P^{-1}AP$ en ese caso. Definimos

$$\underline{\operatorname{tr} T := \operatorname{tr} A} = \operatorname{tr} B.$$

Conclusión: si se define la **traza de** T como la traza de su matriz, con respecto de *cualquier* base $\mathcal{B}$ de V, se ve que $\underline{\operatorname{tr} T}$ está bien definida. Entonces la traza define una forma lineal $\underline{\operatorname{tr} : M_n(\mathbb{F}) \to \mathbb{F}}$ y a la vez una forma lineal $\underline{\operatorname{tr} : \mathcal{L}(V) \to \mathbb{F}}$.

Determinantes

Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Antes de dar una definición formal de su determinante $\underline{\det A} \in \mathbb{F}$, es oportuno examinar los casos $n = 1, 2, 3$.

El caso $n = 1$ es fácil y trivial; $\det [a_{11}] := a_{11}$.

Se define un determinante 2×2 así:

$$\det A \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} := \underline{ad - bc}. \tag{6.3}$$

En general, las barras verticales (en lugar de corchetes o paréntesis) denotan determinantes. La fórmula anterior también puede escribirse así:

$$\det A \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{12}a_{21}.$$

En el capítulo 3, se encontró la cantidad $(ad - bc)$ en la fórmula para el inverso de una matriz 2×2:

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Se sobrentiende que cuando $ad - bc = 0$, la matriz A no es invertible.

► Un determinante 3×3 se define – provisionalmente – mediante una *expansión en la primera fila*:

$$\det A := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} := a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

y al expandir cada "determinante menor" 2×2, se llega a

$$\det A := \underline{a_{11}a_{22}a_{23} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}}\,. \tag{6.4}$$

Este es un polinomio (homogéneo) de grado 3 en las 9 entradas de la matriz A.

También se puede efectuar una *expansión en la primera columna*:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} := a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

que también coincide con (6.4)

Un tercer procedimiento obtiene (6.4) de modo más directo:

$$\det A = \left|\begin{matrix} \searrow^{+} & \searrow^{+} & \searrow^{+} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ \nearrow_{-} & \nearrow_{-} & \nearrow_{-} \end{matrix}\right| \begin{matrix} \\ a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ \\ \end{matrix} \tag{6.5}$$

Consiste en replicar las primeras dos columnas a la derecha (como fantasmas) y luego calcular los seis productos en (6.4) de las "subdiagonales" de este arreglo 3×5: los que "bajan" con signo $+$ y los que "suben" con signo $-$ (siguiendo las flechas).

El Ejemplo 6.7 aplica los tres procedimientos a una típica matriz 3×3:

$$
\text{(a)} \qquad \begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} = -3 \begin{vmatrix} 3 & -4 \\ -2 & 5 \end{vmatrix} - 1 \begin{vmatrix} -1 & -4 \\ 1 & 5 \end{vmatrix} + 2 \begin{vmatrix} -1 & 3 \\ 1 & -2 \end{vmatrix}
$$
$$
\begin{aligned} &= -3(15-8) - 1(-5+4) + 2(2-3) \\ &= -21 + 1 - 2 = -22. \end{aligned}
$$

$$
\text{(b)} \qquad \begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} = -3 \begin{vmatrix} 3 & -4 \\ -2 & 5 \end{vmatrix} - (-1) \begin{vmatrix} 1 & 2 \\ -2 & 5 \end{vmatrix} + 1 \begin{vmatrix} 1 & 2 \\ 3 & -4 \end{vmatrix}
$$
$$
\begin{aligned} &= -3(15-8) + 1(5+4) + 1(-4-6) \\ &= -21 + 9 - 10 = -22. \end{aligned}
$$

$$
\text{(c)} \qquad \begin{vmatrix} -3 & 1 & 2 \\ -1 & 3 & -4 \\ 1 & -2 & 5 \end{vmatrix} = \left| \begin{array}{ccc|cc} -3 & 1 & 2 & -3 & 1 \\ -1 & 3 & -4 & -1 & 3 \\ 1 & -2 & 5 & 1 & -2 \end{array} \right.
$$
$$
\begin{aligned} &= (-45 - 4 + 4) - (6 - 24 - 5) \\ &= -45 - (-23) = -22. \end{aligned}
$$

Ahora se puede definir determinantes de matrices $n \times n$ *por inducción sobre* n. En preparación para eso, se define la **submatriz** $\underline{A_{ij}}$ de A (con A mayúscula) como la matriz $(n-1) \times (n-1)$ obtenido de A, *al borrar su fila # i y su columna # j*. (Hay un total de n^2 submatrices de ese tipo.)

Se define el **menor** $\underline{M_{ij}}$ de A como su determinante: $\underline{M_{ij} := \det A_{ij}}$.

