Elementos de Álgebra Lineal

Miguel Ángel Mota & Beatriz Rumbos

12 de mayo de 2017

Índice general

Capítulo 1

Preliminares

Conjuntos

Un **conjunto**, en su concepción más elemental, es una colección de objetos denominados los **elementos** del conjunto. Típicamente, los conjuntos suelen denotarse por letras mayúsculas: $A, B, C, \ldots$y sus elementos por letras minúsculas: $a, b, c, \ldots$ El símbolo "$\in$" denota la pertenencia de un elemento a un conjunto, por ejemplo

$$a \in A$$

significa que "a" es un elemento del conjunto A. La no pertenencia a un conjunto se denota por "$\notin$" de manera que

$$x \notin A$$

se lee como "x no es un elemento de A". Al conjunto sin elementos se le denomina **conjunto vacío** y se denota por $\emptyset$. Si dados dos conjuntos A y B se tiene que todos los elementos de A son elementos de B, entonces decimos que A es un subconjunto de B y esto se denota como

$$A \subset B.$$

En particular, $A \subset A$. Dado que el conjunto vacío no tiene elementos, por vacuidad, se tiene que $\emptyset \subset A$ para cualquier conjunto A.

Dos conjuntos A y B son iguales si ambos tienen los mismos elementos. En este caso se tiene que $A \subset B$ y $B \subset A$. La igualdad entre los conjuntos se denota simplemente por $A = B$. Si A es un subconjunto de B pero A no es igual a B (en este caso existirá al menos un elemento de B que no pertenece al conjunto A), decimos que A es un subconjunto propio de B (o está contenido propiamente en B). En este caso se utiliza la notación

$$A \subsetneq B.$$

Para describir un conjunto es necesario saber quienes son sus elementos. Para este efecto pueden listarse los elementos -esto es práctico si se trata de un conjunto pequeño o con cierta estructura recursiva- o bien describirlos con una o varias propiedades que los caracterizen. Podemos tener así

$$\begin{aligned} A &= \{a, b, 666, \#, \bigstar\}, \\ B &= \{1, 2, 3, \ldots\}, \\ C &= \{\text{anfibios de color verde}\}, \\ D &= \{x \in \mathbb{R} \mid -1 < x < 1\}. \end{aligned}$$

En el caso del conjunto D, la descripción de sus elementos se lee como: "el conjunto de todos los x que pertenecen a los números reales tales que cumplen $-1 < x < 1$". Sabemos que una forma compacta de expresar este conjunto es el intervalo abierto $(-1, 1)$ en la recta numérica[1].

Dados dos conjuntos A y B, podemos "juntar sus elementos" en un conjunto llamado la "unión de A y B", denotada por $A \cup B$ y definida como

$$A \cup B = \{x | x \in A \text{ o } x \in B\},$$

en donde el "o" indica que x puede ser un elemento de A, de B o de ambos. De forma análoga se define la intersección de los conjuntos A y B como el conjunto de sus elementos comunes y se denota por $A \cap B$. Formalmente,

$$A \cap B = \{x | x \in A \text{ y } x \in B\}.$$

Asimismo, también se define el **complemento relativo**[2] de B en A como

$$A \setminus B = \{x | x \in A \text{ y } x \notin B\}.$$

Así, $A \setminus B$ se refiere al conjunto de todos los elementos de A que no están en B.

Usualmente trabajamos con conjuntos cuyos elementos pertenecen a algún conjunto más grande (universal) que contiene a todos los conjuntos de interés. Si denotamos a este conjunto universal por X, para

[1] Es frecuente que la linea vertical "|" que significa "tales que cumplen" se sustituya por dos puntos ": ". Así, podríamos alternativamente denotar al conjunto C como

$$C = \{x \in \mathbb{R} \colon -1 < x < 1\}.$$

[2] En ocasiones llamada **diferencia de conjuntos** y denotada por $A - B$.

todo conjunto A tendremos que $A \subset X$. El complemento relativo de A con respecto a X se conoce simplemente como **complemento** y se denota por A^c, tenemos así que

$$A^c = X \setminus A = \{x \in X | x \notin A\}.$$

Ejemplo 1.1 *Para cualesquiera A, B conjuntos, se tiene que*

$$A \setminus B = A \cap B^c.$$

En efecto,

$$\begin{aligned} A \setminus B &= \{x | x \in A \ y \ x \notin B\} \\ &= \{x | x \in A \ y \ x \in B^c\} = A \cap B^c \end{aligned}$$

Ejemplo 1.2 *Supongamos que nuestro universo X es el conjunto de letras el alfabeto griego. Si $A = \{\alpha, \beta, \gamma, \gamma, \gamma\}$ y $B = \{\beta, \alpha, \gamma\}$, entonces $A = B$. Para demostrar esto simplemente observamos que dado cualquier $x \in A$ se cumple que $x \in B$, por lo que $A \subset B$, y dado cualquier $x \in B$ se cumple que $x \in A$, con lo cual $B \subset A$; por lo tanto, $A = B$. Observemos que es irrelevante el orden en el que aparecen los elementos, así como también si éstos aparecen o no repetidos.*

Ejemplo 1.3 *Sea X el conjunto de números naturales $\{1, 2, \ldots\}$. Sean $A = \{x \in X \mid x \text{ es impar}\}$, $B = \{x \in X \mid x \text{ es par}\}$, $C = \{6, 66, 666\}$ y $D = \{1, 2, 5, 6, 7\}$. Entre otras, se cumplen las siguientes:*

$$\begin{aligned} A &\subsetneq X, \\ A \cup B &= X, \\ A \cap B &= \varnothing \\ A^c &= B, \\ B^c &= A, \\ A \setminus B &= A, \\ C \cap D &= \{6\}, \\ C \setminus D &= \{66, 666\}. \end{aligned}$$

Conjuntos de números

Seguramente han utilizado los siguientes conjuntos de números:

$\mathbb{N} = \{1, 2, 3, \ldots\}$, conocidos como números Naturales.

$\mathbb{Z} = \{\ldots, -2 - 1, 0, 1, 2, \ldots\}$, conocidos como números Enteros.

$\mathbb{Q} = \left\{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\right\}$, conocidos como números Racionales.

En todos estos conjuntos pueden definirse las operaciones de suma $(+)$ y producto $(\times)$ entre cada pareja de elementos, mismas que satisfacen una serie de propiedades, junto con los elementos 0 y 1 que son elementos "neutros” o "idénticos", para la suma y el producto, respectivamente. Sin entrar en detalle, nos conformaremos con saber que podemos realizar estas operaciones aritméticas correctamente. El álgebra abstracta estudia la estructura algebraica de éstos y otros muchos conjuntos. En particular, los llamados espacios vectoriales son conjuntos que poseen lo que se conoce como una estructura *lineal.* El álgebra lineal estudia a estos espacios y a las funciones -que preservan esta estructura lineal- entre ellos. En este texto nos concentraremos en un caso particular de espacio vectorial: $\mathbb{R}^n$, mismo que se describirá más adelante.

Obsérvese que para los conjuntos de números descritos arriba es intuitivamente claro cómo $\mathbb{Z}$ puede obtenerse a partir de $\mathbb{N}$ y $\mathbb{Q}$ a partir de $\mathbb{Z}$. Asimismo, todos estos conjuntos tienen el mismo número de elementos (o cardinalidad) que $\mathbb{N}$ y, por lo tanto, son numerables. La demostración de este hecho es relativamente simple y puede consultarse en cualquier texto elemental de teoría de conjuntos. La construcción del conjunto de números reales, denotado por $\mathbb{R}$, es mucho más compleja de entender ya que algebraicamente posee la misma estructura que los números racionales; sin embargo, difiere sustancialmente en cuanto a su estructura cardinal y geométrica (o más bien topológica). El detalle de cómo obtener a $\mathbb{R}$ a partir de $\mathbb{Q}$ va más allá de los alcances de este libro y pertenece más bien al ámbito de un curso de análisis matemático.

Recordemos que cualquier elemento de $\mathbb{Q}$ tiene una expresión decimal finita o bien infinita pero con un patrón regular. A diferencia de los racionales, los números irracionales son aquellos tales que su expansión decimal es infinita y no presenta regularidad alguna como, por ejemplo, $\pi = 3.1416\ldots$, $e = 2.7182\ldots$, $\sqrt{2} = 1.4142, \ldots$, etc.Como conjunto, los reales son la unión de los racionales y los irracionales y representan a todos los puntos de una recta infinita, o la “recta real” (ver la Figura 1-1).

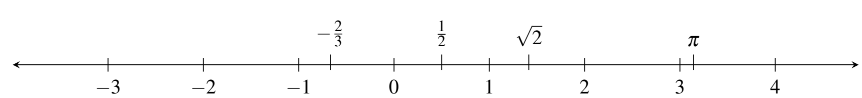

Figura 1-1 Recta Real

Cada uno de estos conjuntos de números extiende al anterior puesto que

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

y su construcción está motivada por la necesidad de resolver ecuaciones.

Por ejemplo, para resolver las ecuaciones

$$\begin{aligned} x + 1 &= 0, \\ 2x &= 1, \\ x^2 &= 2, \end{aligned}$$

es necesaria la construcción de $\mathbb{Z}$, $\mathbb{Q}$ y $\mathbb{R}$, respectivamente. Observemos que la ecuación

$$x^2 = -1, \tag{1.1}$$

en apariencia inofensiva, no tiene solución en ninguno de estos conjuntos de números. Para resolver esta ecuación se requiere del conjunto de los números complejos, denotado por $\mathbb{C}$, mismos que extienden a los números reales de manera que $\mathbb{R} \subset \mathbb{C}$. En este curso no utilizaremos a los números complejos y todos los coeficientes y variables toman valores reales o bien en alguno de los subconjuntos $\mathbb{N}, \mathbb{Z}$ o $\mathbb{Q}$.

Demostraciones matemáticas para dummies

En matemáticas, existen enunciados fundamentales que llamamos axiomas y que asumimos ciertos sin necesidad de demostrarlos. El resto de la disciplina consiste de resultados cuya validez debe demostrarse, ya sea a partir de los axiomas o de otros resultados que ya se han validado previamente. Cada enunciado matemático puede ser verdadero o falso (¡no se permite que a veces sea verdadero y a veces falso!) y en caso de que sea verdadero decimos que el enunciado es válido. Típicamente se tiene un enunciado compuesto del tipo: "Si X, entonces Y" o bien algo equivalente. Este enunciado compuesto es válido si cuando X es válido, entonces Y también lo es. Para demostrar que esta aseveración es cierta, asumimos, por hipótesis, que X es válido y mediante una secuencia de pasos que se siguen lógicamente el uno del otro, llegamos a que Y es válido. Se dice fácil, pero como siempre, ¡el diablo está en los detalles! Veamos un ejemplo sencillo.

Ejemplo 1.4 *En este ejemplo, la palabra número se refiere a un número entero. Se define el siguiente subconjunto de enteros:*

$$Pares = \{x \in \mathbb{Z} \mid x = 2n, \ \text{para algún } n \in \mathbb{Z}\}. \tag{Par}$$

*Demostrar la validez del siguiente enunciado: "**el producto de dos números pares es par**". ¿Por dónde comenzamos? Primero, entendamos que significa la definición Par: si nos dicen que x es un entero par, entonces x puede escribirse como $2n$ para algún entero n, en otras palabras, x es un múltiplo de 2. Segundo, es conveniente reescribir el enunciado de la siguiente forma: "**Para cualesquiera enteros** a **y** b,*

***si** a **y** b **son pares, entonces** ab **es par".** Tomamos como punto de partida que tenemos dos números pares arbitrarios, a y b y procedemos como sigue: como a y b son pares, existen enteros m y n tales que $a = 2m$ y $b = 2n$, por lo tanto,*

$$ab = (2m)(2n) = 2(2mn)$$

y tenemos que ab también es un múltiplo de 2, con lo cual es par y concluímos la demostración. ■ *(Este pequeño cuadrado lo utilizaremos para denotar el final de una demostración).*

Notemos que en la demostración anterior probamos que cuando el enunciado "Si a y b son pares" es verdadero, entonces el enunciado "ab es par" también lo es. Decimos que "ab es par" es una condición **necesaria** para que a y b sean pares. Si ab fuese impar, sería imposible que a y b fuesen pares. No obstante, "ab es par" no es una condición **suficiente** para garantizar que a y b sean pares, por ejemplo, podríamos tener: $a = 2$, $b = 3$ por lo que $ab = (2)(3) = 6$, que es par; sin embargo, $b = 3$ no lo es.

En ocasiones se requiere demostrar la equivalencia de dos enunciados matemáticos, digamos X y Y. Para este propósito se utiliza el siguiente enunciado compuesto:

$$X \text{ si y sólo si } Y.$$

Esto es simplemente una forma abreviada de poner

$$\text{Si } X, \text{ entonces } Y \text{ y si } Y, \text{ entonces } X.$$

Para que este enunciado compuesto sea válido, bajo cualquier escenario los enunciados X y Y deben ser ambos verdaderos o ambos falsos. En esta situación, la validez de X es una condición necesaria y suficiente para la validez de Y.

Ejemplo 1.5 *Para este ejemplo se define el conjunto de los enteros impares (o nones) como sigue:*

$$\textit{Impares} = \{x \in \mathbb{Z} \mid x = 2n + 1, \text{ para algún } n \in \mathbb{Z}\}. \qquad \text{(Impar)}$$

Demostrar que para todo entero x se tiene que $x + 2$ es impar si si sólo si $x + 5$ es par. Probemos primero "Si $x + 2$ es impar, entonces $x + 5$ es par. Como punto de partida, tenemos un número entero arbitrario x tal que $x + 2$ es impar, es decir, podemos escribir a $x + 2$ como

$$x + 2 = 2n + 1$$

para algún entero n. Nuestra meta es demostrar que $x+5$ es par. Para este efecto, sumamos 3 en ambos lados de la expresión anterior para obtener

$$x + 2 + 3 = 2n + 1 + 3,$$

por lo tanto

$$x + 5 = 2n + 4 = 2(n + 2).$$

Esto nos dice que $x+5$ es par (cumple la Definición Par) y concluimos esta parte de la demostración. Falta demostrar que "Si $x+5$ es par, entonces $x+2$ es impar". Procedemos de forma análoga: como $x+5$ es par podemos escribirlo como

$$x + 5 = 2m$$

para algún entero m. Restando 3 en ambos lados de esta expresión obtenemos

$$x + 2 = 2m - 3 = 2m - 4 + 1 = 2(m - 2) + 1.$$

De esta última expresión concluimos que $x+2$ es impar ya que cumple con la Definición Impar, concluyendo así la demostración.■

Para concluir con esta sección veremos como demostrar que un enunciado es verdadero, ¡asumiendo que es falso y llegando a una contradicción! Por ejemplo, podríamos demostrar la validez del enunciado compuesto "si X, entonces Y", asumiendo que Y es falso y X verdadero y, a partir de ahí llegar a una contradicción, concluyendo que no podemos asumir que Y sea falso[3]. Esto puede verse en el caso de la demostración del enunciado "si a y b son pares, entonces ab es par". Antes de comenzar, observemos las siguientes relaciones entre conjuntos:

$$\mathbb{Z} = Pares \cup Impares,$$
$$\varnothing = Pares \ \cap \ Impares.$$

De aquí deducimos que cualquier número entero debe ser par o impar y ningún entero es par e impar simultáneamente. Procedamos ahora a la demostración del enunciado: dados a y b pares, supongamos que ab no es par. Entonces ab es impar y puede escribirse de la forma

$$ab = 2n + 1$$

para algún entero n. Como a y b son pares, podemos escribirlos como $a = 2p$, $b = 2q$, para enteros p y q. En este caso el producto

$$ab = (2p)(2q) = 2(2pq),$$

[3]Formalmente se demuestra que la negación del enunciado original "Si X, entonces Y" no es válida.

es decir, es par. Esto último nos lleva a una contradicción pues habíamos supuesto que ab era impar, con lo cual (aunque usted no lo crea) se concluye la demostración ■

La validez de la demostración anterior se debe a que el enunciado "si X, entonces Y" es equivalente al enunciado "si no Y, entonces no X", en donde "no X" se refiere a la negación del enunciado X. Es común realizar demostraciones por contradicción cuando se trata de probar la existencia de alguna propiedad u objeto. Dos enunciados clásicos (conocidos y demostrados en la antigua Grecia) cuyas demostraciones se efectúan por contradicción son.

- El número de enteros primos es infinito.
- $\sqrt{2}$ es un número irracional.

El lector interesado puede encontrar las demostraciones en múltiples libros e-libros, páginas web, etcétera. La idea general de la demostración del primer enunciado es suponer que el conjunto de todos los números primos es finito, digamos N. A partir de ahí se construye un primo que no pertenece a este conjunto, lo cual es una contradicción. Claramente, dado que no existe una forma recursiva de construir al conjunto de primos, no puede probarse en forma directa que es infinito.

Para la demostración del segundo enunciado, se parte del hecho de que los números reales son racionales o irracionales y no pueden ser ambos simultáneamente. Asumimos que $\sqrt{2}$ no es irracional, por lo tanto es racional y puede expresarse como

$$\sqrt{2} = \frac{a}{b},$$

para a, b enteros, $b \neq 0$. A partir de aquí se llega a una contradicción y concluimos que $\sqrt{2}$ debe ser irracional. Una vez más, sería imposible hacer la demostración en forma directa ya que tendríamos que escribir a $\sqrt{2}$ con una expresión decimal infinita sin patrón alguno.

Principio de Inducción

En matemáticas surge con frecuencia la necesidad de demostrar que una propiedad es válida para todos los números naturales (o para todos los naturales a partir de cierto punto). En este caso se utiliza el llamado **Principio de Inducción** para demostrar la propiedad en cuestión.

Principio de Inducción Sea P una propiedad definida para los números Naturales, supongamos que 1 satisface P y cuando $n-1$ satisface P podemos demostrar que n satisface P. Entonces P es válida para todos los números Naturales.

Observación 1.6 *La demostración de que* 1 *satisface* P *se conoce como la* ***base de la inducción****.*

El siguiente ejemplo es clásico y cuenta la leyenda que el niño K.F. Gauss fue quien encontró esta fórmula cuando asistía a la escuela primaria.

Ejemplo 1.7 *Demostrar que para toda* $n \in \mathbb{N}$ *se cumple la siguiente*

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}, \tag{1.2}$$

es decir, la suma de los primeros n *naturales se calcula simplemente como* $\frac{n(n+1)}{2}$*. De acuerdo al Principio de Inducción, primero demostramos la fórmula (1.2) para* $n = 1$*. Este caso es trivial puesto que cláramente*

$$1 = \frac{1(1+1)}{2}.$$

Ahora bien, suponemos que (1.2) es válida para $n - 1$*, por lo tanto se cumple*

$$1 + 2 + \cdots + (n-1) = \frac{(n-1)((n-1)+1)}{2} = \frac{(n-1)(n)}{2}. \tag{1.3}$$

Sumando n *en ambos lados de (1.3) se tiene que*

$$\begin{aligned} 1 + 2 + \cdots + (n-1) + n &= \frac{(n-1)(n)}{2} + n \\ &= \frac{(n-1)(n) + 2n}{2} \\ &= \frac{n((n-1)+2)}{2} \\ &= \frac{n(n+1)}{2} \end{aligned}$$

y podemos concluir que (1.2) es válida para todo $n \in \mathbb{N}$ $\blacksquare$

Aprender a realizar demostraciones es algo que requiere de mucha práctica, así que la mejor forma de dominar el tema es “aprendiendo haciendo”.

Capítulo 2

Sistemas de Ecuaciones Lineales

Introducción

Como es bien sabido, una ecuación lineal en una variable es una igualdad de la forma

$$ax = b,$$

donde a es el *coeficiente de la variable* x y b es el *término independiente*. La calificación de "lineal" se refiere a que la variable "x" aparece con exponente igual a uno. Resolver la ecuación se refiere a encontrar los valores de x, si es que existen, para los cuales se satisface la igualdad y el conjunto de soluciones se denota po S. La resolución de esta ecuación tiene dos casos posibles.

Caso 1: $a \neq 0$. En esta situación es posible multiplicar ambos lados de la igualdad anterior por $\frac{1}{a}$, de manera que la igualdad se conserva, obteniendose[1]

$$x = \frac{1}{a}(ax) = \frac{b}{a}.$$

Por ejemplo, la solución de la ecuación $4x = 2$ es $x = \frac{2}{4} = \frac{1}{2}$. Así, nuestro conjunto solución es $S = \left\{\frac{1}{2}\right\}$ o bien

$$S = \left\{x \in \mathbb{R} \mid x = \frac{1}{2}\right\},$$

[1]Notemos que la división entre a puede realizarse ya que $a \neq 0$.

que es un conjunto con un único elemento.