Las primeras dos fórmulas para un determinante 3×3 entonces son:

$$
\begin{aligned} \det A &= a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} \\ &= a_{11}M_{11} - a_{21}M_{21} + a_{31}M_{31}\,. \end{aligned}
$$

La Defn. 6.9 ofrece varias recetas inductivas para calcular $\det A$ para expansiones recursivas. Una de ellas (expansión en la primera fila) es:

$$
\det A := \underline{a_{11}M_{11} - a_{12}M_{12} + \cdots + (-1)^{1+n}a_{1n}M_{1n}} = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} M_{1j} \tag{6.6a}
$$

y en cada caso se trata de una *suma alternada* de términos $a_{ij}M_{ij}$.

6.2. Clase 25

Esta es la clase virtual del 8 de julio del 2021. La temática comprende las secciones § 6.1: *Traza y determinante de una matriz cuadrada*, y § 6.2: *Propiedades de determinantes*.

La fórmula de Leibniz para un determinante

La Defn 6.9 expone varias fórmulas para calcular determinantes, que se pueden resumir así:

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \qquad (6.6\text{b,c})$$

donde la primera sumatoria (con i fijo) corresponde a una *expansión en la fila F_i* de A; mientras la segunda sumatoria (con j fijo) corresponde a una *expansión en su columna C_j*.

De hecho, estas expresiones constituyen un total de $2n$ fórmulas diferentes. ¿Por qué, entonces, dan el mismo resultado?

Resulta que cada una de estas expansiones, al componerlas con expansiones secundarias en todos los menores M_{ij}, conducen a *la misma expansión completa*. Esa expansión completa se escribe así:

$$\det A = \sum_{\sigma} (-1)^{\sigma} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \qquad (6.7)$$

En esta **fórmula de Leibniz** la sumatoria recorre las *permutaciones* $\underline{\sigma = (j_1, \dots, j_n)}$ de $(1, 2, \dots, n)$. Cada término es un producto de entradas a_{ij} de A, tomadas de filas diferentes y de columnas diferentes de todas las maneras posibles. Antes de sumar esos productos, se les asigna un *signo* $\underline{(-1)^{\sigma} = \pm 1}$, que vale $+1$ si σ es par, y -1 si σ es impar.

(La fórmula fue explicada verbalmente por Leibniz en su carta de 1693 a l'Hôpital – ese es el epígrafe del capítulo 6 – donde a la vez introdujo la idea de nombrar los coeficientes de un sistema de ecuaciones lineales con un par de índices (i, j) en vez de agotar el abecedario.)

Nota: Una permutación σ se llama *par* [respectivamente, *impar*] según sea igual al producto de un número par [resp. impar] de *transposiciones* $r \leftrightarrow s$. Además, si τ es otra permutación y si $\sigma\tau$ es la permutación compuesta, resulta que $\underline{(-1)^{\sigma\tau} = (-1)^{\sigma}(-1)^{\tau}}$.

Para comprobar la fórmula (6.7), se hace la expansión de $\det A$ en su primera fila, $\det A = \sum_{j_1=1}^{n} \pm a_{1j_1} M_{1j_1}$; luego se expande cada menor M_{1j} en la primera fila de la submatriz A_{1j}, y así sucesivamente.

Por ejemplo, las primeras dos expansiones dan

$$\det A = \sum_{j_1,j_2=1}^{n} \pm a_{1j_1} a_{2j_2} M_{12,j_1 j_2}\,.$$

Solo es necesario comprobar que los signos ± 1 en la expansión completa son los $(-1)^{\sigma}$.

► Es útil observar que los signos $(-1)^{i+j}$ en las expansiones (6.6) siguen el patrón del *tablero de damas (o de ajedrez)*:

$$\begin{vmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{vmatrix} \tag{6.8}$$

con signos + en la diagonal.

(Discusión del Ejemplo 6.11.)

Propiedades generales de determinantes

La propiedad clave de determinantes (Prop. 6.12) es:

$$\det(AB) = (\det A)(\det B). \tag{6.9}$$

La entrada c_{kj} de $C = AB$ es la suma $\sum_{i=1}^{n} a_{ki} b_{ij}$. Se expande $\det C$ con la fórmula de Leibniz; y se usa la multiplicatividad $(-1)^{\sigma\tau} = (-1)^{\sigma}(-1)^{\tau}$ para obtener los signos de los factores $\det A$ y $\det B$ en sus propias fórmulas de Leibniz.

► La Prop. 6.13 muestra cómo se comportan determinantes bajo operaciones de fila:

(a) al intercambiar dos filas de A, $\det A$ *cambia de signo*;

(b) al multiplicar una fila de A por $c \neq 0$, *se multiplica* $\det A$ *también* por c;

(c) al sustraer de una fila de A un múltiplo de otra fila, $\det A$ *permanece igual*.

Al recordar que estas operaciones de fila son *premultiplicaciones* por matrices de tipo (3.12): $\underline{A \mapsto P_{ik}\,A}$; $\underline{A \mapsto M_i(c)\,A}$; $\underline{A \mapsto R_{ik}(c)\,A}$; y al usar (6.9), solo es necesario chequear que:

$$\det P_{ik} = -1, \qquad \det M_i(c) = c, \qquad \det R_{ik}(c) = 1.$$

En cada una de estas matrices, se calcula su determinante por expansión en filas $F = (0, 0, \dots, 1, \dots, 0, 0)$ con el 1 en la diagonal; y de esta forma se reduce al caso 2×2:

$$\det P_{ik} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1, \qquad \det M_i(c) = \begin{vmatrix} 1 & 0 \\ 0 & c \end{vmatrix} = c, \qquad \det R_{ik}(c) = \begin{vmatrix} 1 & 0 \\ -c & 1 \end{vmatrix} = 1.$$

► La Prop. 6.14 dice que la transpuesta A^{t} tiene el mismo determinante que A:

$$\underline{\det A^{\mathsf{t}} = \det A}.$$

Esto también viene de la fórmula de Leibniz, porque

$$\det A^{\mathsf{t}} = \sum_{\sigma \in S_n} (-1)^\sigma a_{j_1 1} a_{j_2 2} \dots a_{j_n n};$$

Los términos de esta sumatoria son productos de entradas a_{ji}, una de cada columna y una de cada columna. En el lado derecho, se nota que $\pi \colon (j_1, \dots, j_n) \mapsto (1, \dots, n)$ es la inversa de $\sigma \colon (1, \dots, n) \mapsto (j_1, \dots, j_n)$. Al escribir π como $\pi \colon (1, \dots, n) \mapsto (p_1, \dots, p_n)$ y al notar que $(-1)^\pi = (-1)^\sigma$, se reorganiza la sumatoria anterior como

$$\det A^{\mathsf{t}} = \sum_{\pi} (-1)^\pi a_{1p_1} a_{2p_2} \cdots a_{np_n} = \det A.$$

Sobre la décima tercera lista de ejercicios: "Tarea 13"

Ejercicios 6.1 La propiedad $\tau(AB) = \tau(BA)$ casi caracteriza la traza: si τ es un formal lineal que la cumple, entonces τ debe ser proporcional a la traza $\underline{\mathrm{tr}}$.

Ejercicio 6.2 En 1925, Werner Heisenberg y Max Born se dieron cuenta que ciertas "observables" Q, P de la (incipiente) mecánica cuántica *no conmutan*, sino que cumplen la relación: $QP - PQ = i\hbar\,1$. Al principio, Born pensaba que Q, P deben ser *matrices* $n \times n$ para algún n finita. Pero eso no es posible: para modelarlos, es necesario usar "matrices infinitas", o sea, operadores sobre un espacio de Hilbert infinitodimensional.

Ejercicios 6.8 a 6.10 Algunas determinantes se calculan fácilmente mediante expansiones en filas o columnas. Pero otras se resuelven más fácilmente por eliminación gaussiana, es decir, al aplicarles ciertas operaciones de fila. La factorización del *determinante de Vandermonde* (Ejercicio 6.9) es una buena ilustración de ese fenómeno.

Ejercicio 6.12 Una nueva manera de comprobar independencia lineal de varios vectores en $\mathbb{R}^n$ es verificar que su *determinante de Gram* no es 0. Pero: ¿por qué funciona esa regla?

6.3. Clase 26

Esta es la clase virtual del 12 de julio del 2021. La temática comprende la segunda parte de la sección § 6.2: *Propiedades de determinantes*.

Otras propiedades de determinantes

En la clase anterior, se ha visto dos propiedades algebraicas de determinantes de matrices cuadradas:

$$\begin{aligned} \det(AB) &= (\det A)(\det B), \\ \det A^{\mathsf{t}} &= \det A. \end{aligned}$$

Además, se notó que las operaciones de fila, aplicadas a una matriz A, tienen estos efectos sobre su determinante:

(a) $F_i \leftrightarrow F_k$: *cambio de signo*, $\det A \mapsto -\det A$;

(b) $F_i \mapsto c\,F_i$: *multiplicar por* c, $\det A \mapsto c(\det A)$;

(c) $F_i \mapsto F_i - c\,F_k$: *no hay cambio*, $\det A \mapsto \det A$.

► Ahora usaremos estas propiedades – en lugar de la definición misma del determinante por la fórmula de Leibniz – para sacar algunas otra consecuencias.

La Prop. 6.15 menciona una circunstancia particularmente útil para el cálculo de determinantes: *si* A *es* ***triangular***, *entonces*

$$\det A = a_{11}a_{22}\cdots a_{nn} \tag{$*$}$$

es el producto de las entradas en la diagonal.

Si A es triangular superior, entonces A^{t} es triangular inferior, con $\det A^{\mathsf{t}} = \det A$; nótese que A y A^{t} tienen las mismas entradas diagonales. Entonces, para verificar esta fórmula, podemos suponer que A es *triangular inferior*.