Caso 2: $a = 0$**.** Aquí existen dos subcasos posibles.

Subcaso 2.1: $b \neq 0$**.** Como ningún número real multiplicado por cero es igual a algo distinto de cero, el conjunto solución asociado es vacío, es decir, $S = \emptyset$. Este es el caso de la ecuación $0x = 7$.

Subcaso 2.2: $b = 0$**.** Como cualquier número real multiplicado por cero es igual a cero, el conjunto solución de la ecuación $0x = 0$ es $S = \mathbb{R}$, es decir, cualquier número real.

Consideremos ahora ecuaciones lineales con dos variables (al igual que antes, los términos que contienen a las variables son de orden uno), o bien, ecuaciones de la forma

$$ax + by = c,$$

donde a es el *coeficiente de la variable* x, b es el *coeficiente de la variable* y y c es el *término independiente*. La resolución de esta ecuación tiene cuatro casos posibles.

Caso 1: $a \neq 0$ **y** $b \neq 0$**.** Aquí, al despejar y obtenemos que $y = \frac{c}{b} - \frac{a}{b}x$ y el correspondiente conjunto solución puede visualizarse en el plano cartesiano como una recta (con pendiente $-\frac{a}{b}$ y que corta al eje vertical en el punto $(0, \frac{c}{b})$) que atraviesa ambos ejes del plano.

Caso 2: $a = 0$ **y** $b \neq 0$**.** En esta situación obtenemos $0x + by = c$. Despejando y obtenemos que $y = \frac{c}{b}$. Por ejemplo, la pareja (x, y) es solución de $0x + 4y = 2$ siempre que $y = \frac{1}{2}$ y sin ninguna restricción sobre x.Un momento de reflexión nos permite percatarnos que, de hecho, x puede tomar cualquier valor por lo que el conjunto solución es

$$S = \left\{ (x, \frac{1}{2}) \mid x \in \mathbb{R} \right\}.$$

Este conjunto puede visualizarse en el plano cartesiano como la recta $y = \frac{1}{2}$ que, evidentemente, es paralela al eje de las x's. Es claro que este caso podría pensarse como un caso particular del anterior cuando la pendiente de la recta es igual a cero.

Caso 3: $a \neq 0$ **y** $b = 0$**.** Aquí obtenemos $ax + 0y = c$. Despejando x obtenemos que $x = \frac{c}{a}$, por lo que el correspondiente conjunto solución es $S = \{(\frac{c}{a}, y) \mid y \in \mathbb{R}\}$. Este conjunto puede visualizarse en el plano cartesiano como la recta $x = \frac{c}{a}$ que, evidentemente, es paralela al eje vertical.

Caso 4: $a = 0$ **y** $b = 0$**.** Existen dos subcasos posibles.

Subcaso 4.1: $c \neq 0$**.** La ecuación correspondiente es $0x + 0y = c$. y el conjunto solución asociado es $S = \emptyset$.

Subcaso 4.2: $c = 0$**.** Como cualquier pareja de reales (x, y) satisface la ecuación $0x + 0y = 0$, el conjunto solución de esta ecuación es $S = \{(x, y) \mid x, y \in \mathbb{R}\}$.

En la Figura 1.2 se representan los conjuntos solución de los casos 1 (si $a, b, c, d > 0$) y 2:

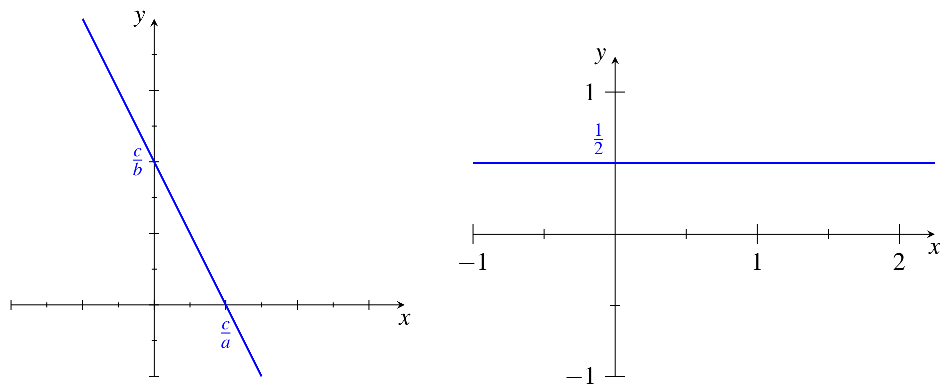

Figura 1.2

Recordemos ahora que $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$ es el conjunto de todas las parejas ordenadas cuyas componentes son números reales de manera que $S = \mathbb{R}^2$ en el subcaso 2. Análogamente, $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$ se define como el conjunto de todas las tripletas ordenadas cuyas componentes son números reales. En general, tenemos la siguiente definición para cualquier número natural n.

Definición 2.1 $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\}$ *es el conjunto de todas las n-tuplas o $n-$adas ordenadas cuyas componentes son números reales. A los elementos de $\mathbb{R}^n$ también se les llama* ***vectores*** *de $\mathbb{R}^n$.*

Sistemas de ecuaciones lineales de dos por dos

Un sistema con dos ecuaciones y dos variables x, y es un sistema de la forma

$$\begin{cases} ax + by = c, \\ a'x + b'y = c', \end{cases}$$

donde a, b, c, a', b' y c' son números reales. Decimos que un par ordenado (t_1, t_2) es solución del sistema si y sólo si ("syss" de aquí en adelante) es solución de cada una de las dos ecuaciones que componen al sistema.

Ejemplo 2.2 *Encontrar el conjunto solución del sistema*

$$\begin{cases} 2x - 2y = 8, & \textit{(I)} \\ -4x + 2y = -6. & \textit{(II)} \end{cases} \tag{A}$$

Si multiplicamos ambos lados de la ecuación I por $\frac{1}{2}$, obtenemos el nuevo sistema

$$\begin{cases} x - y = 4, & \textit{(III)} \\ -4x + 2y = -6. & \textit{(II)} \end{cases} \tag{B}$$

Notemos que los sistemas A y B son equivalentes *en el sentido de que tienen el mismo conjunto solución. De hecho, por la forma en que se obtuvo B, es claro que toda solución de A también satisface B. Para comprobar que toda solución de B también es solución de A basta multiplicar ambos lados de (III) por 2 (equivalentemente $\left(\frac{1}{2}\right)^{-1}$).*

Ahora, sumando 4 veces la ecuación (III) a la ecuación (II) obtenemos el sistema

$$\begin{cases} x - y = 4, & \textit{(III)} \\ -2y = 10. & \textit{(IV)} \end{cases} \tag{C}$$

Observar que como

$$(IV) = 4(III) + (II),$$

se sigue que

$$(II) = (IV) - 4(III),$$

de manera que los sistemas B y C también son equivalentes. Pero como A es equivalente a B concluimos que los tres sistemas tienen exáctamente el mismo conjunto solución.

Como siguiente paso multiplicamos ambos lados de la ecuación (IV) por $-\frac{1}{2}$ obteniéndose el sistema equivalente

$$\begin{cases} x - y = 4, & \textit{(III)} \\ y = -5. & \textit{(V)} \end{cases} \tag{D}$$

Finalmente, sumando 1 por la ecuación (V) a (III)[2] *llegamos a*

$$\begin{cases} x = -1, & \textit{(VI)} \\ y = -5. & \textit{(V)} \end{cases} \tag{E}$$

[2] Evidentemente, esto es lo mismo que sumar (V) a (III), pero queremos presentar esta operación como un caso particular de un método más general.

Este último sistema, cuya solución se lee en forma directa, es equivalente a todos los anteriores. De esta forma, el conjunto solución de todos estos sistemas está dado por

$$S = \{(x, y) \in \mathbb{R}^2 | x = -1, y = -5\}$$

o simplemente $S = \{(-1, -5)$.

Los pasos que hemos empleado en esta resolución pueden sistematizarse de una forma más efectiva. Para ello, basta notar que en los primeros cuatro sistemas involucrados aparecen siempre dos renglones con los cuales realizamos ciertas operaciones que podemos visualizar como sigue:

$$\left[\begin{array}{cc|c} 2 & -2 & 8 \\ -4 & 2 & -6 \end{array}\right] \overset{\frac{1}{2}R_1}{\sim} \left[\begin{array}{cc|c} 1 & -1 & 4 \\ -4 & 2 & -6 \end{array}\right] \overset{4R_1+R_2}{\sim} \left[\begin{array}{cc|c} 1 & -1 & 4 \\ 0 & -2 & 10 \end{array}\right] \overset{-\frac{1}{2}R_2}{\sim}$$

$$\left[\begin{array}{cc|c} 1 & -1 & 4 \\ 0 & 1 & -5 \end{array}\right] \overset{1R_2+R_1}{\sim} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & -5 \end{array}\right].$$

El símbolo $\sim$ significa que hay equivalencia entre sistemas. Adicionalmente, kR_i (donde k es un número real distinto de cero) significa que multiplicaremos el renglón i por k mientras que $kR_j + R_l$ indica que la equivalencia se obtiene sumando un múltiplo (k) de una ecuación (de un renglón R_j) a otra (otro renglón R_l) con el fin de, literalmente, aniquilar variables y poder llegar así a la solución. Este paso se conoce como *método de eliminación*. Más adelante, en el ejemplo 2.16, consideraremos una última operación entre matrices de tal forma que con estas tres operaciones seremos capaces de resolver cualesquier sistema de ecuaciones (sin importar que tan grande sea éste).

Geométricamente, cada una de las ecuaciones originales representa una recta en $\mathbb{R}^2$. Las dos rectas se cortan en el punto $(-1, -5)$, como se ilustra en la Figura 1.3.

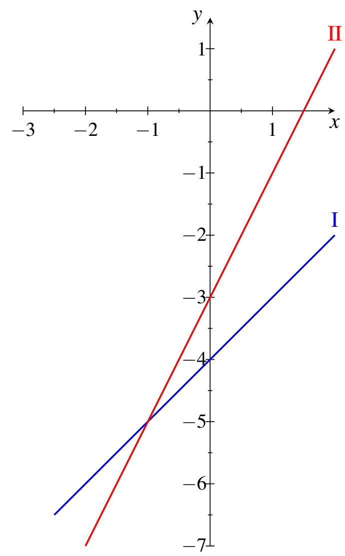

Figura 1.3

El lector se preguntará el por qué utilizar el método anterior para resolver el sistema de dos ecuaciones con dos variables. Después de todo, podríamos utilizar el llamado método de sustitución (despejar una variable de una ecuación, sustituirla en la otra y resolver). Sin embargo, la sustitución se vuelve inmanejable cuando se tienen más ecuaciones y más variables; afortunadamente, el método que acabamos de introducir se extiende de forma inmediata a los casos más complejos. Para poder formalizarlo, introducimos el concepto de "matriz" en la siguiente sección.

Matrices

Una **matriz** A de $m \times n$ es un arreglo rectangular de $m \times n$ reales colocados en m renglones horizontales y n columnas verticales. De esta forma, A puede visualizarse como

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

La notación $A \in M_{m\times n}(\mathbb{R})$ es una abreviatura para decir que es una matriz A de $m \times n$ con entradas o componentes reales. Dado que en estas notas estas entradas siempre serán números reales utilizaremos la notación simplificada $A \in M_{m\times n}$. En general las matrices se denotan por letras mayúsculas: $A, B, C, \ldots$

Definición 2.3 *Dada $A \in M_{m\times n}$, $1 \leq i \leq m$ y $1 \leq j \leq n$, A_i denota al renglón i de A, A^j denota a la columna j de A y a_{ij} denota a la componente de A que se localiza en el renglón i y en la columna j de A (obsérvese que se utiliza la letra minúscula "a" que corresponde a la mayúscula que denota a la matriz). De esta forma,*

$$A_i = \begin{bmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{bmatrix},$$
$$A^j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}.$$

Formalmente, y para evitar posibles confusiones cuando m o n son números mayores a 9, la componente a_{ij} debería denotarse por $a_{i,j}$. Sin embargo, cuando el contexto sea suficientemente claro omitiremos el uso de comas para simplificar la notación.

Ejemplo 2.4 *Sean $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 6 & 3 \end{bmatrix}$ y $C = \begin{bmatrix} 1 & 0 & 6 \\ 7 & 8 & 3 \\ 9 & 6 & 4 \end{bmatrix}$. Utilizando la notación que acabamos de introducir tenemos que $A \in M_{3\times 2}$, $B \in M_{1\times 3}$ y $C \in M_{3\times 3}$. Además,*

$$A_1 = \begin{bmatrix} 1 & 2 \end{bmatrix}, a_{32} = 6, C_2 = \begin{bmatrix} 7 & 8 & 3 \end{bmatrix} \text{ y } C^2 = \begin{bmatrix} 0 \\ 8 \\ 6 \end{bmatrix}.$$

Definición 2.5 *Sean $A, B \in M_{m\times n}$. Decimos que $A = B$ syss ambas matrices coinciden entrada por entrada. Esto es, syss $a_{ij} = b_{ij}$ para toda i y toda j.*

Ejemplo 2.6 *Si*

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}, C = \begin{bmatrix} 2 & 3 \end{bmatrix} \text{ y } D = \begin{bmatrix} 3 & 2 \end{bmatrix},$$

entonces $A \neq B$ (pues $a_{22} \neq b_{22}$) y $C \neq D$ (pues $c_{11} = 2 \neq 3 = d_{11}$).

En lo que sigue haremos un recuento taxonómico de algunas matrices relevantes.

Definición 2.7 *La matriz cero* $\Theta \in M_{m\times n}$ *es la matriz de* $m \times n$ *cuyas entradas son todas iguales a cero. Por ejemplo, si* $\Theta \in M_{2\times 2}$ *y* $\Theta' \in M_{2\times 3}$ *son las correspondientes matriz cero, entonces*

$$\Theta = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \ y \ \Theta' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Definición 2.8 *Sea* $A \in M_{m\times n}$. *Decimos que* A *es una* ***matriz cuadrada*** *de orden* n *syss* $m = n$.

Ejemplo 2.9 *La matriz* $\begin{bmatrix} 1 & 0 \\ 7 & 8 \end{bmatrix}$ *es cuadrada de orden 2.*

Definición 2.10 *Si*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \ddots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

es una matriz cuadrada de orden n, *entonces la* ***diagonal principal*** *se define como la sucesión* $a_{11}, a_{22}, \dots a_{nn}$.

Ejemplo 2.11 *La diagonal principal de* $\begin{bmatrix} 1 & 0 \\ 7 & 8 \end{bmatrix}$ *es 1, 8.*

Definición 2.12 *La matriz identidad* $I \in M_{n\times n}$ *es la matriz cuadrada de orden* n *tal que todos sus elementos en la diagonal principal son 1 y tal que todos sus elementos que están fuera de dicha diagonal son cero. Alternativamente,* $A = I$ *syss*

$$a_{ij} = \begin{cases} 1 \ si \ i = j, \\ 0 \ si \ i \neq j. \end{cases}$$

Así, la matriz identidad $I \in M_{n\times n}$ *se visualiza como*

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

Definición 2.13 *Decimos que* $\vec{v}$ *es un vector de* $\mathbb{R}^n$ *syss* $\vec{v} \in M_{n\times 1}$ *o* $\vec{v} \in M_{1\times n}$ *es decir, syss* $\vec{v}$ *es una columna o un renglón con* n *componentes.*

Aunque $\begin{bmatrix} \pi \\ -3 \end{bmatrix}$ y $\begin{bmatrix} \pi & -3 \end{bmatrix}$ son formalmente objetos distintos (pues el primero es una matriz con dos renglones y una columna mientras que el segundo es una matriz con un renglón y dos columnas), resulta natural decir que ambos coinciden en la primera y segunda componente (tales componenetes son respectivamente π y -3). Por ello, los vectores de $\mathbb{R}^n$ pueden representarse como columnas o como renglones y la Definición 2.13 no entra en conflicto con la Definición 2.1.

Representación matricial

Un sistema lineal de m ecuaciones y n variables es un conjunto con m igualdades de la forma

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{gathered}$$

Decimos que el vector $(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$ es una solución del sistema syss al reemplazar cada variable x_i por el valor t_i se satisfacen las m igualdades que lo componen. Por lo tanto, el conjunto solución es siempre un subconjunto de $\mathbb{R}^n$ (aunque este conjunto puede ser vacío).

Utilizando notación matricial representamos nuestro sistema como

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Esta matriz es conocida como **la matriz aumentada del sistema**. El vector

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

es decir, la última columna de la matriz aumentada se llama el **vector de términos independientes.** Finalmente, la matriz

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

es decir, la submatriz que se obtiene al omitir la última columna de la matriz aumentada se llama **la matriz de coeficientes**. Obsérvese que la columna i corresponde a los coeficientes de la variable x_i, de manera que debe tenerse sumo cuidado si se llega a cambiar el orden de las columnas. Así pues, la matriz aumentada del sistema puede representarse como

$$(A \mid \vec{b})$$

Ejemplo 2.14 *La matriz aumentada del sistema*

$$\begin{aligned} w + 3x + 2y + 6z &= 7, \\ w + 8x + \quad + z &= -1, \\ w + x + y + z &= 1, \end{aligned}$$

es la matriz

$$\left[\begin{array}{cccc|c} 1 & 3 & 2 & 6 & 7 \\ 1 & 8 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

Además,

$$\begin{bmatrix} 7 \\ -1 \\ 1 \end{bmatrix} \quad y \quad \begin{bmatrix} 1 & 3 & 2 & 6 \\ 1 & 8 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

son, respectivamente, el vector de términos independientes y la matriz de los coeficientes de este sistema.

Evidentemente, los sistemas

$$\left\{ \begin{aligned} w + 3x + 2y + 6z &= 7, \\ w + 8x + \quad + z &= -1, \\ w + x + y + z &= 1, \end{aligned} \right. \qquad \text{y} \qquad \left\{ \begin{aligned} w + 3x + 2y + 6z &= 7, \\ w + x + y + z &= 1, \\ w + 8x + \quad + z &= -1 \end{aligned} \right.$$

son equivalentes, por lo que tiene sentido escribir

$$\left[\begin{array}{cccc|c} 1 & 3 & 2 & 6 & 7 \\ 1 & 8 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right] \overset{R_2 \leftrightarrow R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 3 & 2 & 6 & 7 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 8 & 0 & 1 & -1 \end{array}\right].$$

Así pues, contamos con tres diferentes tipos de **operaciones elementales** sobre matrices que preservan el conjunto solución entre los sistemas de ecuaciones subyacentes:

Definición 2.15

1. El **tipo 1** corresponde a **intercambiar renglones**.
2. El **tipo 2** a **multiplicar un renglón por una constante distinta de cero**.
3. El **tipo 3** corresponde a **sumar un múltiplo de un renglón a otro renglón** conocido como el método de eliminación (ver ejemplo 2.2).

Ejemplo 2.16 *Utilizando operaciones elementales encontraremos el conjunto solución del sistema*

$$\begin{aligned} 2x + y + 3z &= 0, \\ x + y + z &= 1, \\ 2x + y - z &= 4. \end{aligned}$$

Para ello notamos que

$$\left[\begin{array}{ccc|c} 2 & 1 & 3 & 0 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & 4 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & 1 & 3 & 0 \\ 2 & 1 & -1 & 4 \end{array}\right] \overset{-2R_1+R_2}{\sim}$$
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & -2 \\ 2 & 1 & -1 & 4 \end{array}\right] \overset{-2R_1+R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & -2 \\ 0 & -1 & -3 & 2 \end{array}\right] \overset{-R_2}{\sim}$$
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & -1 & -3 & 2 \end{array}\right] \overset{1R_2+R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & -4 & 4 \end{array}\right] \overset{-\frac{1}{4}R_3}{\sim}$$
$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 1 & -1 \end{array}\right] \overset{1R_3+R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right] \overset{-1R_3+R_1}{\sim}$$
$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right] \overset{-1R_2+R_1}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

En particular, esta última matriz nos dice que

$$\begin{aligned} 1x + 0y + 0z &= 1, \\ 0x + 1y + 0z &= 1, \\ 0x + 0y + 1z &= -1. \end{aligned}$$

Por lo tanto, nuestro conjunto solución es $S = \{(1, 1, -1)\}$.