La expansión en la primera fila tiene un solo término en cada paso; por ejemplo:

$$\begin{vmatrix} {}^{(+)}a_{11} & {}^{(-)}0 & {}^{(+)}0 & {}^{(-)}0 \\ * & a_{22} & 0 & 0 \\ * & * & a_{33} & 0 \\ * & * & * & a_{44} \end{vmatrix} = a_{11} \begin{vmatrix} {}^{(+)}a_{22} & {}^{(-)}0 & {}^{(-)}0 \\ * & a_{33} & 0 \\ * & * & a_{44} \end{vmatrix} = a_{11}a_{22} \begin{vmatrix} {}^{(+)}a_{33} & {}^{(-)}0 \\ * & a_{44} \end{vmatrix} = a_{11}\,a_{22}\,a_{33}\,a_{44}\,.$$

Cualquier *matriz diagonal* es triangular (de ambos tipos). La fórmula $(*)$ entonces también calcula determinantes de matrices diagonales. En particular, la matriz identidad (de cualquier tamaño) tiene determinante uno: $\underline{\det 1_n = 1}$.

► Hay dos circunstancias (Lema 6.16) que implican $\underline{\det A = 0}$:

(i) Si tiene una fila o una columna de ceros. (Por la expansión en esa fila o columna).

(ii) Si tiene *dos filas iguales*, o *dos columnas iguales*. (Porque entonces resulta que $\det A = -\det A$, o bien $\det A^{\mathsf{t}} = -\det A^{\mathsf{t}}$, lo que es lo mismo.)

► El determinante recibe su nombre porque *determina* si una matriz es invertible o no. La Prop. 6.17 dice que: $\det A = 0$ *si y solo si* A^{-1} *no existe* (se dice que A es **singular**).

Para calcular $\det A$, se puede *intentar reducir* A *a una matriz triangular mediante eliminación gaussiana* – con pivoteo parcial: o sea, se admiten intercambios de filas si fueran necesarios. (Las operaciones de fila de tipo (b) no son necesarios y por ahora es mejor evitarlos.)

Se sabe que A es invertible $\iff$ el sistema de ecuaciones $A\boldsymbol{x} = \mathbf{0}$ tiene solución única, $\iff$ la eliminación termina exitosamente con n pivotes no ceros.

En el caso contrario, se obtiene solo $k < n$ pivotes no ceros y el algoritmo se detiene en un paso intermedio, en donde $A \mapsto A'$ donde:

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

donde U' es *triangular superior* $k \times k$. Sus elementos diagonales son $a_{11}, a_{22}^{(2)}, \dots, a_{kk}^{(k)}$.

La matriz A' cumple $\underline{\det A' = \pm \det A}$ (se ha usado operaciones de tipos (a) y (c) solamente). En este A, 0 es un bloque rectangular de ceros; y *la primera columna de* Y' *es también una columna de ceros* porque no hay más pivotes disponibles. Esto implica que $\underline{\det Y' = 0}$; y luego $\underline{\det A' = 0}$ – por el Ejercicio 6.10, o bien por expansiones en columnas:

$$\begin{aligned} \det A' &= a_{11} M_{11} = a_{11} a_{22}^{(2)} M_{12,12} = \cdots \\ &= a_{11}\, a_{22}^{(2)} \cdots a_{kk}^{(k)} \det Y' = 0. \end{aligned}$$

Ahora supóngase que la eliminación termina con éxito, así que A^{-1} existe. Para verificar que $\det A \neq 0$, se puede *calcular* $\det A$ con la misma eliminación. Si hubo unos r intercambios de filas, premultiplicando por matrices $P_{i_1 k_1}, P_{i_2 k_2}, \dots, P_{i_r k_r}$, *se guarda un factor* $\underline{(-1)^r} = \pm 1$ *para el cálculo final* de $\det A$. Sea $P := P_{i_r k_r} \cdots P_{i_1 k_1}$ (el producto de las $\overline{P_{ik}}$ en el orden inverso). Resulta que *la matriz* PA *admite eliminación gaussiana simple*, $\underline{PA = LV}$, con L triangular unipotente y V triangular. (En palabras: se vuelve a correr el algoritmo, haciendo todos los cambios de filas al inicio, $A \mapsto PA$.)

El determinante de A se obtiene de las relaciones:

$$\underline{(-1)^r \det A} = (\det P)(\det A) = \det(PA) = \det(LV) = (\det L)(\det V) = \underline{\det V},$$

porque $\underline{\det L = 1}$ ya que L es triangular con entradas diagonales iguales a 1. Y las entradas diagonales de V son los pivotes obtenidos. La Prop. 6.18 enuncia la fórmula final:

$$\boxed{\det A = (-1)^r \, a_{11} \, a_{22}^{(2)} \cdots a_{nn}^{(n)}.} \tag{6.11}$$

Cálculo del rango de una matriz, usando determinantes

Una matriz rectangular $m \times n$ no tiene determinante si $m \neq n$. Sin embargo, posee un surtido de *submatrices cuadradas* $k \times k$, donde $k \leqslant \min\{m, n\}$.

La Prop. 6.19 ofrece una nueva manera de *calcular el rango* $\overline{r(A)}$ de una matriz $A \in M_{m \times n}(\mathbb{F})$. Este es *el mayor entero* $\underline{k} \in \mathbb{N}$ tal que *haya una* $\overline{\textit{submatriz}}$ B de A, de tamaño $k \times k$ con $\underline{\det B \neq 0}$.

Para comprobar esto, tomamos $k := \overline{r(A)}$. Se debe mostrar que A tiene una submatriz $k \times k$, llamémosla B, tal que $\underline{\det B \neq 0}$; y si M es una submatriz de A de tamaño $(k+1) \times (k+1)$, entonces $\underline{\det M = 0}$.

Para hallar B, tómese k columnas linealmente independientes de A (por ejemplo, aquellas que se vuelven columnas básicas al reducir A a su forma escalonada). Al borrar las otras columnas, se obtiene una submatriz C de A, de tamaño $m \times k$. Tómese k *filas* linealmente independientes de C – por ejemplo, las que se vuelven columnas básicas en la forma escalonada de C^{t}. Al borrar las demás filas, queda una submatriz B, que es $k \times k$ y que tiene k filas linealmente independientes. Entonces $r(B) = k$, el rango máximo de una matriz $k \times k$, así que B es invertible, con $\det B \neq 0$.

Si M es una submatriz de A de tamaño $(k+1) \times (k+1)$, las columnas de M son porciones de unas $k+1$ columnas de A, que son linealmente *dependientes*. Una de ellas se puede expresar como combinación lineal de las otras. La misma relación entre columnas es aplicable a la submatriz M, al borrar las filas de A que sobran. Esto dice que $r(M) < k+1$; que M no es invertible; y que $\det M = 0$.

6.4. Clase 27

Esta es la clase virtual del 15 de julio del 2021. La temática comprende la sección § 6.3: *La regla de Cramer*.

Cofactores y la matriz adjugada

Los sistemas lineales con n ecuaciones para n variables se resuelven usualmente por eliminación gaussiana, si $n \geqslant 4$. Cuando $n = 2$ o $n = 3$, un método alternativo es el de aplicar *una fórmula explícita* para el vector (columna) de soluciones – si existe. De hecho, esa fórmula explícita está disponible para cualquier n, pero requiere unos cálculos extensos.

Aún así, la fórmula explícita tiene importancia teórica y conviene verla para cualquier n. Fue propuesta por Gabriel Cramer en 1750 (para sistemas 3×3) y se conoce como la "regla de Cramer".

► El ingrediente esencial para obtener la regla de Cramer para una matriz cuadrada A es su **matriz adjugada**, o *matriz de cofactores*, denotada $\underline{\operatorname{adj} A}$. El **cofactor** de a_{ij} (una entrada de A) es el escalar

$$(-1)^{i+j} M_{ji}$$

donde M_{ji} es el menor de la entrada a_{ji} de A (nótese esos índices transpuestos). La matriz $\underline{\operatorname{adj} A}$ tiene este cofactor en la posición (i, j). Se obtiene este matriz en tres pasos:

(i) reemplazando cada elemento a_{ij} de A por el *menor* M_{ij} correspondiente;

(ii) multiplicando cada entrada por el *signo* $(-1)^{i+j} = \pm 1$ del tablero de damas (6.8);

(iii) tomando la *transpuesta* de la matriz resultante.

Una matriz 2×2 tiene la siguiente matriz adjugada:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \overset{\text{(i)}}{\longmapsto} \begin{bmatrix} d & c \\ b & a \end{bmatrix} \overset{\text{(ii)}}{\longmapsto} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \overset{\text{(iii)}}{\longmapsto} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Cabe recordar la fórmula (3.5) para invertir una matriz 2×2:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad \text{si} \quad ad - bc \neq 0.$$

Ahora se puede generalizar esta fórmula así:

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A \qquad \text{si} \quad \det A \neq 0. \tag{6.15}$$

En breve se justificará esta fórmula para la matriz inversa.

Si A es una matriz 3×3, su matriz adjugada es también 3×3. Por ejemplo,

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & -4 & -2 \end{bmatrix} \xmapsto{\text{(i)}} \begin{bmatrix} -4 & 0 & 0 \\ 2 & -2 & -4 \\ -2 & 0 & 2 \end{bmatrix} \xmapsto{\text{(ii)}} \begin{bmatrix} -4 & 0 & 0 \\ -2 & -2 & 4 \\ -2 & 0 & 2 \end{bmatrix} \xmapsto{\text{(iii)}} \begin{bmatrix} -4 & -2 & -2 \\ 0 & -2 & 0 \\ 0 & 4 & 2 \end{bmatrix}.$$

El determinante de la primera matriz es

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

Ahora es fácil calcular la productos de A y $\operatorname{adj} A$ (en los dos órdenes):

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

► Esta es una instancia de la fórmula general:

$$A\,(\operatorname{adj} A) = (\operatorname{adj} A)\,A = (\det\, A)\,1_n. \tag{6.12}$$

Al escribir $\underline{B := \operatorname{adj} A}$, esto es $\underline{AB = BA = (\det\, A)\,1_n}$.