Geométricamente, cada ecuación representa un plano en $\mathbb{R}^3$, mismos que se intersectan en el punto $(1, 1, -1)$. Esto se representaría visualmente como sigue:

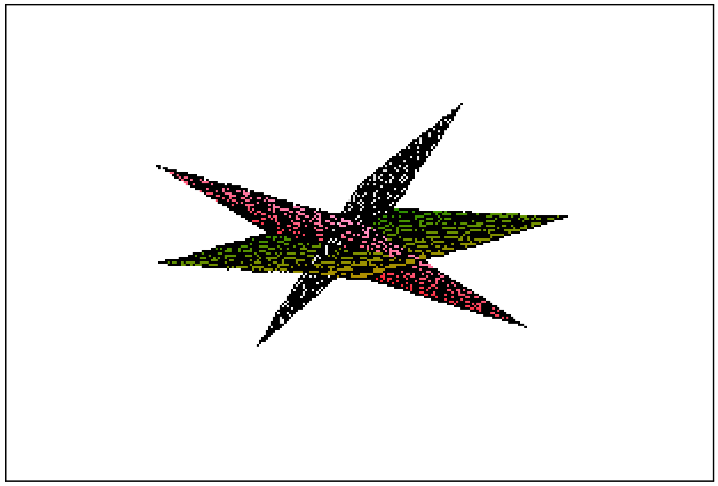

Figura 1.4

El orden y la elección de las operaciones elementales que hemos escogido en el ejemplo anterior no son únicos pero si tienen una cierta lógica. Esencialmente, intentamos llevar a la matriz original a lo que se conoce como su *forma escalonada reducida*. Este concepto será definido formalmente en la siguiente sección pero por lo pronto adelantamos que, para llegar a esta forma, buscamos crear (mediante operaciones elementales de los tres tipos) una escalera descendente de altura máxima y con escalones tan estrechos como sea posibe.

Utilizamos primero el método de eliminación de arriba hacia abajo y de izquierda a derecha para aniquilar a todas las componentes que están por debajo del inicio de cada uno de los futuros escalones. Las operaciones elementales de tipo dos son empleadas a conveniencia y sirven, entre otras cosas, para asegurarnos que si un renglón tiene al menos una componenete distinta de cero, entonces su primera componenete distinta de cero es igual a uno. Este uno es llamado un **uno principal**[3]. Finalmente, recurrimos al método de eliminación - esta vez de abajo hacia arriba y de derecha a izquiera - con el fin de garantizar que si una columna tiene un uno principal, entonces todas las otras entradas de esa columna son cero.

Definición 2.17 *Decimos que un sistema lineal con m ecuaciones y n variables es* ***homogéneo*** *syss su vector correspondiente de términos independientes es igual al vector cero de $\mathbb{R}^m$.*

[3] En general, un **elemento principal** se refiere a la primera componente distinta de cero de un renglón, independientemente de si es o no igual a uno.

Definición 2.18 *Un sistema lineal es* ***consistente*** *syss tiene al menos una solución. Si el conjunto solución es vacío, entonces diremos que el sistema es* ***inconsistente****.*

Proposición 2.19 *Todo sistema lineal homogéneo es consistente.*

Demostración. Basta notar que si un sistema homogéneo tiene m ecuaciones y n variables, entonces el vector cero de $\mathbb{R}^n$ siempre pertenece al conjunto solución de dicho sistema de manera que este conjunto nunca es vacío. ■

Como el vector cero de $\mathbb{R}^n$ siempre es solución de un sistema homogéneo con n variables, decimos que el vector cero es **la solución trivial** de cualesquiera de estos sistemas. Por supuesto que algunos sistemas homogéneos tienen muchas soluciones (este es el caso del sistema cuya única ecuación es $x - y = 0$ y cuyo conjunto solución es $\{(x, x) \mid x \in \mathbb{R}\}$) y algunos otros sólo tienen la solución trivial.

Ejemplo 2.20 *Usando la heurística previa a la definición, resolveremos el sistema homogéneo*

$$\begin{aligned} 2y + 3z &= 0, \\ 2x - 3y + 2z &= 0, \\ 3x + \frac{1}{2}y - z &= 0. \end{aligned}$$

Para ello notamos que

$$\left[\begin{array}{ccc|c} 0 & 2 & 3 & 0 \\ 2 & -3 & 2 & 0 \\ 3 & \frac{1}{2} & -1 & 0 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 2 & 0 \\ 0 & 2 & 3 & 0 \\ 3 & \frac{1}{2} & -1 & 0 \end{array}\right] \overset{-\frac{3}{2}R_1 + R_3}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 2 & 0 \\ 0 & 2 & 3 & 0 \\ 0 & 5 & -4 & 0 \end{array}\right]$$

$$\overset{-\frac{5}{2}R_2 + R_3}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 2 & 0 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -\frac{23}{2} & 0 \end{array}\right] \overset{-\frac{2}{23}R_3}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 2 & 0 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \overset{-3R_3 + R_2}{\sim}$$

$$\left[\begin{array}{ccc|c} 2 & -3 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \overset{-2R_3 + R_1}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \overset{\frac{1}{2}R_2}{\sim} \left[\begin{array}{ccc|c} 2 & -3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

$$\overset{3R_2 + R_1}{\sim} \left[\begin{array}{ccc|c} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \overset{\frac{1}{2}R_1}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

Por lo tanto, $x = 0$, $y = 0$, $z = 0$. Dicho de otra forma, nuestro conjunto solución consiste únicamente de la solución trivial $\{(0, 0, 0)\}$.

Forma escalonada reducida y teorema de Gauss-Jordan

El lector observador habrá notado que en todos los ejemplos anteriores, la forma final de la matriz tiene ciertas características, mismas que resumimos a continuación.

Definición 2.21 *Una matriz está en **forma escalonada** syss se cumplen las siguientes tres condiciones:*

1. *Si la matriz tiene renglones que consisten únicamente de ceros, esos renglones están en su parte inferior.*
2. *Si un renglón tiene componentes distintas de cero, entonces la primera (de izquierda a derecha) de esas componentes distintas de cero es igual a uno. Ese uno es llamado un **uno principal**.*
3. *Para cada renglón que tenga un uno principal, éste aparaece a la derecha y por debajo de los unos principales que aparecen en los renglones precedentes.*

Definición 2.22 *Una matriz está en **forma escalonada reducida** syss satisface las tres condiciones de la Definición 2.21 junto con la siguiente condición adicional.*

4. *Cada uno principal es el único elemento diferente de cero dentro de la columna en la cual se encuentra.*

Observación 2.23 *Cuando una matriz $A \in M_{m\times n}$ está en forma escalonada reducida, por construcción, el número de unos principales es menor o igual al mínimo entre m y n.*

Ejemplo 2.24 *Las matrices*

$$\begin{bmatrix} 0 & 1 & \pi & 2 & 7 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

están en forma escalonada, sin embargo no están en forma escalonada reducida pues no cumplen la propiedad (4). Las matrices

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

no están en forma escalonada, en el primer caso no se satisface (2) *y en el segundo no se cumple* (1).

Ejemplo 2.25 *Las matrices*

$$A = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 7 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \Theta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

están en forma escalonada reducida. Sin embargo, las siguientes matrices no están en forma escalonada reducida:

$$C = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 1 & 5 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad F = \begin{bmatrix} 0 & 1 & 8 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Nótese que C no satisface (1)*, D no satisface* (2)*, E no satisface* (3) *y F no satisface* (4) *(aunque F sí está en forma escalonada).*

El método de tener un arreglo rectangular de los coeficientes (tipo matriz) y llegar a la forma escalonada para encontrar la solución era ya conocido en la antigua China desde 179 dC. En el siglo XIX, K.F.Gauss utilizó este método por lo cual se denomina *eliminación gaussiana* a la obtención de la forma escalonada. En 1887, W. Jordan extendió este procedimiento para llegar a la forma escalonada reducida. Cuando se resuelven los sistemas lineales numéricamente en una computadora, se prefiere la eliminación Gaussiana ya que contiene aproximadamente la mitad de los pasos que el método de Gauss-Jordan.

EL siguiente teorema nos garantiza la existencia de la FER para cualquier matriz.

Teorema 2.26 (Gauss-Jordan) *Para cualquier matriz $A \in M_{m\times n}$, existe una única matriz B tal que B está en FER y B se obtiene a partir de A aplicando operaciones elementales. La matriz B se conoce como "la forma escalonada reducida de A".*

Este teorema asegura, entre otras cosas, que dada la matriz aumentada de un sistema de ecuaciones siempre es posible encontrar el conjunto solución de dicho sistema mediante el uso de operaciones elementales adecuadas. El **método de Gauss-Jordan** consiste en reducir una matriz aumentada a su FER y nos permite determinar la consistencia o inconsistencia de cualquier sistema.

Demostración. Se proporciona un esbozo de la demostración (que es más bien un algoritmo general para obtener la FER) y se dejan los detalles al lector. Sea $A \in M_{m\times n}$, si $A = \Theta$ (la matriz cero), el resultado es obvio pues A es su propia FER. Si $A \neq \Theta$, se hace inducción sobre el número de renglones de A : Si $m = 1$, el resultado es inmediato. Si $m > 1$, entonces tomamos la primera columna (de izquierda a derecha) que es distinta de cero. Supongamos que se trata de la columna A^j, entonces existe alguna entrada $a_{ij} \neq 0$ en A^j. Realizando la operación elemental $\frac{1}{a_{ij}} R_i$ sobre $i - ésimo$ renglón la entrada a_{ij} se transforma en "1". Las entradas $a_{lj} \neq 0$ se convierten en cero realizando operaciones elementales del tipo $-a_{lj} R_i + R_l$. En este punto A^j se ha transformado en la columna

$$i \longrightarrow \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Realizando la operación elemental $R_1 \longleftrightarrow R_i$ la $j - ésima$ columna se transforma en

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Sea $B \in M_{m \times n}$, la matriz que se ha obtenido a partir de A transformando la columna A^j como arriba. Ahora se considera la matriz $\hat{B} \in M_{(m-1) \times n}$ que se obtiene eliminando el primer renglón de B. Por hipótesis de inducción existe la FER de la matriz $\hat{B}$. Realizando las mismas operaciones elementales sobre B que se utilizaron para obtener la FER de $\hat{B}$ obtenemos que la matriz A se ha transformado en una matriz $\hat{A}$ que se visualiza como sigue:

$$\hat{A} = \begin{array}{c} \begin{array}{ccccccc} & & & j & & & \\ & & & \downarrow & & & \end{array} \\ \begin{bmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{bmatrix} \end{array},$$

en donde la matriz que se obtiene eliminando el primer renglón de $\hat{A}$ está en FER. Si la $k - ésima$ columna de esta FER tiene un uno principal en el renglón l y $a_{1k} \neq 0$, se realiza la operación elemental $-a_{1k} R_l + R_1$ sobre la matriz $\hat{A}$. Una vez realizado esto para todos los casos posibles, la matriz resultante será la FER de A. ∎

Observación 2.27 *La FER de cualquier matriz no sólamente existe sino que es única. Este resultado se demostrará más adelante, en la sección 11.*

Ejemplo 2.28 *Utilizando el método de Gauss-Jordan verificaremos que el sistema*

$$\begin{aligned} x + 2y &= 0, \\ -x - 2y &= 1, \end{aligned}$$

es inconsistente. Para ello, basta notar que

$$\left[\begin{array}{cc|c} 1 & 2 & 0 \\ -1 & -2 & 1 \end{array}\right] \overset{1R_1+R_2}{\sim} \left[\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 0 & 1 \end{array}\right].$$

El segundo renglón de la matriz nos dice que $0x + 0y = 1$, *por lo que este sistema no tiene solución.*

Ejemplo 2.29 *Considérese el sistema*

$$\begin{aligned} 3x - y &= a, \\ x - y &= b, \\ 2x + y &= c. \end{aligned}$$

Consideramos su matriz aumentada y obtenemos la FER de la matriz de coeficientes como sigue:

$$\left[\begin{array}{cc|c} 3 & -1 & a \\ 1 & -1 & b \\ 2 & 1 & c \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{cc|c} 1 & -1 & b \\ 3 & -1 & a \\ 2 & 1 & c \end{array}\right] \overset{-3R_1+R_2}{\sim}$$
$$\left[\begin{array}{cc|c} 1 & -1 & b \\ 0 & 2 & a-3b \\ 2 & 1 & c \end{array}\right] \overset{-2R_1+R_3}{\sim} \left[\begin{array}{cc|c} 1 & -1 & b \\ 0 & 2 & a-3b \\ 0 & 3 & c-2b \end{array}\right] \overset{\frac{R_2}{2};\frac{R_3}{3}}{\sim}$$
$$\left[\begin{array}{cc|c} 1 & -1 & b \\ 0 & 1 & \frac{a-3b}{2} \\ 0 & 1 & \frac{c-2b}{3} \end{array}\right] \overset{1R_2+R_1;\,-1R_2+R_3}{\sim} \left[\begin{array}{cc|c} 1 & 0 & b + \frac{a-3b}{2} \\ 0 & 1 & \frac{a-3b}{2} \\ 0 & 0 & \frac{c-2b}{3} - \frac{a-3b}{2} \end{array}\right].$$

Observemos que este sistema será consistente syss

$$\frac{c-2b}{3} - \frac{a-3b}{2} = 0$$

o bien syss

$$3a - 5b - 2c = 0.$$

En este caso se tendrá que

$$x = b + \frac{a-3b}{2}, \;\; y = \frac{a-3b}{2}.$$

El ejemplo anterior nos sugiere el siguiente resultado:

Proposición 2.30 *Si* $A \in M_{m\times n}$ *es la matriz de coeficientes de un sistema de ecuaciones y* $m > n$*, entonces la FER de* A *tiene, al menos un renglón de ceros.*

Demostración. El resultado es consecuencia directa de la definición de la FER, observando que el número de unos principales es, a lo más igual al número de columnas (equivalentemente, variables). Como $m > n$, necesariamente existirán renglones de ceros. ■

En algunas ocasiones los sistemas de ecuaciones no sólo son consistentes (i.e., su conjunto solución no es vacío), sino que tienen además un número infinito de soluciones. Cuando esto ocurre, el método de Gauss Jordan también es capaz de detectar este fenómeno.

Ejemplo 2.31 *Si deseamos encontrar el conjunto solución del sistema*

$$\begin{aligned} x + 2y - 3z &= -4, \\ 2x + y - 3z &= 4, \end{aligned}$$

basta notar que

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & -4 \\ 2 & 1 & -3 & 4 \end{array}\right] \overset{-2R_1+R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 2 & -3 & -4 \\ 0 & -3 & 3 & 12 \end{array}\right] \overset{-\frac{1}{3}R_2}{\sim}$$

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & -4 \\ 0 & 1 & -1 & -4 \end{array}\right] \overset{-2R_2+R_1}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 4 \\ 0 & 1 & -1 & -4 \end{array}\right].$$

De esta forma obtenemos que $x - z = 4$*,* $y - z = -4$ *y que* z *es una variable libre (i.e., puede tomar cualquier valor real). Despejando a* x*,* y *en términos de esta variable libre obtenemos que el conjunto solución del sistema puede expresarse como*

$$\begin{aligned} S &= \{(4 + z, -4 + z, z) \mid z \in \mathbb{R}\} \\ &= \{(4, -4, 0) + (z, z, z) \mid z \in \mathbb{R}\}, \end{aligned}$$

por lo que existe un número infinito de soluciones (tantas como los números reales). Algunas de ellas son

$$\ldots, (4 - 1, -4 - 1, -1),\ (4 + 0, -4 + 0, 0),\ (4 + 1, -4 + 1, 1), \ldots$$

Definición 2.32 *Supongamos que* L *es un sistema con* m *ecuaciones y* n *variables y que* x *es una de estas variables. Decimos que*

- x *es una* ***variable restringida*** *de* L *syss la columna correspondiente a* x *en la FER tiene un uno principal.*
- x *es una* ***variable libre*** *de* L *syss* x *no es una variable restringida de* L*.*

Notamos ahora que el número infinito de soluciones para el sistema

$$\begin{aligned} x + 2y - 3z &= -4, \\ 2x + y - 3z &= 4, \end{aligned}$$

que consideramos en el ejemplo anterior implica un número infinito de soluciones para el sistema homogéneo asociado (el que se obtiene a partir del sistema original igualando todos los términos independientes a cero)

$$\begin{aligned} x + 2y - 3z &= 0, \\ 2x + y - 3z &= 0. \end{aligned}$$

En efecto, el vector $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ es invariante bajo operaciones elementales. Entonces, sin necesidad de volver a realizar las operaciones del ejemplo 2.31, obtenemos que

$$\left[\begin{array}{ccc|c} 1 & 2 & -3 & 0 \\ 2 & 1 & -3 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right].$$

Por lo tanto, la variable z que estaba libre en el conjunto solución S del sistema original vuelve a estar libre en el conjunto solución S_H, del sistema homogéneo asociado, de hecho,

$$S_H = \{(z, z, z) \mid z \in \mathbb{R}\}.$$

Proposición 2.33 *Si un sistema lineal L tiene un número infinito de soluciones, entonces su sistema homogéneo asociado H también tiene un número infinito de soluciones.*

Demostración. Supongamos que nuestro sistema original L tiene m ecuaciones y n variables. Sea $(A|a)$ la matriz aumentada asociada al sistema, donde A y a son, respectivamente, la matriz de coeficientes y el vector de términos independientes de L (en el ejemplo anterior $A = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 1 & -3 \end{bmatrix}$ y $a = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$). Supongamos, además, que $(B|b)$ es la FER de $(A|a)$. Por la hipótesis de infinitud, hay al menos una variable libre en esta FER. Notemos ahora que la matriz aumentada del sistema homogéneo asociado H es $(A|\vec{0})$, donde $\vec{0}$ es el vector cero de $\mathbb{R}^m$. Como $\vec{0}$ es invariante bajo operaciones elementales, la FER de $(A|\vec{0})$ es $(B|\vec{0})$. Por lo tanto, esta segunda FER es consistente y también tiene al menos una variable libre. Esto da lugar a un número infinito de soluciones para H. ■

Dado un sistema lineal, este es inconsistente o bien tiene al menos una solución. En este segundo caso cabe la posibilidad de que existan o no variables libres. Este sencillo razonamiento justifica el siguiente resultado.

Proposición 2.34 *Si L es un sistema lineal, entonces ocurre una y sólo una de las siguientes alternativas:*

(a) L es inconsistente. Esto es, la FER tiene al menos un renglón del tipo

$$[0\ 0, \ldots, 0 \mid c]$$

donde c es un real distinto de cero.

(b) L tiene exáctamente una solución. Esto es, no sucede (a) y la FER de L no tiene variables libres.

(c) L tiene un número infinito de soluciones. Esto es, no suceden ni (a) ni (b) y la FER de L tiene al menos una variable libre.

Proposición 2.35 *Si H es un sistema lineal homogéneo con mas variables que ecuaciones, entonces H tiene un número infinito de soluciones.*

Demostración. Supongamos que H tiene m ecuaciones y n variables. Por hipótesis $m < n$. Como todo sistema homogéneo es consistente, H debe satisfacer alguna de las dos últimas opciones citadas en la proposición anterior. Ahora bien, como

$$\#\text{variables restringidas de } H = \#\text{unos principales} \leq m < n,$$

H tiene al menos una variable libre. Esto da lugar a un número infinito de soluciones. ■

Ejemplo 2.36 *Sin necesidad de cálculos engorrosos, la proposición anterior nos asegura que el sistema homogéneo*

$$\begin{aligned} \pi x + 8{,}9y + 99z &= 0, \\ 26x + 666y - 3.9z &= 0, \end{aligned}$$

tiene un número infinito de soluciones.

La demostración de la siguiente proposición está autocontenida en el enunciado de la misma

Proposición 2.37 *Supongamos que H es un sistema lineal homogéneo con n ecuaciones y n variables. Las siguientes afirmaciones son equivalentes:*

(a) H tiene solución única.

(b) *H no tiene variables libres.*[4]

(c) *La FER de la matriz de coeficientes de H (la cual es una matriz cuadrada) tiene n unos principales.*

(d) *La FER de la matriz de coeficientes de H es la matriz identidad $I \in M_{n\times n}$.*

Ejemplo 2.38 *Es sencillo notar que el sistema homogéneo*

$$\begin{aligned} x - y + z &= 0, \\ 2x - 2y + z &= 0, \\ 3x - 3y + z &= 0, \end{aligned}$$

tiene un número infinito de soluciones. Así pues, la proposición anterior garantiza que la FER de la matriz

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

no es la matriz identidad.

La teoría que hemos aprendido ahora tiene varias aplicaciones operativas. Entre ellas, tenemos al siguiente problema.