La Prop. 6.22 demuestra esta fórmula general (6.12). Para hacerlo, cabe recordar el cálculo de $\underline{\det\, A}$ por expansión en la fila $\boldsymbol{a}'_i$ o bien en la columna $\boldsymbol{a}_j$:

$$\det\, A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \tag{6.6b,c}$$

Por la definición de los cofactores $b_{ij} := (-1)^{i+j} M_{ji}$, esto es lo mismo que

$$\det\, A = \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{i=1}^{n} b_{ji} a_{ij} \tag{6.13}$$

o bien, como productos punto de filas por columnas:

$$\det\, A = \boldsymbol{a}'_i \cdot \boldsymbol{b}_i = \boldsymbol{b}'_j \cdot \boldsymbol{a}_j\,.$$

En breve: las entradas diagonales de los AB y BA son todas iguales a $\underline{\det\, A}$.

Para las entradas no diagonales de AB, fíjese que si $i \neq k$, vale

$$\boldsymbol{a}'_k \cdot \boldsymbol{b}_i = \sum_{j=1}^{n} a_{kj} b_{ji} = \sum_{j=1}^{n} (-1)^{i+j} a_{kj} M_{ij} \,. \tag{6.14}$$

La segunda sumatoria es la expansión en la fila #i de $\underline{\det A'}$, donde $A \mapsto A'$ se obtiene al reemplazar la fila $\boldsymbol{a}'_i$ por la fila $\boldsymbol{a}'_k$. Pero entonces A' es una matriz con dos filas iguales y por lo tanto $\underline{\det A' = 0}$. O sea, vale $\boldsymbol{a}'_k \cdot \boldsymbol{b}_i = 0$ si $i \neq k$. Dicho de manera más sencilla: *la matriz AB es diagonal*.

Del mismo modo se obtiene $\boldsymbol{b}'_j \cdot \boldsymbol{a}_l = 0$ cuando $j \neq l$, así que BA es también diagonal. Se ha mostrado que tanto AB como BA es un múltiplo de la matriz identidad 1_n, por el factor $\underline{\det A}$.

La regla de Cramer

Volvamos al planteamiento del sistema de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{b}$ (con n ecuaciones para n variables, o sea: $A \in M_n(\mathbb{F})$, $\boldsymbol{b} \in \mathbb{F}^n$, $\boldsymbol{x} \in \mathbb{F}^n$. Este sistema tiene *solución única* si y solo si A es invertible, si y solo si $\det A \neq 0$. (La solución única es $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.)

Supongamos que $\det A \neq 0$. La fórmula $(\operatorname{adj} A)\, A = (\det A)\, 1_n$ implica que

$$(\det A)\, \boldsymbol{x} = (\operatorname{adj} A)\, A\, \boldsymbol{x} = (\operatorname{adj} A)\, \boldsymbol{b}.$$

Esta es una igualdad entre dos vectores de columna; sus coordenadas j son:

$$(\det A)\, x_j = (\text{fila } \#j \text{ de } \operatorname{adj} A)\, \boldsymbol{b} = \sum_{i=1}^{n} (-1)^{i+j} M_{ij}\, b_i = \det B_j$$

para cierta matriz B_j. La sumatoria es una expansión de $\overline{\det B_j}$ en la columna #j de B_j, si esa columna es una copia de $\boldsymbol{b}$ (y sus otras columnas son las de A). En resumen: *defínase* $\underline{B_j := [\boldsymbol{a}_1 \cdots \boldsymbol{a}_{j-1}\, \boldsymbol{b}\, \boldsymbol{a}_{j+1} \cdots \boldsymbol{a}_n]}$. De ahí (Prop. 6.24) se sigue la **regla de Cramer**:

$$x_j = \frac{\det B_j}{\det A} \quad \text{para} \quad j = 1, 2, \dots, n. \tag{6.16}$$

El caso 2×2 de la regla de Cramer es el siguiente:

$$\begin{Bmatrix} ax_1 + bx_2 = p \\ cx_1 + dx_2 = q \end{Bmatrix} \implies x_1 = \frac{pd - bq}{ad - bc}\,, \quad x_2 = \frac{aq - pc}{ad - bc}\,.$$

Sobre la décima cuarta lista de ejercicios: "Tarea 14"

Ejercicios 6.17 Algunas fórmulas de la geometría analítica del plano $\mathbb{R}^2$ o del espacio $\mathbb{R}^3$ pueden expresarse con determinantes, aprovechando las propiedades algebraicas de estos. Como muestra, se ofrece la ecuación del círculo que pasa por tres puntos dados en el plano.

Ejercicio 6.18 El determinante de la matriz adjugada $\underline{\operatorname{adj} A}$ es una potencia de $\underline{\det A}$. La matriz adjugada de la matriz adjugada de A es un múltiplo de la matriz original.

Ejercicios 6.20 y 6.21 Se presenta una matriz simétrica 8×8 que depende de dos parámetros reales, b y c. Esta matriz viene de una tesis de licenciatura, en donde era necesario obtener condiciones necesarias y suficientes sobre b y c para que la matriz sea definida positiva. Esto se puede hacer con métodos del capítulo 5 (Ejercicio 6.20) y bien por el cálculo de ciertos menores (Ejercicio 6.21).

El Ejercicio 6.21 aprovecha un par de *criterios* (no demostrados en el curso): (i) una matriz simétrica real A es *positiva* (semidefinida) si todos sus *menores principales* son no negativos; (ii) A es *definida positiva* si sus *menores principales delanteros* son positivos. El segundo es un criterio bastante útil en varias aplicaciones.

6.5. Clase 28

Esta es la clase virtual del 19 de julio del 2021. La temática comprende la sección §6.4: *Autovalores y autovectores*.

Autovalores de una matriz cuadrada

Si $T\colon V \to V$ es un operador lineal, puede ocurrir que T lleva un vector $\boldsymbol{x} \in V$ a un vector $T(\boldsymbol{x})$ paralelo (esto es, proporcional) a xx:

$$\underline{T(\boldsymbol{x}) = \lambda\,\boldsymbol{x}} \quad \text{con} \quad \boldsymbol{x} \neq \mathbf{0}. \tag{6.17a}$$

(La solución $\boldsymbol{x} = \mathbf{0}$ sería trivial: cualquier operador lineal T lleva $\mathbf{0}$ en $\mathbf{0}$, y $\lambda\mathbf{0} = \mathbf{0}$ es cierto para cualquier λ. Por lo tanto, *se excluye el vector nulo*.)

Cuando esta relación (6.17a) se cumple, se dice que $\underline{\lambda \in \mathbb{F}}$ es un **autovalor** o *valor propio* de T y que $\underline{\boldsymbol{x} \in V}$, $\boldsymbol{x} \neq \mathbf{0}$, es un **autovector** o *vector propio* de T asociada a λ.

El caso de interés especial es $V = \mathbb{F}^n$ y $T = T_A$, donde A es una matriz $n \times n$ con entradas en $\mathbb{F}$. Entonces se propone la ecuación:

$$\underline{A\boldsymbol{x} = \lambda\,\boldsymbol{x}} \quad \text{para algún} \quad \boldsymbol{x} \neq \mathbf{0} \text{ en } \mathbb{F}^n. \tag{6.17b}$$

Resulta que esta es un sistema *homogéneo* de ecuaciones lineales:

$$\underline{(\lambda\, 1_n - A)\boldsymbol{x} = \mathbf{0}}$$

y se busca una solución no trivial $\boldsymbol{x} \neq \mathbf{0}$. Esto ocurre solo si el sistema *no tiene solución única*. Ya se sabe que esto depende de la naturaleza de la matriz $(\lambda\, 1_n - A)$.

Concretamente, λ es un autovalor de A si y solo si esta ecuación tiene una solución $\boldsymbol{x} \neq \mathbf{0}$, si y solo si $\underline{\ker(\lambda\, 1_n - T_A) \neq \{\mathbf{0}\}}$, si y solo si

$$\underline{\det(\lambda\, 1_n - A) = 0}. \tag{6.18}$$

► Para obtener información sobre los posibilidades para λ, se puede desarrollar este determinante con la fórmula de Leibniz. De los $n!$ sumandos, el mas importante es el primero:

$$\underline{(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})} + (\text{términos de grado menor que } n \text{ en } \lambda).$$

En síntesis: el lado izquierdo de (6.18) es "un polinomio en λ".

Este polinomio, que llamaremos $p_A(t)$, se llama el **polinomio característico** de A (en la Defn. 6.27) y es de grado n:

$$\underline{p_A(t) := \det(t\, 1_n - A)} = t^n + (\text{términos de grado menor que } n). \tag{6.19}$$

Este polinomio de grado n tiene a lo sumo n *raíces* – podría tener menos – al plantear la ecuación $\underline{p_A(t) = 0}$. Los autovalores de A son estas raíces $\underline{t = \lambda_1}$, $\underline{t = \lambda_2}, \dots, \underline{t = \lambda_k}$ con $k \leqslant n$.

► El Ejemplo 6.28 toma una matriz real típica 2×2:

$$A = \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix}.$$

Se plantea la ecuación

$$p_A(t) = \det(t\, 1_2 - A) = \begin{vmatrix} t+5 & -3 \\ 6 & t-4 \end{vmatrix} = 0,$$

que se simplifica en:

$$t^2 + t - 2 = 0, \quad \text{o bien} \quad (t-1)(t+2) = 0,$$

y las raíces son $\lambda = 1, -2$.