Ejemplo 2.39 *Encontrar todos los valores h y k tales que el sistema*

$$\begin{aligned} x_1 - 3x_2 &= 1, \\ 2x_2 + hx_2 &= k. \end{aligned}$$

(a) *No tiene solución.*

(b) *Tiene solución única.*

(c) *Tiene un número infinito de soluciones.*

Para resolver los tres casos primero notamos que

$$\left[\begin{array}{cc|c} 1 & -3 & 1 \\ 2 & h & k \end{array}\right] \overset{-2R_1+R_2}{\sim} \left[\begin{array}{cc|c} 1 & -3 & 1 \\ 0 & h+6 & k-2 \end{array}\right].$$

De acuerdo a la Proposición 2.34, este sistema es inconsistente syss

$$h + 6 = 0 \ \ y \ \ k - 2 \neq 0.$$

Así, la solución para (a) es $h = -6$, $k \neq 2$. Con respecto a (b), tenemos que el sistema es consistente syss sus dos variables están restringidas. Pero, para que $h+6$ pueda "convertirse en un uno" deberíamos de poder utilizar la operación elemental $\frac{1}{h+6}R_2$. Para ello, basta que $h + 6 \neq 0$ (i.e, $h \neq -6$)). Finalmente, notamos que (c) ocurre syss (b) y (a) no ocurren. De esta forma, la solución para (c) es $h = -6$ junto con $k = 2$.

---

[4] Recuérdese que todo sistema homogéneo tiene al menos una solución.

Capítulo 3

El espacio vectorial $\mathbb{R}^n$

Introducción

El método de Gauss Jordan visto en el capítulo anterior proporciona un algoritmo para la resolución de sistemas de ecuaciones lineales. Sin embargo, una pregunta natural es si se trata simplemente de un algoritmo aislado y puramente mecánico o bien si estos sistemas lineales y sus soluciones son parte de un universo más amplio de estructuras lineales. La respuesta va en la segunda dirección y es por eso que ahora se introducirá el concepto de espacio vectorial, que constituye uno de los objetos más fundamentales dentro de la matemática y la física.

En general, a los elementos de un espacio vectorial se les llama vectores, en particular, $\mathbb{R}^n$ es un espacio vectorial por lo cual la Definición 2.1 en el capítulo anterior, en la cual $(x_1, x_{2,} \ldots, x_n) \in \mathbb{R}^n$ es un vector es correcta. La estructura algebraica relevante para los vectores es que éstos pueden sumarse entre si y multiplicarse por elementos de cierto conjunto de números (los números reales en el caso de $\mathbb{R}^n$). Adicionalmente, estas operaciones cumplen ciertas propiedades que daremos más adelante.

En el caso particular de $\mathbb{R}^n$, un vector[1]

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

puede visualizarse geométricamente como una "flecha" que va del origen al punto $(x_1, x_{2,}, \ldots, x_n)$. La Figura 2.1 muestra esta representación para los vectores $\vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ y $\vec{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

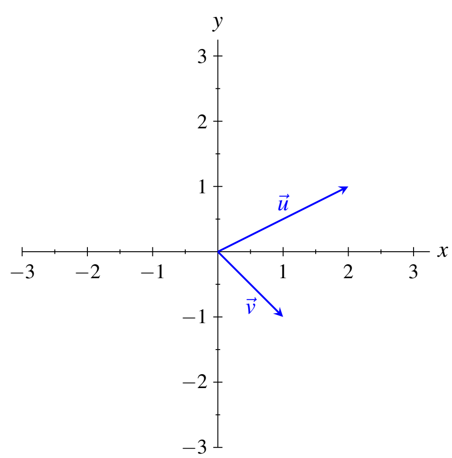

Figura 2.1

En la figuras 2.2 y 2.3 se representan la suma $\vec{u} + \vec{v} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, así como

[1] Por conveniencia utilizaremos vectores columna.

también el producto $2\vec{v} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$.

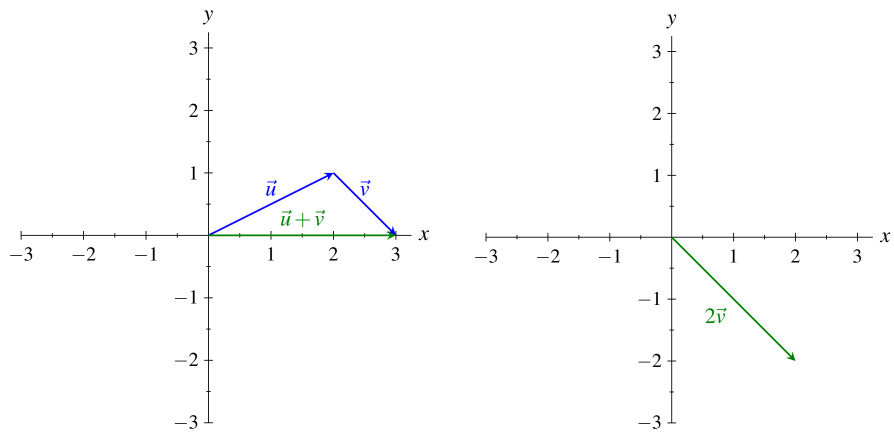

Figura 2.2

Obsérvese que en la Figura 2.2 el punto inicial del vector $\vec{v}$ se traslada al punto $(2,1)$. En la Figura 2.3 se representa el vector vector $\vec{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ en $\mathbb{R}^3$.

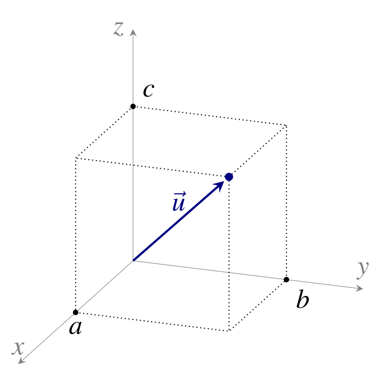

Figura 2.3

Estructura algebraica de $\mathbb{R}^n$

Recordemos que dados

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{y} \quad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

vectores de $\mathbb{R}^n$, decimos que $\vec{u} = \vec{v}$ syss coinciden entrada por entrada. Esto es, $\vec{u} = \vec{v}$ syss $u_i = v_i$ para $i = 1, 2, \ldots, n$. En lo que sigue introduciremos dos operaciones sobre $\mathbb{R}^n$.

Definición 3.1 *Sean*

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad y \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

vectores de $\mathbb{R}^n$ y sea c un número real. La ***suma*** *de $\vec{a}$ y $\vec{b}$ (denotada por $\vec{a} + \vec{b}$) y el* ***producto*** *del escalar c con $\vec{a}$ (denotado por $c\vec{a}$) se definen como*

$$\vec{a} + \vec{b} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} \quad y \quad c\vec{a} = \begin{bmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{bmatrix},$$

es decir, las operaciones se realizan "coordenada por coordenada".

Observación 3.2 *Se hace énfasis en que el producto que se ha definido, no es entre dos vectores, sino entre un escalar $c \in \mathbb{R}$ y un vector $\vec{a} \in \mathbb{R}^n$.*

Notación 3.1 *En el contexto de $\mathbb{R}^n$ utilizaremos las siguientes convenciones.*

- *El vector cero (aquel cuyas componentes son todas cero) es denotado por $\vec{0}$.*
- *$-\vec{a}$ denota al vector $(-1)\vec{a}$.*
- *$\vec{a} - \vec{b}$ significa $\vec{a} + (-\vec{b})$.*

- *A partir de ahora la palabra* ***escalar*** *será sinónimo de número real, o bien,* c es un escalar syss $c \in \mathbb{R}$.

Ejemplo 3.3 *Si* $\vec{a} = \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$ *y* $\vec{b} = \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}$ *son vectores en* $\mathbb{R}^3$*, entonces*

$$2\vec{a} - 3\vec{b} = \begin{bmatrix} -6 \\ 8 \\ 2 \end{bmatrix} + \begin{bmatrix} -9 \\ -15 \\ 0 \end{bmatrix} = \begin{bmatrix} -15 \\ -7 \\ 2 \end{bmatrix}.$$

Ejemplo 3.4 *Si* $\vec{a} = \begin{bmatrix} -1 \\ 7 \end{bmatrix}, \vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ *son vectores en* $\mathbb{R}^2$*,entonces* $\vec{a} + \vec{0} = \vec{a}$.

Ejemplo 3.5 *Geométricamente, la suma puede visualizarse con el llamado método del paralelogramo como se muestra en la figura 2.4.*

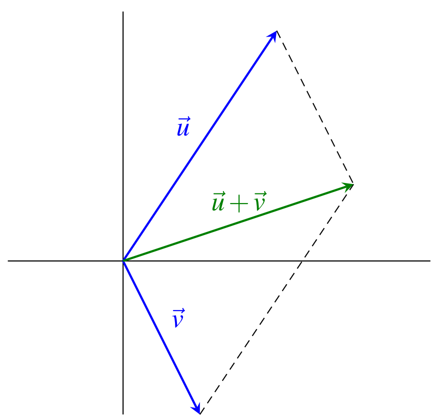

Figura 2.4: Suma de vectores

La siguiente proposición es sencilla (pero tediosa) de demostrar apelando a las propiedades básicas de $\mathbb{R}$.

Proposición 3.6 $\mathbb{R}^n$ *junto con las operaciones de suma entre vectores y de producto escalar que acabamos de definir satisface que:*

V1 *Si* $\vec{u}, \vec{v} \in \mathbb{R}^n$*, entonces* $\vec{u}+\vec{v} = \vec{v}+\vec{u}$ *(conmutatividad de la suma).*

V2 *Si* $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$*, entonces* $\vec{u}+(\vec{v}+\vec{w}) = (\vec{u}+\vec{v})+\vec{w}$ *(asociatividad de la suma).*

V3 *Si* $\vec{u} \in \mathbb{R}^n$*, entonces* $\vec{u} + \vec{0} = \vec{u}$ *(existencia de neutro aditivo).*

V4 *Si* $\vec{u} \in \mathbb{R}^n$*, entonces* $\vec{u} - \vec{u} = \vec{0}$ *(existencia de inversos aditivos).*

V5 Si $\vec{u} \in \mathbb{R}^n$, entonces $1\vec{u} = \vec{u}$.

V6 Si $\vec{u} \in \mathbb{R}^n$ y β y γ son escalares, entonces $(\beta\gamma)\vec{u} = \beta(\gamma\vec{u})$.

V7 Si $\vec{u} \in \mathbb{R}^n$ y β y γ son escalares, entonces $(\beta+\gamma)\vec{u} = \beta\vec{u} + \gamma\vec{u}$ (primera forma de distributividad).

V8 Si $\vec{u}, \vec{v} \in \mathbb{R}^n$ y β es un escalar, entonces $\beta(\vec{u}+\vec{v}) = \beta\vec{u} + \beta\vec{v}$ (segunda forma de distributividad).

Demostración. A manera de ejemplo probamos $(V3)$ y $(V7)$. En el caso de $(V3)$ notamos que

$$\vec{u} + \vec{0} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} u_1 + 0 \\ u_2 + 0 \\ \vdots \\ u_n + 0 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \vec{u}$$

De forma similar, en el caso de $(V7)$ tenemos que

$$(\beta+\gamma)\vec{u} = \begin{bmatrix} (\beta+\gamma)u_1 \\ (\beta+\gamma)u_2 \\ \vdots \\ (\beta+\gamma)u_n \end{bmatrix} = \begin{bmatrix} \beta u_1 + \gamma u_1 \\ \beta u_2 + \gamma u_2 \\ \vdots \\ \beta u_n + \gamma u_n \end{bmatrix}$$

$$= \begin{bmatrix} \beta u_1 \\ \beta u_2 \\ \vdots \\ \beta u_n \end{bmatrix} + \begin{bmatrix} \gamma u_1 \\ \gamma u_2 \\ \vdots \\ \gamma u_n \end{bmatrix} = \beta \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \gamma \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \beta\vec{u} + \gamma\vec{u}$$

■

A partir de estas ocho propiedades, conocidas como **propiedades de espacio vectorial**, se deducen otros resultados cada vez más complejos. A continuación se proporcionan algunos ejemplos.

Proposición 3.7 (Ley de la cancelación) *Si $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ y $\vec{a} + \vec{b} = \vec{a} + \vec{c}$, entonces $\vec{b} = \vec{c}$.*

Demostración. Supongamos $\vec{a} + \vec{b} = \vec{a} + \vec{c}$. Sumando $-\vec{a}$ a ambos lados de esta igualdad y aplicando sucesivamentelas propiedades de asociatividad, del inverso y del neutro aditivo, obtenemos

$$\begin{gathered} -\vec{a} + (\vec{a} + \vec{b}) = -\vec{a} + (\vec{a} + \vec{c}) \text{ syss} \\ (-\vec{a} + \vec{a}) + \vec{b} = (-\vec{a} + \vec{a}) + \vec{c} \text{ syss} \\ \vec{0} + \vec{b} = \vec{0} + \vec{c} \text{ syss} \\ \vec{b} = \vec{c}. \end{gathered}$$

■

Corolario 3.8 *Si* $\vec{a}, \vec{b} \in \mathbb{R}^n$ *y* $\vec{a} + \vec{b} = \vec{a}$*, entonces* $\vec{b} = \vec{0}$*.*

Demostración. Por hipótesis tenemos que $\vec{a} + \vec{b} = \vec{a}$ o lo que es lo mismo, $\vec{a}+\vec{b} = \vec{a}+\vec{0}$. Por la ley de la cancelación concluimos que $\vec{b} = \vec{0}$. ■

Proposición 3.9 *Si* $\vec{u} \in \mathbb{R}^n$ *y* β *es un escalar, entonces* $\beta\vec{u} = \vec{0}$ *syss* $\beta = 0$ *o* $\vec{u} = \vec{0}$*.*

Demostración. Si $\beta = 0$ o $\vec{u} = \vec{0}$, entonces es claro que $\beta\vec{u} = \vec{0}$. Supongamos ahora que

$$\beta\vec{u} = \vec{0}; \tag{$\bigstar$}$$

debemos demostar que $\beta = 0$ o $\vec{u} = \vec{0}$. Si $\beta = 0$, la demostración queda terminada. Supongamos, entonces, que $\beta \neq 0$ (y, por lo tanto, tiene sentido hablar de $\frac{1}{\beta}$). Multiplicando ambos lados de ($\bigstar$) por $\frac{1}{\beta}$ obtenemos que

$$\vec{u} = 1\vec{u} = (\frac{1}{\beta}\beta)\vec{u} = \frac{1}{\beta}(\beta\vec{u}) = \frac{1}{\beta}\vec{0} = \vec{0}.$$

■

Producto punto o escalar

Dados dos vectores $\vec{x}$ y $\vec{y}$ en $\mathbb{R}^n$, el **producto escalar** o **producto punto** entre ellos se define como sigue:

Definición 3.10 *Sean*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \ \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n.$$

El ***producto punto (o producto escalar)*** *entre* $\vec{x}$ *y* $\vec{y}$ *se define como*[2]

$$\vec{x} \cdot \vec{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i.$$

[2] Alternativamente se utiliza la notación

$$\langle \vec{x}, \vec{y} \rangle$$

para el producto escalar de los vectores $\vec{x}$ y $\vec{y}$.

Observemos que este producto asocia a cada pareja de vectores de $\mathbb{R}^n$ un escalar, es decir, un número real.

El producto punto cumple las siguientes propiedades para $\vec{x}$, $\vec{y}$, $\vec{z}$ en $\mathbb{R}^n$ y $r \in \mathbb{R}$. Las demostraciones de todas ellas son inmediatas a partir de la definición y se dejan como ejercicio al lector.

P1 $\vec{x} \cdot \vec{x} = 0$ syss $\vec{x} = \vec{0}$.

P2 $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$.

P3 $\vec{x} \cdot (\vec{y} + \vec{z}\ ) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}$.

P4 $(r\vec{x}) \cdot \vec{y} = r(\vec{x} \cdot \vec{y}) = \vec{x} \cdot (r\vec{y})$.

Ejemplo 3.11 *Dados*

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \in \mathbb{R}^2,$$

se tiene que

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 1 \times 3 + (-1) \times 2 = 1.$$

Ejemplo 3.12

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \in \mathbb{R}^3,$$

se tiene que

$$\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} = 5 - 5 + 0 = 0,$$

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

por lo tanto

$$\begin{aligned}
&\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \left(2 \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right) \\
&= \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \left(2 \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} \right) + \left(\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right) \\
&= 2 \left(\begin{matrix} 1 \\ -1 \\ 2 \end{matrix} \cdot \begin{bmatrix} 5 \\ 5 \\ 0 \end{bmatrix} \right) + \left(\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right) \\
&= 2 \times 0 + 4 = 4.
\end{aligned}$$

Observación 3.13 *Sea* $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ *y sea* $\vec{x}^{\perp} = \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix}$ *un vector perpendicular (u ortogonal) a* $\vec{x}$*. Observamos que*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} = x_1x_2 + x_2(-x_1) = 0.$$

El producto punto será utilizado más adelante para generalizar la noción de ángulo recto a espacios vectoriales más generales.

Combinaciones lineales

Definición 3.14 *Sean* $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m, \vec{v}$ *vectores de* $\mathbb{R}^n$*. Decimos que el vector* $\vec{v}$ ***es una combinación lineal de los vectores*** $\vec{u}_1, \vec{u}_2, \ldots \vec{u}_m$*, syss existen escalares* $\beta_1, \beta_2, \ldots, \beta_m$ *tales que*

$$\vec{v} = \beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \ldots + \beta_m\vec{u}_m.$$

Podríamos expresar la suma anterior en forma compacta como

$$\vec{v} = \sum_{i=1}^{m} \beta_i\vec{u}_i,$$

sin embargo, para mayor claridad en este libro se preferirá el uso de la notación extendida.

Ejemplo 3.15 *Consideremos los vectores*

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \vec{u_1} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \vec{u_2} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \end{bmatrix}$$

en $\mathbb{R}^3$*. Claramente,*

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = -2 \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \end{bmatrix} = -2\vec{u}_1 + 2\vec{u}_2,$$

por lo que $\vec{v}$ *es una combinación lineal de* $\vec{u}_1$ *y* $\vec{u}_2$*.*

Ejemplo 3.16 *Consideremos los vectores* $\vec{v}, \vec{u}_1, \vec{u}_2$ *del ejemplo anterior junto con el vector*

$$\vec{u}_3 = \begin{bmatrix} 666 \\ \pi \\ -765.8 \end{bmatrix}$$

Como $v = -2\vec{u}_1 + 2\vec{u}_2$*, es obvio que* v *es combinación lineal de* $\vec{u}_1$*,* $\vec{u}_2$*,* $\vec{u}_3$ *(pues* $v = -2\vec{u}_1 + 2\vec{u}_2 + 0\vec{u}_3$*).*

Definición 3.17 *Sean $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ vectores de $\mathbb{R}^n$. Definimos el* ***conjunto generado por*** *$\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$, denotado por span$\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$, como el conjunto de todas las combinaciones lineales de estos m vectores. Esto es,*

$$span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\} = \{\beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \ldots + \beta_m\vec{u}_m \mid \beta_1, \beta_2, \ldots, \beta_m \in \mathbb{R}\}.$$

Ejemplo 3.18 *Sea $\vec{v} \in \mathbb{R}^n$, entonces, span$\{\vec{v}\} = \{\beta\vec{v} \in \mathbb{R}^n \mid \beta \in \mathbb{R}\}$. Si $\vec{v} = \vec{0}$, este conjunto contiene únicamente al vector $\vec{0}$. Si $\vec{v} \neq \vec{0}$, Podemos visualizar a span$\{\vec{v}\}$ como la recta infinita en $\mathbb{R}^n$ (que pasa por el origen) en la dirección de $\vec{v}$ (ver figura 2.5)*

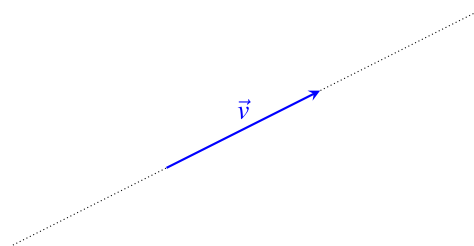

Figura 2.5

Observación 3.19 *El ejemplo anterior junto con la interpretación geométrica de la suma de dos vectores proporcionada en el ejemplo (3.5), permiten visualizar a las combinaciones lineales de uno y dos vectores como rectas y planos, respectivamente. Lo relevante es que se trata de estructuras lineales o planas. En general, las combinaciones lineales de $n-1$ vectores $(n > 3)$ originan también estructuras lineales en $\mathbb{R}^n$ llamadas* ***hiperplanos****.*

Ejemplo 3.20 *Consideremos, de nuevo, los vectores $\vec{v}, \vec{u}_1, \vec{u}_2 \in \mathbb{R}^3$ del Ejemplo 3.15. Ya sabemos que $\vec{v} \in$ span$\{\vec{u}_1, \vec{u}_2\}$ (pues $v = -2\vec{u}_1 + 2\vec{u}_2$) y que $\vec{0} \in$ span$\{\vec{u}_1, \vec{u}_2\}$ (pues $\vec{0} = 0\vec{u}_1 + 0\vec{u}_2$). Ahora nos gustaría describir, en general, span$\{\vec{u}_1, \vec{u}_2\}$. Para ello, notamos que un vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix} \in$ span$\{\vec{u}_1, \vec{u}_2\}$ syss existen reales β_1, β_2 tales que*

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \beta_1\vec{u}_1 + \beta_2\vec{u}_2 = \begin{bmatrix} 0 \\ -\beta_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{\beta_2}{2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 + \frac{1}{2}\beta_2 \\ -\beta_1 + 0 \\ 0 + 0 \end{bmatrix}.$$

Como dos vectores son iguales syss coinciden componente a componente, esto da lugar a un sistema de ecuaciones en el cual las variables *son los escalares x_1 y x_2. La matriz aumentada y su FER se obtienen como*

$$\left[\begin{array}{cc|c} 0 & \frac{1}{2} & a \\ -1 & 0 & b \\ 0 & 0 & c \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{cc|c} -1 & 0 & b \\ 0 & \frac{1}{2} & a \\ 0 & 0 & c \end{array}\right] \overset{-R_1}{\sim}$$
$$\left[\begin{array}{cc|c} 1 & 0 & -b \\ 0 & \frac{1}{2} & a \\ 0 & 0 & c \end{array}\right] \overset{2R_2}{\sim} \left[\begin{array}{cc|c} 1 & 0 & -b \\ 0 & 1 & 2a \\ 0 & 0 & c \end{array}\right].$$

Por lo tanto, el sistema es consistente (i.e., existen tales β_1 y β_2) syss $c = 0$ y

$$span\{\vec{u}_1, \vec{u}_2\} = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3 \mid c = 0 \right\} = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid a, b \in \mathbb{R} \right\}.$$

Además,

$$\begin{bmatrix} a \\ b \\ 0 \end{bmatrix} = \beta_1 \vec{u}_1 + \beta_2 \vec{u}_2 \quad syss \quad \beta_1 = -b, \beta_2 = 2a$$

Así, en el caso concreto de $\begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}$ *tenemos que* $\begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix} = -4u_1 - 6u_2$.

Observemos que si tomamos los vectores $u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $u_4 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, *entonces* $span\{\vec{u}_1, \vec{u}_2\} = span\{\vec{u}_3, \vec{u}_4\}$. *En efecto, si*

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \beta_1 \vec{u}_3 + \beta_2 \vec{u}_4 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ 0 \end{bmatrix},$$

Se tiene que la FER del sistema asociado es simplemente

$$\left[\begin{array}{cc|c} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & c \end{array}\right].$$

De aquí que los escalares β_1 y β_2 existen siempre y cuando $c = 0$ y

$$span\{\vec{u}_3, \vec{u}_4\} = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid a, b \in \mathbb{R} \right\} = span\{\vec{u}_1, \vec{u}_2\},$$

que puede expresarse como

$$\left\{ a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid a, b \in \mathbb{R} \right\}.$$

Geométricamente, las combinaciones lineales de ambos conjuntos de vectores: span$\{\vec{u}_1, \vec{u}_2\}$ y span$\{\vec{u}_3, \vec{u}_4\}$, generan el mismo subconjunto de vectores en el espacio, en este caso el plano xy, es decir el conjunto de todas las combinaciones lineales de los vectores

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad y \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

De esta forma, span$\{\vec{u}_1, \vec{u}_2\}$ = span$\{\vec{u}_3, \vec{u}_4\}$ = span$\{\vec{e}_1, \vec{e}_2\}$.

Ejemplo 3.21 *En general, dados dos vectores $\vec{u}$ y $\vec{v}$ (digamos en $\mathbb{R}^2$ o $\mathbb{R}^3$), tenemos que span$\{\vec{u}, \vec{v}\}$ puede ser una recta o un plano, dependiendo de si los vectores son LD o LI. Esto se ilustra en la Figura 2.6.*

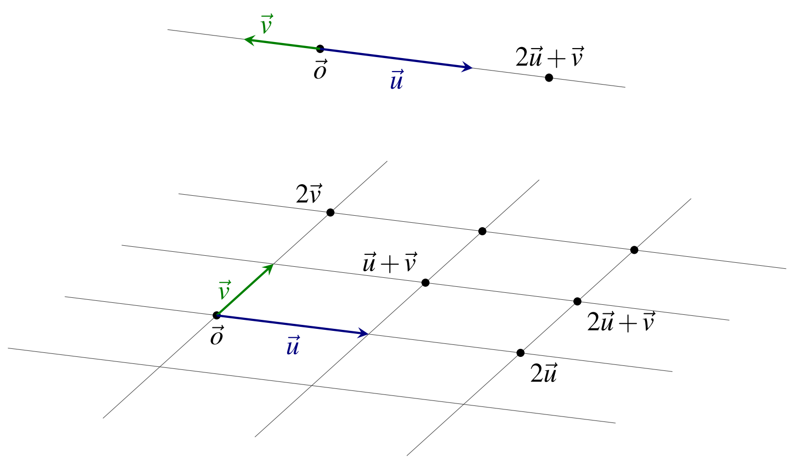

Figura 2.6

Ejemplo 3.22 *Consideremos ahora a los vectores*

$$\vec{v_1} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}$$

y supongamos que deseamos describir a los vectores del conjunto span$\{\vec{v}_1, \vec{v}_2\}$.

Para ello notamos que, por definición,

$$span\{\vec{v_1},\vec{v_2}\} = \left\{ x_1 \begin{bmatrix}1\\2\\0\end{bmatrix} + x_2 \begin{bmatrix}0\\-2\\1\end{bmatrix} \mid x_1,x_2 \in \mathbb{R} \right\}$$
$$= \left\{ \begin{bmatrix}x_1\\2x_1-2x_2\\x_2\end{bmatrix} \mid x_1,x_2 \in \mathbb{R} \right\}.$$

Por lo tanto,

$$\begin{bmatrix}a\\b\\c\end{bmatrix} \in span\{\vec{v}_1,\vec{v}_2\}$$

syss existen $x_1, x_2 \in \mathbb{R}$ *tales que*

$$\begin{bmatrix}a\\b\\c\end{bmatrix} = x_1 \begin{bmatrix}1\\2\\0\end{bmatrix} + x_2 \begin{bmatrix}0\\-2\\1\end{bmatrix} = \begin{bmatrix}x_1\\2x_1-2x_2\\x_2\end{bmatrix}.$$

Como dos vectores son iguales syss coinciden componente a componente, se tiene el siguiente sistema de ecuaciones

$$\begin{aligned} x_1 &= a,\\ 2x_1-2x_2 &= b,\\ x_2 &= c, \end{aligned}$$

cuya matriz aumentada es

$$\left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & 2 & b\\ 1 & -1 & c \end{array}\right].$$

Obtenemos la FER como sigue.

$$\left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & 2 & b\\ 1 & -1 & c \end{array}\right] \overset{-R_1+R_3}{\sim} \left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & 2 & b\\ 0 & -1 & c-a \end{array}\right]$$

$$\overset{R_2\leftrightarrow R_3}{\sim} \left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & -1 & c-a\\ 0 & 2 & b \end{array}\right] \overset{2R_2+R_3}{\sim} \left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & -1 & c-a\\ 0 & 0 & b+2c-2a \end{array}\right] \overset{-R_2}{\sim}$$
$$\left[\begin{array}{cc|c} 1 & 0 & a\\ 0 & 1 & a-c\\ 0 & 0 & b+2c-2a \end{array}\right].$$

Por lo tanto, el sistema es consistente (i.e., existen tales x_1 y x_2) syss

$$b + 2c - 2a = 0.$$

Esto es, syss

$$b = 2a - 2c.$$

En tal caso, tenemos que la matriz anterior está dada como

$$\left[\begin{array}{cc|c} 1 & 0 & a \\ 0 & 1 & a-c \\ 0 & 0 & 0 \end{array}\right].$$

De aquí se obtiene que $x_1 = a, x_2 = a - c$. Así, en el caso concreto del vector

$$\vec{w} = \begin{bmatrix} 8 \\ 14 \\ 1 \end{bmatrix},$$

tenemos que $\vec{w} \in span\{\vec{v}_1, \vec{v}_2\}$ puesto que $2(8) - 2(1) = 14$ y podemos expresar a $\vec{w}$ como

$$\vec{w} = 8\vec{v}_1 + (8-1)\vec{v}_2 = 8\vec{v}_1 + 7\vec{v}_2.$$

Finalizamos esta sección con tres resultados bastante intuitivos acerca del comportamiento del *span* de un conjunto de vectores.

Proposición 3.23 *Si $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ son vectores de $\mathbb{R}^n$, entonces $\vec{0} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$.*

Demostración. Basta notar que $\vec{0} = 0\vec{u}_1 + 0\vec{u}_2 + \ldots + 0\vec{u}_m$. ■

Proposición 3.24 *Si $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}$ son vectores de $\mathbb{R}^n$, entonces*

$$span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} \subset span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}.$$

Demostración. Sea $\vec{v} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$; debemos probar que

$$\vec{v} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}.$$

Pero, por definición, existen escalares $\beta_1, \beta_2, \ldots, \beta_p$ tales que $v = \beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \ldots + \beta_p\vec{u}_p$. Por lo tanto,

$$v = \beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \ldots + \beta_p\vec{u}_p + 0\vec{u}_{p+1} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}.$$

■

Antes de pasar al siguiente resultado recordemos que dos conjuntos M y N son iguales syss tienen exáctamente los mismos elementos. Así, para demostrar que $M = N$, basta demostrar que todos los elementos de M son elementos de N ($M \subset N$) y que todos los elementos de N son elementos de M ($N \subset M$).

Ejemplo 3.25 *Sean $\vec{v}_1, \vec{v}_2$ y $\vec{v}_3$ vectores en $\mathbb{R}^n$ y $\vec{v}_3 \in span\{\vec{v}_1, \vec{v}_2\}$. Entonces se tiene que*

$$span\{\vec{v}_1, \vec{v}_2\} = span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}.$$

En este ejemplo hay que probar la igualdad entre dos conjuntos. La contención $span\{\vec{v}_1, \vec{v}_2\} \subset span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ es inmediata por la proposición anterior. Para probar la otra contención, tomamos cualquier elemento $\vec{w} \in span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. Por definición, tenemos que existen números reales (o escalares) $\gamma_1, \gamma_2, \gamma_3$ tales que

$$\vec{w} = \gamma_1\vec{v}_1 + \gamma_2\vec{v}_2 + \gamma_3\vec{v}_3. \tag{3.1}$$

Por hipótesis, $\vec{v}_3 \in span\{\vec{v}_1, \vec{v}_2\}$ y existen $\beta_1, \beta_2 \in \mathbb{R}$ tales que, $v_3 = \beta_1\vec{v}_1 + \beta_2\vec{v}_2$. Sustituyendo esta expresión para v_3 en (3.1) se obtiene

$$\vec{w} = \gamma_1\vec{v}_1 + \gamma_2\vec{v}_2 + \gamma_3\left(\beta_1\vec{v}_1 + \beta_2\vec{v}_2\right)$$

y podemos reescribir esta igualdad como

$$\vec{w} = \left(\gamma_1 + \gamma_3\beta_1\right)\vec{v}_1 + \left(\gamma_2 + \gamma_3\beta_2\right)\vec{v}_2,$$

es decir, $\vec{w} \in span\{\vec{v}_1, \vec{v}_2\}$ y se tiene que $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subset span\{\vec{v}_1, \vec{v}_2\}$ con lo cual concluye la demostración.

A continuación enunciamos y demostramos el caso general.

Proposición 3.26 *Si $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}$ son vectores de $\mathbb{R}^n$ y $\vec{u}_{p+1} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$, entonces*

$$span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} = span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}.$$

Demostración. Gracias al resultado anterior sabemos que

$$span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} \subset span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}.$$

Por lo tanto, resta verificar que

$$span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\} \subset span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$$

Por hipótesis, existen escalares $\beta_1, \beta_2, \ldots, \beta_p$ tales que

$$\vec{u}_{p+1} = \beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \cdots + \beta_p\vec{u}_p.$$

Sea ahora $\vec{v} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p, \vec{u}_{p+1}\}$; debemos probar que $\vec{v} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$. Pero, por definición, existen escalares $\gamma_1, \gamma_2, \ldots, \gamma_{p+1}$ tales que

$$\vec{v} = \gamma_1\vec{u}_1 + \gamma_2\vec{u}_2 + \cdots + \gamma_{p+1}\vec{u}_{p+1}$$

Por lo tanto,

$$\vec{v} = \gamma_1\vec{u}_1 + \gamma_2\vec{u}_2 + \cdots + \gamma_{p+1}(\underbrace{\beta_1\vec{u}_1 + \beta_2\vec{u}_2 + \cdots + \beta_p\vec{u}_p}_{\vec{u}_{p+1}}) =$$

$$(\gamma_1 + \gamma_{p+1}\beta_1)\vec{u}_1 + (\gamma_2 + \gamma_{p+1}\beta_2)\vec{u}_2 + \cdots + (\gamma_p + \gamma_{p+1}\beta_p)\vec{u}_p.$$

Concluimos pues, que $\vec{v} \in span\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$. ■

Algunas representaciones útiles

Consideremos el siguiente sistema con m ecuaciones y n variables.

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases}$$

Como siempre, denotamos a la matriz correspondiente de coeficientes por A, al vector de términos independientes por $\vec{b}$ y por S al conjunto solución del sistema lineal. Notamos que $(t_1, t_2, \ldots, t_n) \in S$ syss

$$\begin{bmatrix} a_{11}t_1 + a_{12}t_2 + \ldots + a_{1n}t_n \\ a_{21}t_1 + a_{22}t_2 + \ldots + a_{2n}t_n \\ \vdots \\ a_{m1}t_1 + a_{m2}t_2 + \ldots + a_{mn}t_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Esto es, syss

$$\begin{bmatrix} t_1a_{11} \\ t_1a_{21} \\ \vdots \\ t_1a_{m1} \end{bmatrix} + \begin{bmatrix} t_2a_{12} \\ t_2a_{22} \\ \vdots \\ t_2a_{m2} \end{bmatrix} + \ldots + \begin{bmatrix} t_na_{1n} \\ t_na_{2n} \\ \vdots \\ t_na_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

O lo que es lo mismo,

$$t_1A^1 + t_2A^2 + \ldots t_nA^n = \vec{b}.$$

De esta forma, hemos probado el siguiente resultado.

Proposición 3.27 *$(t_1, t_2, \ldots, t_n)$ es solución del sistema cuya matriz aumentada es $(A|\vec{b})$ syss*

$$t_1A^1 + t_2A^2 + \ldots + t_nA^n = \vec{b}.$$

En particular, dicho sistema es consistente syss $\vec{b} \in span\{A^1, A^2, \ldots, A^n\}$.

Definición 3.28 (Producto de una matriz por un vector) *Sea $A \in M_{m\times n}$ y sea*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

un vector de $\mathbb{R}^n$, pensado como una matriz de $n \times 1$.[3] *Definimos el producto de A y $\vec{x}$, denotado por $A\vec{x}$ como*

$$A\vec{x} = x_1 A^1 + x_2 A^2 + \ldots + x_n A^n.$$

Ejemplo 3.29 *Si*

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 6 \\ 4 & 0 \end{bmatrix}, \ \vec{x} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \ \vec{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad y \quad \vec{z} = \begin{bmatrix} 7 \\ 0 \\ 5 \end{bmatrix},$$

entonces

$$A\vec{x} = -1 \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 10 \\ -4 \end{bmatrix},$$

$$A\vec{y} = 0 \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

y $A\vec{z}$ no está definido.

Utilizando esta nueva operación, podemos traducir la Proposición 3.27 como sigue.

Corolario 3.30 *Supongamos que*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

es un vector de $\mathbb{R}^n$, pensado como una matriz de $n \times 1$. Las siguientes afirmaciones son equivalentes:

1. *$\vec{x}$ es solución del sistema lineal cuya matriz aumentada es $(A|\vec{b})$.*
2. *$A\vec{x} = \vec{b}$.*
3. *$x_1 A^1 + x_2 A^2 + \ldots + x_n A^n = \vec{b}$*

Así pues, el sistema de ecuaciones

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

[3] Notar que n es el tamaño del vector $\vec{x}$ y el número de columnas de la matriz A.

también puede reescribirse en forma matricial como

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

o alternativamente en forma vectorial como

$$x_1 A^1 + x_2 A^2 + \cdots + x_n A^n = \vec{b},$$

en donde $A^j \in \mathbb{R}^m$.

Ejemplo 3.31 *El sistema de ecuaciones*

$$\begin{aligned} 3x_2 + 6x_3 &= 3, \\ x_1 + x_2 + x_3 &= 0, \end{aligned}$$

puede representarse como

$$x_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

o bien, como

$$\begin{bmatrix} 0 & 3 & 6 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

En lo que sigue sistematizaremos algunos de los pasos que empleamos en los Ejemplos 3.20 y 3.22.

El siguiente corolario es una reinterpretación del Corolario 3.30.

Corolario 3.32 *Supongamos que*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

es un vector de $\mathbb{R}^n$*, pensado como una matriz de* $n \times 1$*, que* $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n, \vec{v}$ *son vectores de* $\mathbb{R}^m$ *y que* $B \in M_{m\times n}$ *es la matriz cuya columna* B^j *coincide con* $\vec{u}_j$ *. Las siguientes afirmaciones son equivalentes:*

(a) $\vec{x}$ *es solución del sistema cuya matriz aumentada es* $(B|\vec{v})$*.*

(b) $x_1\vec{u}_1 + x_2\vec{u}_2 + \ldots + x_n\vec{u}_n = \vec{v}$*.*

Demostración. Gracias al Corolario 3.30 tenemos que $\vec{x}$ es solución del sistema $B\vec{x} = \vec{v}$ (cuya matriz aumentada es $(B|\vec{v})$) syss

$$\begin{gathered} x_1\vec{u}_1 + x_2\vec{u}_2 + \ldots + x_n\vec{u}_n = \\ x_1B^1 + x_2B^2 + \ldots + x_nB^n = \vec{v}. \end{gathered}$$

■

Ejemplo 3.33 *Encontrar todos los pares ordenados $(c_1, c_2) \in \mathbb{R}^2$ tales que*

$$c_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$

es lo mismo que encontrar todos los pares $(c_1, c_2) \in \mathbb{R}^2$ que satisfacen

$$\begin{bmatrix} 1 & 0 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}.$$

Pero,

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 4 & 3 & 7 \end{array}\right] \overset{-4R_1+R_2}{\sim} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 3 & 3 \end{array}\right] \overset{\frac{1}{3}R_2}{\sim} \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right],$$

de manera que nuestro conjunto solución es $S = \{(1, 1)\}$.

Ejemplo 3.34 *Supongamos que se nos pide determinar si $\vec{w}$ es combinación lineal de $\vec{v}_1, \vec{v}_2, \vec{v}_3$, donde*

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Una forma sería resolver el sistema

$$x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

encontrando la FER de la matriz aumentada del sistema como sigue

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 1 & -1 & 2 & 2 \end{array}\right] \overset{-R_1+R_2}{\sim} \left[\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

y notando que este sistema es inconsistente. Por lo tanto, $\vec{w} \notin span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

Alternativamente, podríamos notar que $\vec{v}_2, \vec{v}_3 \in span\{\vec{v}_1\}$ y por la Proposición 3.26 se tiene que

$$span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{v}_1\}.$$

Pero

$$span\{\vec{v}_1\} = \left\{ \beta \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2 \mid \beta \in \mathbb{R} \right\} \subsetneq \mathbb{R}^2$$

y, claramente

$$\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \notin span\{\vec{v}_1\} = span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}.$$

Observemos que geométricamente, $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *es simplemente la recta* $y = x$ *en el plano* xy.

El símbolo $\subsetneq$ significa "contención propia". Esto es, $A \subsetneq B$ quiere decir que A está contenido en B, pero que B contiene elementos que no están en A. El ejemplo anterior muestra que si $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ son vectores de $\mathbb{R}^n$ con $p \geq n$, entonces no es necesariamente cierto que $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = \mathbb{R}^n$ (hay algunos casos donde esta igualdad es verdadera y otros, como el que acabamos de ver, donde es falsa). ¿Cuántos vectores son necesarios para generar todo $\mathbb{R}^n$? ¿Qué deben satisfacer estos vectores? Estas preguntas y otras similares serán abordadas en el capítulo de Bases y Dimensión. Para este efecto, es necesario el concepto de independencia lineal que exponemos en el siguiente capítulo.