Para los autovectores, se debe resolver (para cada λ) un sistema homogéneo: $(1_2 - A)\boldsymbol{x} = \mathbf{0}$ y $(-2\,1_2 - A)\boldsymbol{y} = \mathbf{0}$:

$$\begin{Bmatrix} 6x_1 - 3x_2 = 0 \\ 6x_1 - 3x_2 = 0 \end{Bmatrix}, \qquad \begin{Bmatrix} 3y_1 - 3y_2 = 0 \\ 6y_1 - 6y_2 = 0 \end{Bmatrix}.$$

Nótese la ecuación redundante en cada caso. Las soluciones son:

$$\boldsymbol{x} = \begin{bmatrix} a \\ 2a \end{bmatrix} = a\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} b \\ b \end{bmatrix} = b\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Se recomienda un chequeo de que estos son autovectores de A:

$$\begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \qquad \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix} = -2\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

► Cuando $n > 2$, es menos fácil resolver la ecuación polinomial $p_A(t) = 0$, porque es necesario factorizar $p_A(t)$. Sin embargo, hay una circunstancia en donde eso es factible: cuando A es una *matriz triangular*. En tal caso, la matriz $(t\,1_n - A)$ es también triangular, y el determinante se calcula en seguida:

$$p_A(t) = \det(t\,1_n - A) = \begin{vmatrix} t - a_{11} & -a_{12} & \dots & -a_{1n} \\ 0 & t - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t - a_{nn} \end{vmatrix} = (t - a_{11})(t - a_{22}) \cdots (t - a_{nn})$$

si A es triangular superior. De ahí se ve que *los autovalores de una matriz triangular son sus elementos diagonales*: $\lambda \in \{a_{11}, a_{22}, \dots, a_{nn}\}$.

► Si dos matrices $n \times n$, A y B son semejantes: $B = P^{-1}AP$ para alguna P invertible, entonces

$$\det B = \det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A.$$

Se sigue que $\underline{\det A}$ es insensible a cambios de base de $\mathbb{F}^n$. Es recomendable – pero no siempre es posible – buscar una base que consiste de autovectores de A.

Para ver un caso factible, se puede retomar el ejemplo anterior:

$$A = \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} \quad \text{tiene autovectores} \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Los dos autovectores son linealmente independientes (pues no son proporcionales).

Entonces forman las columnas de una matriz invertible:

$$P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad \text{con inverso} \quad P^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}.$$

Ahora calculamos la matriz $B := P^{-1}AP$, así:

$$P^{-1}AP = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}.$$

Esta es una *matriz diagonal* B; sus entradas diagonales, 1 y -2, son los autovalores de A.

Esto sucede porque

$$t\,1_n - B = P^{-1}(t\,1_n - A)P$$

y al tomar determinantes, $p_B(t) = p_A(t)$. Por lo tanto, *dos matrices semejantes tienen los mismos autovalores*. Y cuando B es diagonal, estos son las entradas diagonales de B.

► Hay matrices *reales* sin autovalor alguno. El Ejemplo 6.32 estudia la matriz

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in M_2(\mathbb{R}).$$

Su polinomio característico es

$$p_J(t) := \begin{vmatrix} t & -1 \\ 1 & t \end{vmatrix} = t^2 + 1.$$

Y la ecuación $\underline{t^2 + 1 = 0}$ no tiene soluciones reales.

Pero si tiene soluciones *complejas*:

$$p_J(t) = t^2 + 1 = (t - i)(t + i),$$

así que los autovalores complejas de J son $\underline{\lambda_1 = i}$ y $\underline{\lambda_2 = -i}$.

Los autovectores son también complejos, en $\mathbb{C}^2$:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} i \\ -1 \end{bmatrix} = i \begin{bmatrix} 1 \\ i \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} -i \\ -1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Para *diagonalizar* J, se usa la matriz invertible

$$P = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \quad \text{con inverso} \quad P^{-1} = -\frac{1}{2i} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} = \frac{i}{2} \begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{1}{2} & \frac{i}{2} \end{bmatrix}.$$

Por lo tanto,

$$P^{-1}JP = \begin{bmatrix} \frac{1}{2} & -\frac{i}{2} \\ \frac{1}{2} & \frac{i}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} \frac{i}{2} & \frac{1}{2} \\ -\frac{i}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Conclusión: J es *diagonalizable en* $M_2(\mathbb{C})$ *pero no es diagonalizable en* $M_2(\mathbb{R})$.

Algo más sobre la "Tarea 14"

Ejercicios 6.23 Una matriz 3×3 puede tener 3 autovalores distintos $\lambda_1, \lambda_2, \lambda_3$. Para cada uno, se debe plantear y resolver un sistema de ecuaciones lineales homogéneas $(\lambda_j 1_3 - A)\boldsymbol{x}_j = \mathbf{0}$ para hallar un autovector correspondiente. Si esos autovectores son linealmente independientes, forman las columnas de una matriz invertible P tal que $P^{-1}AP$ es diagonal.

Ejercicio 6.25 Las dos matrices 2×2 de este ejercicio tiene efectos geométricos diferentes sobre el plano $\mathbb{R}^2$. La primera es una rotación que cambia la dirección de cualquier vector. La segunda es una distorsión que estira el plano en una dirección y lo encoge en otra. Estos fenómenos se reflejan en los autovalores y autovectores que poseen las dos matrices.