Capítulo 4

Independencia Lineal

Introducción

En el capítulo anterior definimos el producto de una matriz $A \in M_{m\times n}$ y de un vector $\vec{t} \in \mathbb{R}^n$ (donde $\vec{t}$ es pensado como una matriz de $n \times 1$) como

$$A\vec{t} = t_1A^1 + t_2A^2 + \ldots + t_nA^n.$$

Este producto es, por definición, una combinación lineal de las columnas A^j de la matriz[1] A y sirve, entre otras cosas, para proporcionar una forma alternativa de representar sistemas de ecuaciones lineales.

En este capítulo responderemos a la siguiente pregunta: **Dado un sistema lineal consistente, ¿bajo qué condiciones podemos asegurar que hay exáctamente una solución?** Antes de intentar contestarla, analicemos los siguientes ejemplos.

Ejemplo 4.1 *Consideremos el sistema lineal L dado por*

$$\begin{aligned} 3x_1 - x_2 &= 6, \\ x_1 + x_2 + x_3 &= 1, \end{aligned}$$

[1]Por ejemplo, $\begin{bmatrix} 1 & 0 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 4 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 17 \end{bmatrix} \in span\left\{\begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}\right\}$.

mismo que puede expresarse en forma matricial como

$$\begin{bmatrix} 3 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}.$$

Supongamos que deseamos, primero encontrar su conjunto solución y segundo, determinar de cuantas formas distintas podemos expresar al vector $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ *como combinación lineal de los vectores*

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad y \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

es decir, de los vectores columna de la matriz del sistema. Obtenemos la FER de la matriz aumentada del sistema como sigue:

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & 6 \\ 1 & 1 & 1 & 1 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 3 & -1 & 0 & 6 \end{array}\right] \overset{-3R_1+R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & -4 & -3 & 3 \end{array}\right]$$

$$\overset{-\frac{1}{4}R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & \frac{3}{4} & -\frac{3}{4} \end{array}\right] \overset{-R_2+R_1}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{4} & \frac{7}{4} \\ 0 & 1 & \frac{3}{4} & -\frac{3}{4} \end{array}\right].$$

De aquí que el conjunto solución está dado por

$$S_L = \left\{ \begin{bmatrix} \frac{7}{4} - \frac{1}{4}x_3 \\ -\frac{3}{4} - \frac{3}{4}x_3 \\ x_3 \end{bmatrix} \in \mathbb{R}^3 \mid x_3 \in \mathbb{R} \right\}.$$

Observemos que este conjunto puede escribirse alternativamente como

$$S_L = \left\{ \begin{bmatrix} \frac{7}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix} \in \mathbb{R}^3 \mid x_3 \in \mathbb{R} \right\},$$

mismo que puede visualizarse como una recta que va en dirección del vector $\begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix}$ *y que pasa por el punto* $(\frac{7}{4}, -\frac{3}{4}, 0)$ *en* $\mathbb{R}^3$.

Finalmente, queremos determinar cuantas soluciones existen para el sistema homogéneo de ecuaciones

$$x_1 \begin{bmatrix} 3 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Este sistema puede reescribirse en forma matricial como

$$\begin{bmatrix} 3 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

que es simplemente, el sistema homogéneo H asociado a nuestro sistema original.

Por la proposicón (2.33) al final del Capítulo 1, sabemos que, dado que L tiene un número infinito de soluciones, H también tiene un número infinito de soluciones. De hecho, el conjunto solución de este sistema homogéneo es

$$S_H = \left\{ \begin{bmatrix} -\frac{1}{4}x_3 \\ -\frac{3}{4}x_3 \\ x_3 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\}.$$

Este conjunto puede escribirse alternativamente como

$$S_H = \left\{ x_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix} \in \mathbb{R}^3 \mid x_3 \in \mathbb{R} \right\},$$

mismo que representa la recta en dirección de $\begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix}$ *y que pasa por el origen, es decir, S_L y S_H son rectas paralelas en $\mathbb{R}^3$ (ver Figura 2.7).*

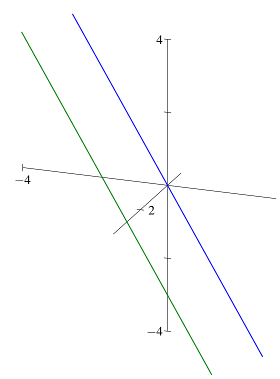

Figura 2.7

La conclusión es que hay un número infinito de formas distintas de expresar al vector cero como combinación lineal de los tres vectores en cuestión.

Observación 4.2 *En el ejemplo anterior, los conjuntos de vectores*

$$S_L = \left\{ \begin{bmatrix} \frac{7}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix} \in \mathbb{R}^3 \mid x_3 \in \mathbb{R} \right\},$$

$$S_H = \left\{ x_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \end{bmatrix} \in \mathbb{R}^3 \mid x_3 \in \mathbb{R} \right\}$$

pueden reescribirse como sigue

$$S_L = \left\{ \begin{bmatrix} \frac{7}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + r \begin{bmatrix} -1 \\ -3 \\ 4 \end{bmatrix} \in \mathbb{R}^3 \mid r \in \mathbb{R} \right\},$$

$$S_H = \left\{ r \begin{bmatrix} -1 \\ -3 \\ 4 \end{bmatrix} \in \mathbb{R}^3 \mid r \in \mathbb{R} \right\},$$

simplemente definiendo $r = \frac{1}{4}x_3$.

El resultado, aparentemente inocuo, de que el vector cero puede ser expresado como combinación lineal de un conjunto de vectores de varias formas, nos lleva a uno de los conceptos más importantes del álgebra lineal: el de independencia (y dependencia) lineal, que abordamos a continuación.

Dependencia e independencia lineal

Definición 4.3 *Sean* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ *vectores de* $\mathbb{R}^n$. *Decimos que el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *es* ***linealmente dependiente (LD)*** *syss existen escalares* $\beta_1, \beta_2, \ldots, \beta_p$, ***no todos cero,*** *tales que*

$$\beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p = \vec{0}.$$

Esto es, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *es linealmente dependiente syss el vector cero de* $\mathbb{R}^n$ *puede expresarse como una combinación lineal* ***no trivial*** *(con algún coeficiente diferente del cero) de estos* p *vectores.*

Ejemplo 4.4 *Consideremos a los vectores de* $\mathbb{R}^3$

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -2 \\ -4 \\ 0 \end{bmatrix} \text{ y } \vec{u}_3 = \begin{bmatrix} 666 \\ \pi \\ -1 \end{bmatrix}.$$

Entonces el conjunto $\{\vec{u}_1, \vec{u}_2\}$ *es LD pues* $2\vec{u}_1 + 1\vec{u}_2 = \vec{0}$. *Asimismo, el conjunto* $\{\vec{u}_1, \vec{u}_2, u_3\}$ *también es linealmente dependiente puesto que* $2\vec{u}_1 + 1\vec{u}_2 + 0\vec{u}_3 = \vec{0}$.

Ejemplo 4.5 *Como ya hemos visto en el Ejemplo 4.1, existe un número infinito de soluciones para la ecuación*

$$x_1 \begin{bmatrix} 3 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

por lo que el conjunto

$$\left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

es LD.

Definición 4.6 *Sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ vectores de $\mathbb{R}^n$. Decimos que el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es* ***linealmente independiente (LI)*** *syss $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ no es linealmente dependiente.*

Observación 4.7 *Si $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ son vectores de $\mathbb{R}^n$ y B es la matriz de $n \times p$ definida de tal forma que la columna B^j coincide con v_j , es decir,*

$$B = [\vec{v}_1 \ \vec{v}_2 \cdots \vec{v}_p],$$

entonces las siguientes afirmaciones son equivalentes:

1. *$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto linealmente independiente.*
2. *La única forma de expresar al vector cero como una combinación lineal de $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ es la trivial.*
3. *$x_1\vec{v}_1 + x_2\vec{v}_2, \ldots, +x_p\vec{v}_p = \vec{0}$ implica que el vector $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ está forzado a ser el vector cero de R^p.*
4. *La ecuación matricial $B\vec{x} = \vec{0}$ tiene exáctamente una solución (la solución trivial)*

Ejemplo 4.8 *Considérense los vectores de $\mathbb{R}^3$*

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \ y \ \vec{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

y supongamos que deseamos determinar si el conjunto $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es linealmente independiente. Para ello, consideramos a la matriz de 3×3, definida como $B = [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3]$ y al sistema homogéneo $B\vec{x} = \vec{0}$. Notamos que la matriz aumentada de este sistema está dada por

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right]$$

y tiene como FER a la matriz

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] = (I \mid \vec{0}).$$

Por lo tanto hay exáctamente una solución (la solución trivial) y el conjunto $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ *es LI.*

Observemos que el conjunto de vectores

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

en $\mathbb{R}^2$ y el conjunto análogo,

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

en $\mathbb{R}^3$, tienen, entre otras, las siguientes carácteristicas:

1. Ambos son linealmente independientes (el caso general se demuestra abajo en la proposición 4.10).

2. Cualquier vector en $\mathbb{R}^2$ o, alternativamente, en $\mathbb{R}^3$ puede expresarse trivialmente como combinación lineal de ellos.

En efecto, dado cualquier vector $\begin{bmatrix} a \\ b \end{bmatrix}$ en $\mathbb{R}^2$, se tiene que

$$\begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Asimismo si $\begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3$, entonces

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Evidentemente lo anterior puede generalizarse a cualquier $\mathbb{R}^n$ y esto nos lleva a otro concepto muy importante, el de base, mismo que se desarrollará formalmente en el siguiente capítulo. Por lo pronto, consideramos un caso particular, el de base estándar que se define a continuación.

Definición 4.9 *La **base estándar** para $\mathbb{R}^n$ es el conjunto $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$, donde e_i es el vector de $\mathbb{R}^n$ cuya i-ésima entrada es igual a uno y cuyas otras entradas son todas cero. Es decir, podemos visualizar a e_i como*

$$\begin{array}{r} \\ \\ \\ i \longrightarrow \\ \\ \\ \\ \end{array} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

en donde el "1" aparace en la i-ésima coordenada.

Proposición 4.10 *La base estándar para $\mathbb{R}^n$ es LI.*

Demostración. Notemos que el conjunto

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

es LI syss el sistema homogéneo

$$\begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & & \ddots & \\ 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

cuya FER es $(I \mid \vec{0})$, tiene solución única, lo cual es obvio. ■

Proposición 4.11 *Sea $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$, entonces $\vec{v} \in span\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$.*

Demostración. La demostración es inmediata ya que

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n.$$

■

Observación 4.12 *En ocasiones es necesario identificar coordenadas específicas de algún vector*

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n,$$

para este propósito la siguiente relación, que es inmediata a partir de la definición del producto punto, es sumamente útil.

$$\vec{e}_i \cdot \vec{v} = v_i.$$

Algunos resultados

Observemos que dado el conjunto $\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$ de vectores en $\mathbb{R}^3$, se tiene que

$$1 \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

de manera que el conjunto es LD. Esto nos lleva a la siguiente proposición.

Proposición 4.13 *Si uno de los vectores $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in \mathbb{R}^n$ es el vector cero, entonces el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es LD.*

Demostración. Siempre pueden reordenarse los vectores de manera que, sin pérdida de generalidad, podemos suponer que $\vec{v}_1 = \vec{0}$. Entonces

$$1\vec{v}_1 + 0\vec{v}_2 + \ldots + 0\vec{v}_p = \vec{0},$$

por lo que es posible expresar al vector cero como una combinación lineal no trivial de $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$. Concluimos pues que el conjunto formado por estos p vectores es LD. ■

Ejemplo 4.14 *Consideremos los vectores*

$$\vec{u}_1 = \begin{bmatrix} 1 \\ \pi \\ 7 \\ -1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -7 \\ -1 \\ 0 \\ 6 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

en $\mathbb{R}^4$. *Como*

$$0\vec{u}_1 + 0\vec{u}_2 + 107{,}5\vec{u}_3 = \vec{0},$$

tenemos que el conjunto $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ *es LD. Observar que es irrelevante quien sea el escalar, distinto del cero, que multiplica a* $\vec{u}_3$*, pues éste es el vector nulo.*

Proposición 4.15 *Si el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *de vectores de* $\mathbb{R}^n$ *es LD y* $\vec{w} \in \mathbb{R}^n$ *es cualquier otro vector, entonces el conjunto*

$$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{w}\}$$

también es LD.

Demostración. Por hipótesis, existen escalares $\beta_1, \beta_2, \ldots, \beta_p$ **no todos cero** tales que

$$\beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p = \vec{0}.$$

Por lo tanto,

$$\beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p + 0w = \vec{0}$$

es una combinación lineal no trivial (pues al menos uno de los escalares allí involucrados es distinto de cero) y el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, w\}$ es LD. ■

Este resultado simplemente nos dice que si un conjunto de vectores es LD, entonces cualquier conjunto que lo contenga también lo será.

Ejemplo 4.16 *Consideremos a los vectores de* $\mathbb{R}^4$,

$$\vec{w}_1 = \begin{bmatrix} 0 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \vec{w}_2 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \vec{w}_3 = \begin{bmatrix} 9 \\ 4 \\ 6 \\ 0 \end{bmatrix} \text{ y } \vec{w}_4 = \begin{bmatrix} 0.89 \\ 3\pi \\ 7 \\ -79 \end{bmatrix}$$

y supongamos que deseamos determinar si el conjunto formado por estos cuatro vectores es LD o LI. Para ello, primero notamos que $2\vec{w}_1 + 3\vec{w}_2 - \vec{w}_3 = \vec{0}$*, por lo que el conjunto* $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ *es LD. Esto último, junto con la proposición anterior, implican que* $\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\}$ *también es LD (de hecho,* $2\vec{w}_1 + 3\vec{w}_2 - \vec{w}_3 + 0\vec{w}_4 = \vec{0}$*).*

Proposición 4.17 *Sean* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}_{p+1}$ *vectores de* $\mathbb{R}^n$*. El conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}_{p+1}\}$ *es LD, syss al menos uno de estos vectores es combinación lineal de los otros* p *vectores.*

Para ejemplificar este resultado, y antes de pasar a su demostración, observamos que, en el ejemplo anterior, el conjunto $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ era LD, por lo que al menos uno de esos tres vectores es combinación lineal de los otros dos. En este caso, a partir de la igualdad

$$2\vec{w}_1 + 3\vec{w}_2 - \vec{w}_3 = \vec{0},$$

se sigue que $\vec{w}_3 \in span\{\vec{w}_1, \vec{w}_2\}$ (pues $\vec{w}_3 = 2\vec{w}_1 + 3\vec{w}_2$). De igual manera, $\vec{w}_1 \in span\{\vec{w}_2, \vec{w}_3\}$ (pues $\vec{w}_1 = -\frac{3}{2}\vec{w}_2 + \frac{1}{2}\vec{w}_3$) y $\vec{w}_2 \in span\{\vec{w}_1, \vec{w}_3\}$ (pues $\vec{w}_2 = -\frac{2}{3}\vec{w}_1 + \frac{1}{3}\vec{w}_3$). Regresemos ahora a la prueba de la Proposición 4.17.

Demostración. Supongamos primero que el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}_{p+1}\}$ es LD. Por definición, existen escalares $\beta_1, \beta_2, \ldots, \beta_p, \beta_{p+1}$ no todos cero tales que

$$\vec{0} = \beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p + \beta_{p+1}\vec{v}_{p+1}.$$

Reordenando los vectores, de ser necesario, podemos suponer que $\beta_{p+1} \neq 0$ y, por lo tanto, existe $\frac{1}{\beta_{p+1}}$y $\vec{v}_{p+1}$ puede despejarse como

$$\vec{v}_{p+1} = -\frac{\beta_1}{\beta_{p+1}}\vec{v}_1 - \frac{\beta_2}{\beta_{p+1}}\vec{v}_2 + \ldots - \frac{\beta_p}{\beta_{p+1}}\vec{v}_p,$$

es decir, $\vec{v}_{p+1} \in span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$.

Supongamos ahora que al menos uno de los $p+1$ vectores del conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}_{p+1}\}$ es combinación lineal de los otros p vectores. Sin perdida de generalidad, podemos suponer que $\vec{v}_{p+1} \in span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$. Entonces existen escalares $\alpha_1, \ldots, \alpha_p$ tales que

$$\alpha_1\vec{v}_1 + \ldots + \alpha_p\vec{v}_p = \vec{v}_{p+1}$$

o lo que es lo mismo,

$$\alpha_1\vec{v}_1 + \ldots + \alpha_p\vec{v}_p - 1\vec{v}_{p+1} = \vec{0}.$$

Por lo tanto, el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}_{p+1}\}$ es LD. ■

Ejemplo 4.18 *En el Ejemplo 4.17 el conjunto $\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\}$ es LD. En este caso, $\vec{w}_1 \in span\{\vec{w}_2, \vec{w}_3, \vec{w}_4\}$, $\vec{w}_2 \in span\{\vec{w}_1, \vec{w}_3, \vec{w}_4\}$, $\vec{w}_3 \in span\{\vec{w}_1, \vec{w}_2, \vec{w}_4\}$, pero $\vec{w}_4 \notin span\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. Observemos que la Proposición 4.17 dice que al menos un vector puede despejarse en términos de los demás. Este ejemplo ilustra que NO todos los vectores pueden despejarse, como es el caso de $\vec{w}_4$.*

Corolario 4.19 *Sean $\vec{v}_1, \vec{v}_2$ vectores de $\mathbb{R}^n$. El conjunto $\{\vec{v}_1, \vec{v}_2\}$ es LD syss al menos uno de los vectores es multiplo escalar del otro (combinación lineal del otro).*

Ejemplo 4.20 *Consideremos a los vectores*

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$$

en $\mathbb{R}^2$*. cláramente se tiene que* $\vec{v}_2 = -2\vec{v}_1$*, por lo que el conjunto* $\{\vec{v}_1, \vec{v}_2\}$ *es LD.*

Ejemplo 4.21 *Consideremos los vectores*

$$\vec{z}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \vec{z}_2 = \begin{bmatrix} -\pi \\ 0 \end{bmatrix}$$

en $\mathbb{R}^2$*. Entonces* $\{\vec{z}_1, \vec{z}_2\}$ *es un conjunto LI ya que no existe un escalar c tal que* $c\vec{z}_1 = \vec{z}_2$*. En efecto, si tal escalar existiese, tendríamos simultáneamente que* $1c = -\pi$ *y que* $2c = 0$*, lo cual es imposible.*

Ejemplo 4.22 *Considérense* $\vec{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *y* $\vec{v}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ *en* $\mathbb{R}^2$*, entonces*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} -1 \\ 1 \end{bmatrix};$$

sin embargo, el vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *no es múltiplo de* $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$*. En este caso* $\{\vec{v}_1, \vec{v}_2\}$ *es un conjunto LD, en el cual* $\vec{v}_2$ *es múltiplo de* $\vec{v}_1$*, pero* $\vec{v}_1$ *no es múltiplo de* $\vec{v}_2$*.*

Ejemplo 4.23 *Consideremos a los vectores de* $\mathbb{R}^4$

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 4 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -2 \\ 56 \\ 8.8 \\ 9 \end{bmatrix} \text{ y } \vec{u}_3 = \begin{bmatrix} 9 \\ 18 \\ 36 \\ 0 \end{bmatrix}.$$

Como $\vec{u}_3$ *es claramente un multiplo de* $\vec{u}_1$*,* $\{\vec{u}_1, \vec{u}_3\}$ *es LD, por lo que* $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ *también es LD en virtud de la Proposición 4.15.*

Proposición 4.24 *Si el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *en* $\mathbb{R}^n$ *es LI y* $v_{p+1} \in \mathbb{R}^n$ *es cualquier otro vector con la propiedad de que*

$$v_{p+1} \notin span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\},$$

entonces el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, v_{p+1}\}$ *también es LI.*

Demostración. Supongamos que existen escalares $\beta_1, \beta_2, \ldots, \beta_{p+1}$ tales que

$$\beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p + \beta_{p+1}\vec{v}_{p+1} = \vec{0}. \quad (\bigstar)$$

Debemos verificar que $\beta_1 = \ldots = \beta_p = \beta_{p+1} = 0$.

Para ello, comenzamos afirmando que $\beta_{p+1} = 0$. En efecto, si $\beta_{p+1} \neq 0$, entonces $\frac{1}{\beta_{p+1}}$ es un número real y, utilizando la igualdad ($\bigstar$), se sigue que

$$\vec{v}_{p+1} = -\frac{\beta_1}{\beta_{p+1}}\vec{v}_1 - \frac{\beta_2}{\beta_{p+1}}\vec{v}_2 - \ldots - \frac{\beta_p}{\beta_{p+1}}\vec{v}_p,$$

lo cual contradice la hipótesis de que $v_{p+1} \notin span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$. Por lo tanto, $\beta_{p+1} = 0$. Esto último junto con la igualdad ($\bigstar$) implican que

$$\beta_1\vec{v}_1 + \beta_2\vec{v}_2 + \ldots + \beta_p\vec{v}_p = \vec{0}.$$

Finalmente, como el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es LI.,concluimos que $\beta_1 = \beta_2 = \ldots = \beta_p = 0$. ■

Ejemplo 4.25 *Sean*

$$\vec{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} \pi \\ 0 \\ 6 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ 6 \end{bmatrix}.$$

vectores en $\mathbb{R}^3$*y supongamos que deseamos comprobar si el conjunto* $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *es LI. Existen al menos dos métodos posibles para efectuar tal comprobación.*

El primero consiste en verificar que el sistema homogéneo

$$\left[\begin{array}{ccc|c} 2 & \pi & 1 & 0 \\ 0 & 0 & \sqrt{2} & 0 \\ 0 & 6 & 6 & 0 \end{array}\right]$$

tiene solución única, lo cual no haremos aquí.

El segundo utiliza algunas de las propiedades que ya hemos demostrado: Es claro que $\vec{v}_1$ *no es un multiplo escalar de* $\vec{v}_2$*, por lo que el conjunto* $\{\vec{v}_1, \vec{v}_2\}$ *es LI. También es más o menos sencillo (puesto que* $\vec{v}_1$ *y* $\vec{v}_2$ *tienen ceros en su segunda coordenada y este no es el caso para* v_3*) notar que no existen escalares* x_1 *y* x_2 *tales que* $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{v}_3$*. El resultado anterior asegura que el conjunto* $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *es LI.*

Proposición 4.26 *Si* $\vec{v}$ *es un vector de* $\mathbb{R}^n$ *distinto de cero, entonces el conjunto* $\{\vec{v}\}$ *es LI.*

Demostración. Supongamos que existe $c \in \mathbb{R}$ tal que $c\vec{v} = \vec{0}$; debemos comprobar que $c = 0$. Pero, $c\vec{v} = \vec{0}$ syss $c = 0$ o $\vec{v} = \vec{0}$. Como $\vec{v} \neq \vec{0}$, concluimos que $c = 0$. ■

Ahora demostraremos que tres vectores en $\mathbb{R}^2$ siempre forman un conjunto LD. A manera de ejemplo supongamos que los vectores en cuestión son

$$\vec{v_1} = \begin{bmatrix} 1 \\ 77 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 66 \\ 2 \end{bmatrix}, \vec{v_3} = \begin{bmatrix} 96 \\ 2.8 \end{bmatrix}.$$

Para comprobar dicha afirmación debemos verificar que el sistema homogéneo

$$\left[\begin{array}{ccc|c} 1 & 66 & 96 & 0 \\ 77 & 2 & 2.8 & 0 \end{array}\right]$$

tiene una infinidad de soluciones, lo cual se sigue de la Proposición 2.35 en el capítulo I, pues se trata de un sistema homogéneo con más variables que ecuaciones que tiene al menos una variable libre. Este es el argumento central detrás del siguiente resultado.

Proposición 4.27 *Un conjunto*$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ *de vectores de* $\mathbb{R}^n$ *con* $m > n$, *siempre es un conjunto LD.*

Demostración. La demostración es consecuencia inmediata de la Proposición 2.35 ya que el sistema homogéneo $A\vec{x} = \vec{0}$, en donde la matriz A tiene como columnas a estos m vectores, tiene una infinidad de soluciones en virtud de que $m > n$. ■

Como corolario podemos mencionar los siguientes casos.

Corolario 4.28 *Sean* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ *vectores de* $\mathbb{R}^n$.

- *Si* $m > n$, *entonces el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ *es LD.*
- *Si* $m < n$, *entonces span*$\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\} \subsetneq \mathbb{R}^n$.

Demostración. La primera parte del corolario es simplemente la Proposición 4.27. Para probar la segunda parte, tomamos la matriz $A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_m]$ y suponemos que su FER está dada por B. Por lo tanto, la Proposición 2.30 nos dice que, como $m < n$, el $n-$ ésimo renglón de B es un renglón de ceros. Si tomamos

$$\vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n,$$

se tiene que $(B \mid \vec{e}_n)$ corresponde a un sistema inconsistente. Sea $\vec{b} \in \mathbb{R}^n$ tal que[2]

$$(A \mid \vec{b}) \sim (B \mid \vec{e}_n);$$

[2]La matriz aumentada $(A \mid \vec{b})$ se obtiene a partir de $(B \mid \vec{e}_n)$ haciendo las operaciones elementales inversas que se utilizaron para obtener B a partir de A. Este proceso no es enteramente trivial y más adelante se probará este resultado con mayor detalle.

entonces, $\vec{b} \notin span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ y $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\} \subsetneq \mathbb{R}^n$. ■

De manera metafórica podríamos decir que en el primer caso pecamos por exceso (de vectores) y la penitencia es la dependencia lineal, mientras que en el segundo pecamos por defecto y la penitencia se traduce en no poder generar a todo $\mathbb{R}^n$. En lo que sigue estudiaremos la situación en la cual tenemos exáctamente n vectores en $\mathbb{R}^n$.

Ejemplo 4.29 *Ilustraremos con un ejemplo simple el caso $m < n$ del corolario anterior. Supongamos que* $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} \subset \mathbb{R}^3$, *entonces la matriz cuyas columnas son estos vectores está dada por*

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

Su FER se obtiene como sigue:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \overset{R_2 \leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \overset{-R_1 + R_3}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \overset{-R_2 + R_3}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = B.$$

Observemos que si tomamos $\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3$, *el sistema cuya matriz aumentada es $(B \mid \vec{e}_3)$ es inconsistente. Podemos obtener un sistema equivalente $(A \mid \vec{b})$ para algún $\vec{b} \in \mathbb{R}^3$, simplemente revirtiendo las operaciones elementales utilizadas para obtener la equivalencia $A \sim B$. En efecto,*

$$(B \mid \vec{e}_3) = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \overset{R_2 + R_3}{\sim} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{array}\right] \overset{R_1 + R_3}{\sim}$$

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{array}\right] \overset{R_3 \leftrightarrow R_2}{\sim} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{array}\right] = (A \mid \vec{b}).$$

De aquí obtenemos que

$$\vec{b} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \notin span \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

y, por lo tanto,

$$span \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} \subsetneq \mathbb{R}^3.$$

Más adelante estudiaremos este caso con más detalle.

Ser o no ser la matriz identidad, esa es la cuestión

A lo largo de un buen número de ejemplos nos hemos encontrado con sistemas de ecuaciones cuyo conjunto solución es: vacío, es decir sin elementos; alternativamente, tiene un elemento único o bien, contiene un número infinito de elementos. Asimismo, el determinar si un sistema lineal tiene solución equivale a determinar si un vector es combinación lineal de otros. Concretamente, si el vector de términos independientes es combinación lineal de los vectores columna de la matriz de coeficientes del sistema lineal.

En el caso en el que se tiene el mismo número de ecuaciones y de variables, el lector observador habrá notado que la FER de la matriz de coeficientes es, en ocasiones, la matriz identidad. Esta observación es de suma importancia puesto que, como veremos a continuación, será lo que determine la estructura del conjunto solución o, alternativamente, del conjunto de vectores columna de la matriz de coeficientes.

Sean $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ vectores de $\mathbb{R}^n$ y sea $A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_n]$ la matriz cuyas columnas son estos vectores. Resumiendo lo visto en los capítulos anteriores, tenemos dos escenarios posibles en cuanto a la FER de A :

Escenario 1 *La FER de A es la matriz identidad $I \in M_{n\times n}$. Entonces para cualquier vector $\vec{b} \in \mathbb{R}^n$, la matriz aumentada $[A|\vec{b}]$ del sistema $A\vec{x} = \vec{b}$ es equivalente a $[I|\vec{c}]$ para algún vector $\vec{c} \in \mathbb{R}^n$. Por lo tanto, este sistema tiene solución única. En particular,*

$$[A|\vec{0}] \sim [I|\vec{0}]$$

*y el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es **LI**.*

Escenario 2 *Si la FER de A (llamémosle C) **no** es la matriz identidad, entonces $[A|\vec{0}]$ es equivalente a $[C|\vec{0}]$ y el sistema homogéneo asociado tiene al menos una variable libre. Por lo tanto, el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ es **LD**.*

En este caso, podemos apelar a la Proposición 4.17 para concluir que, al menos uno de los n vectores en cuestión, debe ser combinación lineal de los otros $n-1$. Supongamos, por ejemplo, que

$$v_1 \in span\{\vec{v}_2, \ldots, \vec{v}_n\}.$$

Gracias a la Proposición 3.26 y al Corolario 4.28 sabemos que

$$span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = span\{\vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\} \subsetneq \mathbb{R}^n.$$

Observación 4.30 *Sean $A, C \in M_{n\times n}$ una matriz y su FER, respectivamente. Si $C \neq I$, entonces C tiene, al menos, un renglón de ceros pues, de no ser así, sería la matriz identidad.*

Estos dos escenarios se resumen con los siguientes dos corolarios:

Corolario 4.31 *Sean $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ vectores de $\mathbb{R}^n$. Las siguientes afirmaciones son equivalentes.*

1. *El conjunto $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ es LI.*
2. *$span\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} = \mathbb{R}^n$.*
3. *Si $A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_n] \in M_{n\times n}$, entonces $A\vec{x} = \vec{b}$ tiene una solución única para cualquier $\vec{b} \in \mathbb{R}^n$.*

Corolario 4.32 *Sean $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ vectores de $\mathbb{R}^n$ y $V = span\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$. Las siguientes afirmaciones son equivalentes.*

1. *El conjunto $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ es LD.*
2. *$V \subsetneq \mathbb{R}^n$.*
3. *Si $A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_n] \in M_{n\times n}$, entonces $A\vec{x} = \vec{b}$ es un sistema inconsistente si $\vec{b} \notin V$ y es consistente con una infinidad de soluciones cuando $\vec{b} \in V$.*

Ejemplo 4.33 *Supongamos que deseamos determinar si el sistema*

$$\begin{aligned} x &= b_1, \\ 2x + y &= b_2 \end{aligned}$$

tiene solución para cualquier elección de b_1 y b_2. Esto es, nos preguntamos si span $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ es o no todo $\mathbb{R}^2$. Para ello, basta observar que el vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ no es multiplo escalar de $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ Por lo tanto, el conjunto $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ es LI y, por el Corolario 4.31, el sistema tiene solución única.

Ejemplo 4.34 *Sean*

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ y } \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

vectores en $\mathbb{R}^3$. *Nos preguntamos si* $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *es o no todo* $\mathbb{R}^3$. *Para ello, basta notar que la matriz*

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

cuyas columnas son estos vectores, tiene como FER a la matriz identidad. Por lo tanto, $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *es un conjunto*[3] *LI. y* $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \mathbb{R}^3$.

Ejemplo 4.35 *Consideremos a los vectores de* $\mathbb{R}^2$: $\vec{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ *y* $\vec{v}_2 = \begin{bmatrix} -3 \\ 6 \end{bmatrix}$. *Claramente son LD pues* $\vec{v}_2 = -3\ \vec{v}_1$. *Si*

$$A = \begin{bmatrix} 1 & -3 \\ -2 & 6 \end{bmatrix} \in M_{2\times 2}$$

y $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \in \mathbb{R}^2$, *es fácil verificar que la FER del sistema* $A\vec{x} = \vec{b}$ *está dada por*

$$\left[\begin{array}{cc|c} 1 & -3 & b_1 \\ 0 & 0 & b_2 + 2b_1 \end{array}\right];$$

con lo cual, el sistema es inconsistente siempre y cuando $b_2 + 2b_1 \neq 0$. *En cambio, si* $b_2 + 2b_1 = 0$, *el sistema tiene una infinidad de soluciones. Concretamente, el conjunto solución, en este caso, está dado por*

$$S = \left\{ \begin{bmatrix} b_1 + 3y \\ y \end{bmatrix} \in \mathbb{R}^2 \mid y \in \mathbb{R} \right\}.$$

Finalmente,

$$\begin{aligned} span\{\vec{v}_1, \vec{v}_2\} &= \left\{ \begin{bmatrix} b_1 \\ -2b_1 \end{bmatrix} \in \mathbb{R}^2 \mid b_1 \in \mathbb{R} \right\} \\ &= \left\{ b_1 \begin{bmatrix} 1 \\ -2 \end{bmatrix} \in \mathbb{R}^2 \mid b_1 \in \mathbb{R} \right\}, \end{aligned}$$

es decir, $span\{\vec{v}_1, \vec{v}_2\}$ *consiste en todos los múltiplos del vector* $\vec{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, *lo cual puede visualizarse como una recta en la dirección de* $\vec{v}_1$ *y que pasa por el origen.*

---

[3] Una forma alternativa de determinar la independencia lineal del este conjunto es notando que $\vec{v}_2$ no es multiplo escalar de $\vec{v}_1$ y que $\vec{v}_3 \notin span\{\vec{v}_1, \vec{v}_2\}$.

El caso general

En esta última sección probaremos que si un sistema lineal, con m ecuaciones y n variables, es consistente, entonces el sistema tiene solución única syss las columnas de la matriz de coeficientes forman un conjunto LI. Para este fin, se encontrará una relación entre el conjunto solución del sistema lineal y el conjunto solución del sistema homogéneo asociado.

En el Ejemplo 2.31 del Capítulo 1 observamos que el conjunto solución del sistema

$$\begin{cases} x + 2y - 3z = -4, \\ 2x + y - 3z = 4 \end{cases}$$

está dado por

$$S = \left\{ \begin{bmatrix} 4+z \\ -4+z \\ z \end{bmatrix} \mid z \in \mathbb{R} \right\},$$

que puede escribirse alternativamente como

$$S = \left\{ \begin{bmatrix} 4 \\ -4 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mid z \in \mathbb{R} \right\}.$$

Podemos obtener elementos específicos o **soluciones particulares** de este conjunto solución tomando distintos valores de z. Por ejemplo, si $z = 0$, se obtiene la solución particular

$$\vec{s}_p = \begin{bmatrix} 4 \\ -4 \\ 0 \end{bmatrix}.$$

También notamos que la solución del sistema homogéneo asociado es el conjunto

$$S_H = \left\{ \begin{bmatrix} z \\ z \\ z \end{bmatrix} \mid z \in \mathbb{R} \right\} = \left\{ z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mid z \in \mathbb{R} \right\}.$$

Concluimos que cualquier elemento del conjunto solución S **puede expresarse como la suma de una solución particular de dicho sistema y una solución del sistema homogéneo asociado.** El resultado formal de esta aseveración se enuncia y demuestra a continuación; sin embargo, antes de formularlo se requiere de la siguiente definición y de un lema referente a la distributividad del producto de matrices.

Definición 4.36 *Sea $A\vec{x} = \vec{b}$ un sistema lineal con $A \in M_{m\times n}$ y $\vec{b} \in \mathbb{R}^m$. Supongamos que el sistema es consistente, de manera que tiene al menos una solución y su conjunto solución S no es vacío. Sea $\vec{s}_p \in S$, cualquier elemento de este conjunto (al cual llamamos una* ***solución particular****). Si $A\vec{x} = \vec{0}$ es el sistema homogéneo asociado (que siempre es consistente) y S_H su conjunto solución, se define el conjunto $\vec{s}_p + S_H$ como sigue:*

$$\vec{s}_p + S_H = \{\vec{s}_p + \vec{h} \mid \vec{h} \in S_H\},$$

es decir, $\vec{s} \in \vec{s}_p + S_H$ syss existe una solución $\vec{h}$ del sistema homogéneo tal que $\vec{s} = \vec{s}_p + \vec{h}$.

Lema 4.37 *Sean $A \in M_{m\times n}$ y sean $\vec{u}$ y $\vec{z}$ vectores de $\mathbb{R}^n$ (vistos como matrices de $n \times 1$. Entonces:*

(a) $A(\vec{u} + \vec{z}) = A\vec{u} + A\vec{z}$.

(b) $A(\vec{u} - \vec{z}) = A\vec{u} - A\vec{z}$.

Demostración. Para (a) notamos que:

$$\begin{aligned} A(\vec{u} + \vec{z}) &= A \begin{bmatrix} u_1 + z_1 \\ \vdots \\ u_n + z_n \end{bmatrix} \\ &= (u_1 + z_1)A^1 + \ldots + (u_n + z_n)A^n = (u_1A^1 + \ldots u_nA^n) + (z_1A^1 + \ldots z_nA^n) \\ &= A \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} + A \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = A\vec{u} + A\vec{z}. \end{aligned}$$

La prueba de (b) es análoga y por lo tanto la omitimos. ■

Proposición 4.38 *Utilizando los mismos supuestos de la definición anterior, se cumple la siguiente igualdad entre conjuntos:*

$$S = \vec{s}_p + S_H.$$

Demostración. Sean $\vec{s}_p \in S$ una solución particular y $\vec{s} \in S$ cualquier elemento de S. Notemos que, utilizando el Lema 4.37,

$$A(\vec{s} - \vec{s}_p) = A\vec{s} - A\vec{s}_p = \vec{b} - \vec{b} = \vec{0},$$

por lo tanto, $\vec{s} - \vec{s}_p \in S_H$ y existe $\vec{h} \in S_H$ tal que $\vec{s} - \vec{s}_p = \vec{h}$, equivalentemente,

$$\vec{s} = \vec{s}_p + \vec{h} \in \vec{s}_p + S_H,$$

y hemos probado que $S \subset \vec{s}_p + S_H$. En forma análoga, si $\vec{s}_p + \vec{h}$ cualquier elemento de $\vec{s}_p + S_H$, entonces, una vez más apelando al Lema 4.37,

$$A(\vec{s}_p + \vec{h}) = A\vec{s}_p + A\vec{h} = \vec{b} + \vec{0} = \vec{b},$$

concluyéndose que $\vec{s}_p + \vec{h} \in S$ y $\vec{s}_p + S_H \subset S$. Por lo tanto, hemos demostrado ambas inclusiones y tenemos que $S = \vec{s}_p + S_H$. ■

Observación 4.39 *Es importante hacer notar que si el sistema*

$$A\vec{x} = \vec{b}$$

es inconsistente, la proposición anterior es evidentemente falsa. En efecto, en dicha situación $S = \varnothing$ *pero* $S_H = \{\vec{0}\}$.

Observación 4.40 *Si* $\vec{h}_1$ *y* $\vec{h}_2$ *son dos elementos distintos de* S_H, *entonces* $\vec{s}_p + \vec{h}_1$ *y* $\vec{s}_p + \vec{h}_2$ *son dos elementos distintos de* $\vec{s}_p + S_H$. *Esto es una consecuencia directa de la ley de cancelación.*

Corolario 4.41 *Sea* $A \in M_{m\times n}$, *sea* $\vec{b}$ *un vector de* $\mathbb{R}^m$ *y supongamos que el sistema* $A\vec{x} = \vec{b}$ *es consistente (es decir,* $S \neq \varnothing$*). Entonces se tiene que:*

1. *La solución del sistema es única syss las columnas de* A *forman un conjunto LI.*

2. *El sistema tiene un número infinito de soluciones syss las columnas de* A *forman un conjunto LD.*

Demostración. Es suficiente probar la primera parte ya que la segunda es un enunciado equivalente. Supongamos que existe s una solución única del sistema, entonces

$$S = \{\vec{s}\} = \vec{s}_p + S_H.$$

Como $\vec{s}_p \in S$ debe tenerse que $\vec{s} = \vec{s}_p$ y, por lo tanto, por la Proposición 4.38 se tiene que dado $\vec{h} \in S_H$ debe cumplirse la igualdad $\vec{s}_p = \vec{s}_p + \vec{h}$. Aplicando la ley de la cancelación llegamos a que $\vec{h} = \vec{0}$ de manera que $S_H = \{\vec{0}\}$. De aquí que el sistema homogéneo $A\vec{x} = \vec{0}$ tiene una única solución y por la Observación 4.7, las columnas de A forman un conjunto LI.

Ahora bien, si asumimos que el conjunto de las columnas de A es LI, una vez más, por la Observación 4.7 el sistema homogéneo asociado tiene como conjunto solución a $S_H = \{\vec{0}\}$ y, por la Proposición 4.38

$$S = \vec{s}_p + S_H = \left\{\vec{s}_p + \vec{0}\right\} = \{\vec{s}_p\},$$

de manera que $S = \{\vec{s}_p\}$ es un conjunto con un único elemento. ■

Observación 4.42 *La demostración directa del segundo enunciado del corolario anterior resulta ser interesante debido a la cardinalidad infinita. Veamos, si el sistema* $A\vec{x} = \vec{b}$ *tiene un número infinito de soluciones, entonces* S *es un conjunto infinito y, por lo tanto también lo será* S_H*. De aquí que* $\vec{0}$ *puede escribirse como combinación lineal no trivial de las columnas de* A *y éstas forman un conjunto LD.*

Ahora bien, si partimos de que las columnas de A *son LD, entonces* S_H *cuenta con un número infinito de elementos. Se podría pensar que podemos concluir que* $\vec{s}_p + S_H$ *y, por lo tanto* S*, son infinitos. En efecto lo serán pero tenemos que apelar a la ley de la cancelación pues hay que garantizar que al sumar* $\vec{s}_p$ *al conjunto infinito* S_H*, el conjunto resultante sigue siendo infinito. La Observación 4.40 nos garantiza que, dados dos vectores distintos* $\vec{h}_1$ *y* $\vec{h}_2$ *en* S_H*, los vectores* $\vec{s}_p + \vec{h}_1$ *y* $\vec{s}_p + \vec{h}_2$ *serán elementos distintos en* S*. Entonces el sistema* $A\vec{x} = \vec{b}$ *tiene más de una solución, por lo tanto* S *tendrá un número infinito de elementos.*

Ejemplo 4.43 *Sean*

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -3 & 0 \\ 1 & 1 & -1 \\ 5 & -1 & 1 \end{bmatrix}, \ \vec{b} = \begin{bmatrix} 2 \\ 3 \\ -2 \\ 2 \end{bmatrix} \ y \ \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Es fácil ver que $\vec{x}_p = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$ *es una solución particular del sistema* $A\vec{x} = \vec{b}$*. Consideremos el sistema homogéneo* $A\vec{x} = \vec{0}$*. La FER de la matriz aumentada correspondiente está dada como*

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

por lo que la única solución es $\vec{x} = \vec{0}$*. De aquí que las columnas de* A *son LI y* $\vec{x}_p$ *es la única solución. al sistema* $A\vec{x} = \vec{b}$*.*

Ejemplo 4.44 *Ahora bien, puede verificarse fácilmente que* $\vec{s}_p = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$ *es una solución particular del sistema*

$$\begin{aligned} 4x + 2z &= -4, \\ -3x - 3y &= 3, \\ x - y + z &= -1. \end{aligned}$$

La matriz aumentada del sistema homogéneo asociado está dada por

$$\left[\begin{array}{rrr|r} 4 & 0 & 2 & 0 \\ -3 & -3 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{array}\right]$$

cuya FER es

$$\left[\begin{array}{rrr|r} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Entonces, el conjunto solución del sistema homogéneo es

$$\begin{aligned} S_H &= \left\{ \begin{bmatrix} -\frac{z}{2} \\ \frac{z}{2} \\ z \end{bmatrix} \in \mathbb{R}^3 \mid z \in \mathbb{R} \right\} \\ &= \left\{ \frac{z}{2} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \in \mathbb{R}^3 \mid z \in \mathbb{R} \right\} \\ &= \left\{ s \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \in \mathbb{R}^3 \mid s \in \mathbb{R} \right\}. \end{aligned}$$

Dado que existe un número infinito de soluciones, las columnas de la matriz de coeficientes del sistema son LD. Asimismo, el sistema original tiene una infinidad de soluciones y el conjunto solución puede expresarse como

$$\begin{aligned} S &= \left\{ \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \in \mathbb{R}^3 \mid s \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} -1-s \\ s \\ 2s \end{bmatrix} \in \mathbb{R}^3 \mid s \in \mathbb{R} \right\}. \end{aligned}$$

Observación 4.45 *Evidentemente el parámetro s en el conjunto*

$$\left\{ \begin{bmatrix} -1-s \\ s \\ 2s \end{bmatrix} \in \mathbb{R}^3 \mid s \in \mathbb{R} \right\}$$

puede denotarse de cualquier forma.

Capítulo 5

Transformaciones lineales

Introducción

Cuando se tienen objetos matemáticos es natural querer estudiar las relaciones que existen entre ellos. Para este efecto, se utilizan las funciones entre estos objetos que preservan las propiedades que los caracterizan. Por ejemplo, si se tienen conjuntos de números A, B, para los cuales nos interesa el hecho de que todos sus elementos puedan ordenarse, entonces las funciones

$$f : A \longrightarrow B$$

que serán de interés son aquellas que preserven el orden, es decir, si $x, y \in A$ y $x < y$, entonces queremos que se cumpla $f(x) < f(y)$. Asimismo, en cálculo nos interesan ciertas propiedades geométricas (o topológicas) del conjunto de números reales (o de $\mathbb{R}^n$ en general) y, por lo tanto, las funciones de interés son aquellas que preservan propiedades de "cercanía" entre los puntos.

En los ejemplos anteriores se desea que las funciones preserven una relación de orden, en el primer caso, y alguna relación de distancia entre elementos, en el segundo caso. Para los espacios vectoriales, los que nos interesa preservar es la estructura algebraica lineal, es decir, la suma entre vectores y la multiplicación por escalares. Geométricamente, las estructuras lineales originadas por las combinaciones lineales de vectores en un espacio vectorial (rectas, planos o hiperplanos, en general), se transformarán en estructuras lineales de otro espacio vectorial. Por lo tanto, las funciones relevantes entre espacios vectoriales (en particular entre los espacios $\mathbb{R}^n$) son aquellas que preservan estas operaciones de suma y producto por escalares.

Transformaciones lineales

Definición 5.1 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una función. Decimos que T es una transformación lineal syss se satisfacen las siguientes dos condiciones:*

1. *T preserva sumas. Esto es, si $\vec{x}, \vec{y} \in \mathbb{R}^n$, entonces $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$.*

2. *T preserva multiplicación escalar. Esto es, si $\vec{x} \in \mathbb{R}^n$ y c es un escalar, entonces $T(c\vec{x}) = cT(\vec{x})$.*

Ejemplo 5.2 *Sea $T : \mathbb{R} \longrightarrow \mathbb{R}$ la función cuya regla de correspondencia es*

$$T(x) = 5x.$$

Notemos que dados $x, y \in \mathbb{R}$ se tiene que

$$T(x + y) = 5(x + y) = 5x + 5y = T(x) + T(y).$$

Adicionalmente, si $c \in \mathbb{R}$,

$$T(cx) = 5(cx) = c(5x) = cT(x),$$

por lo tanto T es una transformación lineal.

Ejemplo 5.3 *Sea $S : \mathbb{R} \longrightarrow \mathbb{R}$ dada por $S(x) = x^2$. Obsérvese que dados $x, y \in \mathbb{R}$ tenemos*

$$\begin{aligned} S(x + y) &= (x + y)^2 = x^2 + 2xy + y^2, \\ S(x) + S(y) &= x^2 + y^2, \end{aligned}$$

de manera que si $xy \neq 0$ se tendrá que $S(x + y) \neq S(x) + S(y)$ y S no es lineal pues, en general, no preserva sumas.

Ejemplo 5.4 *Sea $T_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ la función dada por*

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 3x_1 \end{bmatrix}.$$

Para familiarizarnos con esta función notemos, por ejemplo, que

$$T_1\left(\begin{bmatrix}2\\4\\5\end{bmatrix}\right)=\begin{bmatrix}0\\6\end{bmatrix}$$

Supongamos que deseamos determinar si T_1 es o no una transformación lineal. Para ello, veamos primero que T_1 preserva sumas:

$$T_1\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}\right)=T_1\left(\begin{bmatrix}x_1+y_1\\x_2+y_2\\x_3+y_3\end{bmatrix}\right)=\begin{bmatrix}0\\3(x_1+y_1)\end{bmatrix}=\begin{bmatrix}0\\3x_1+3y_1\end{bmatrix}$$

$$=\begin{bmatrix}0\\3x_1\end{bmatrix}+\begin{bmatrix}0\\3y_1\end{bmatrix}=T_1\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)+T_1\left(\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}\right).$$

El segundo paso consiste en comprobar que T_1 preserva multiplicación escalar. En efecto,

$$T_1\left(c\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)=T_1\left(\begin{bmatrix}cx_1\\cx_2\\cx_3\end{bmatrix}\right)=\begin{bmatrix}0\\3(cx_1)\end{bmatrix}$$

$$=c\begin{bmatrix}0\\3x_1\end{bmatrix}=cT_1\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right).$$

Por lo tanto, T_1 sí es una transformación lineal.

Ejemplo 5.5 *Sea $T_2:\mathbb{R}^2\longrightarrow\mathbb{R}^4$ la función dada por*

$$T_2\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x\\0\\y\\2x+3y\end{bmatrix}.$$

Notemos, por ejemplo, que

$$T_2\left(\begin{bmatrix}-5\\4\end{bmatrix}\right)=\begin{bmatrix}-5\\0\\4\\2\end{bmatrix}$$

Supongamos ahora que deseamos determinar si T_2 es o no una transformación lineal. Para ello, comprobemos primero que T_2 preserva *sumas:*

$$T_2\left(\begin{bmatrix}a\\b\end{bmatrix}+\begin{bmatrix}c\\d\end{bmatrix}\right)=T_1\left(\begin{bmatrix}a+c\\b+d\end{bmatrix}\right)=$$

$$\begin{bmatrix}a+c\\0\\b+d\\2(a+c)+3(b+d)\end{bmatrix}=\begin{bmatrix}a+c\\0\\b+d\\(2a+3b)+(2c+3d)\end{bmatrix}=$$

$$\begin{bmatrix}a\\0\\b\\2a+3b\end{bmatrix}+\begin{bmatrix}c\\0\\d\\2c+3d\end{bmatrix}=T_2\left(\begin{bmatrix}a\\b\end{bmatrix}\right)+T_2\left(\begin{bmatrix}c\\d\end{bmatrix}\right).$$

El segundo paso consiste en comprobar que T_2 preserva multiplicación escalar:

$$T_2\left(c\begin{bmatrix}a\\b\end{bmatrix}\right)=T_2\left(\begin{bmatrix}ca\\cb\end{bmatrix}\right)=\begin{bmatrix}ca\\0\\cb\\2(ca)+3(cb)\end{bmatrix}$$

$$=c\begin{bmatrix}a\\0\\b\\2a+3b\end{bmatrix}=cT_2\left(\begin{bmatrix}a\\b\end{bmatrix}\right).$$

Por lo tanto, T_2 sí es una transformación lineal.

Ejemplo 5.6 *Sea $S:\mathbb{R}^2\longrightarrow\mathbb{R}^4$ la función dada por*

$$S\left(\begin{bmatrix}x\\y\end{bmatrix}\right)=\begin{bmatrix}x\\1\\y\\2x+3y\end{bmatrix}.$$

y supongamos ahora que deseamos determinar si S es o no una transformación lineal. Para ello, observamos que S no preserva multiplicación escalar. En efecto,

$$S\left(2\begin{bmatrix}1\\0\end{bmatrix}\right)=S\left(\begin{bmatrix}2\\0\end{bmatrix}\right)=\begin{bmatrix}2\\1\\0\\4\end{bmatrix}\neq$$

$$\begin{bmatrix}2\\2\\0\\4\end{bmatrix}=2\begin{bmatrix}1\\1\\0\\2\end{bmatrix}=2\left(S\begin{bmatrix}1\\0\end{bmatrix}\right).$$

Por lo tanto, S no es una transformación lineal.

Ejemplo 5.7 *La función identidad* $I : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *(dada por* $I(\vec{x}) = \vec{x}$*, para todo* $\vec{x} \in \mathbb{R}^n$*) es una transformación lineal. En efecto, si* $\vec{x}, \vec{y} \in \mathbb{R}^n$ *y* c *es un escalar, entonces* $I(\vec{x} + \vec{y}) = \vec{x} + \vec{y} = I(\vec{x}) + I(\vec{y})$ *(por lo que* I *preserva sumas) e* $I(c\vec{x}) = c\vec{x} = cI(\vec{x})$ *(por lo que* I *preserva multiplicación escalar).*

Ejemplo 5.8 *La función constante cero* $\Theta : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *(dada por* $\Theta(\vec{x}) = \vec{0}$*, para todo* $\vec{x} \in \mathbb{R}^n$*) es una transformación lineal. En efecto, si* $\vec{x}, \vec{y} \in \mathbb{R}^n$ *y* c *es un escalar, entonces* $\Theta(\vec{x}+\vec{y}) = \vec{0} = \vec{0}+\vec{0} = \Theta(\vec{x})+\Theta(\vec{y})$ *(por lo que* Θ *preserva sumas) y* $\Theta(c\vec{x}) = \vec{0} = c\vec{0} = c\Theta(\vec{x})$ *(por lo que* Θ *preserva multiplicación escalar).*

El siguiente ejemplo es de naturaleza más práctica.

Ejemplo 5.9 *Una empresa polaca produce tres tipos de pelucas: A, B y* C*. Los costos (en zlotys) de producir cada unidad (de tipo A, B y* C*) aparecen en la siguiente tabla:*

	A	B	C
MO	*3*	*4*	*5*
MT	*6*	*2*	*1*

donde MO *y* MT *hacen referencia, respectivamente a mano de obra y a materiales. Así, si esta empresa desea producir* 5 *pelucas de tipo A, 3 de tipo B y 6 de tipo* C*, el vector de* $\mathbb{R}^2$ *cuya primera y segunda entrada son los costos totales de* MO *y de* MT *asociados a estas 14 pelucas está dado por*

$$\begin{bmatrix} 57 \\ 42 \end{bmatrix} = 5 \begin{bmatrix} 3 \\ 6 \end{bmatrix} + 3 \begin{bmatrix} 4 \\ 2 \end{bmatrix} + 6 \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 6 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 6 \end{bmatrix}.$$

Supongamos ahora que la empresa desea producir x *pelucas de tipo A,* y *de tipo B y* z *de tipo* C*. En general, el vector de* $\mathbb{R}^2$ *cuya primera y segunda entrada son los costos totales de* $M0$ *y de* MT *asociados a estas* $x + y + z$ *pelucas está dado por*

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

De forma natural esta multiplicación define una función de $\mathbb{R}^3$ *en* $\mathbb{R}^2$*. La siguiente proposición asegura que dicha función es, de hecho, una transformación lineal.*

Proposición 5.10 *Sea* A *una matriz de* $m \times n$*. Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *es la función dada por* $T(\vec{x}) = A\vec{x}$*, para todo* $\vec{x} \in \mathbb{R}^n$*, entonces* T *es una transformación lineal.*

Demostración. Para verificar que T preserva sumas simplemente notamos que si $\vec{x}, \vec{y} \in \mathbb{R}^n$, entonces

$$T(\vec{x}+\vec{y}) = A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y} = T(\vec{x}) + T(\vec{y}).$$

Por otro lado, para verificar que T preserva multiplicación escalar simplemente notamos que si $\vec{x} \in \mathbb{R}^n$ y c es un escalar, entonces

$$T(c\vec{x}) = A(c\vec{x}) = A \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix} = cx_1A^1 + \dots cx_nA^n$$

$$= c(x_1A^1 + \dots x_nA^n) = c\,(A\vec{x}) = cT(\vec{x}).$$

■

El resultado anterior dice que toda matriz determina una transformación lineal[1] mientras que el siguiente dice que toda transformación lineal da lugar a una matriz. Antes de enunciarlo formalmente recordemos que la base estándar para $\mathbb{R}^n$ es el conjunto $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$, donde e_i es el vector de $\mathbb{R}^n$ cuya i-ésima entrada es igual a uno y cuyas otras entradas son todas cero. Recordemos también que si $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ es un vector de $\mathbb{R}^n$, entonces $\vec{x}$ puede representarse como combinación lineal de $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ como

$$\vec{x} = x_1\vec{e}_1 + \dots x_n + \vec{e}_n.$$

Proposición 5.11 *Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal, $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ es la base estándar* ***ordenada*** *de $\mathbb{R}^n$ y $A \in M_{m\times n}$ es la matriz cuya i-ésima columna es igual al vector $T(e_i)$, para todo $i \in \{1, \dots n\}$, entonces $T(\vec{x}) = A\vec{x}$, para todo $\vec{x} \in \mathbb{R}^n$. Dicha matriz A es conocida como* ***la matriz estándar que representa a*** *T y está dada por $A = [T(\vec{e}_1)\ T(\vec{e}_2) \cdots T(\vec{e}_n)]$.*

Demostración. Como mencionamos anteriormente, para cualquier vector $\vec{x}$ de $\mathbb{R}^n$ ocurre que

$$\vec{x} = x_1\vec{e}_1 + \dots + x_n\vec{e}_n.$$

Aplicando T a ambos lados de esta igualdad se tiene entonces que

$$T(\vec{x}) = T(x_1\vec{e}_1 + \dots + x_n\vec{e}_n)$$

[1] La matriz $A \in M_{m\times n}$ determina a la transformación lineal como función, es decir, determina su dominio ($\mathbb{R}^n$) su codominio ($\mathbb{R}^m$) y su regla de correspondencia: a cada $\vec{x} \in \mathbb{R}^n$ se le asigna $A\vec{x} \in \mathbb{R}^m$.

$$\begin{aligned}
&= T(x_1\vec{e}_1) + \ldots + T(x_n\vec{e}_n) \\
&= x_1T(e_1) + \ldots + x_nT(e_n) \\
&= x_1A^1 + \ldots + x_nA^n \\
&= A\vec{x},
\end{aligned}$$

justo como se quería probar. Aquí es importante mencionar que la segunda y tercera igualdad se obtuvieron utilizando que T preserva sumas y multiplicación escalar, respectivamente. Por lo tanto, la hipótesis de linealidad es fundamental en esta demostración. ■

Observación 5.12 *El lector debe observar que el orden de los elementos de la base estándar es relevante. Cualquier reordenamiento de los mismos resultaría en una matriz distinta asociada a la transformación T. Concretamente, el orden de las columnas sería distinto. Por esta razón, aclaramos que el conjunto $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ es un conjunto ordenado para el cual el orden de los elementos es relevante. Por ejemplo, como conjuntos ordenados se tiene que $\{1, 2, 3\} \neq \{2, 1, 3\}$.*

Ejemplo 5.13 *Consideremos la transformación lineal $T_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por*

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 3x_1 \end{bmatrix}$$

e introducida previamente en el ejemplo (5.4). Como

$$T_1\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, T_1\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, T_1\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

se tiene que la matriz estándar que representa a T_1 es

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

Por lo tanto, para todo vector $\vec{x} \in \mathbb{R}^3$ se tiene que $T(\vec{x}) = A\vec{x}$. En particular,

$$\begin{aligned}
T_1\left(\begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}\right) &= \begin{bmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} \\
&= 2\begin{bmatrix} 0 \\ 3 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 0 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix},
\end{aligned}$$

lo cual coincide con lo que mencionamos en el ejemplo (4.1).

Ejemplo 5.14 *La matriz estándar que representa a la transformación lineal identidad* $I : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *es la matriz cuya columna* j *es el vector* $I(\vec{e}_j) = \vec{e}_j$. *Dicha matriz es, por supuesto, la matriz identidad de* $n \times n$.

Ejemplo 5.15 *Consideremos la transformación lineal* $T_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ *dada por*

$$T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ 0 \\ y \\ 2x+3y \end{bmatrix}$$

e introducida previamente en el ejemplo (5.5). Como

$$T_2\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix}, T_2\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix},$$

se tiene que la matriz estándard que representa a T_2 *es igual a*

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

Por lo tanto, para todo vector $\vec{x} \in \mathbb{R}^2$ *se tiene que* $T_2(\vec{x}) = A\vec{x}$. *En particular,*

$$T_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

y también

$$T_2\left(\begin{bmatrix} -5 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} -5 \\ 4 \end{bmatrix} = -5 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} -5 \\ 0 \\ 4 \\ 2 \end{bmatrix},$$

lo cual coincide con lo que mencionamos en el ejemplo (4.16).

Proposición 5.16 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *es una transformación lineal,* $\vec{0}_n$ *denota al vector cero de* $\mathbb{R}^n$ *y* $\vec{0}_m$ *denota al vector cero de* $\mathbb{R}^m$, *entonces* $T(\vec{0}_n) = \vec{0}_m$.

Demostración. Supongamos que $A \in M_{m\times n}$ es la matriz estándar que representa a T. Entonces, $T(\vec{x}) = A\vec{x}$, para todo vector $\vec{x}$ de $\mathbb{R}^n$. En particular,

$$T(\vec{0}_n) = A\vec{0}_n = 0A^1 + \ldots 0A^n = \vec{0}_m.$$

Otra demostración alternativa consiste, primero, en observar que, gracias al hecho de que T preserva sumas, se tiene

$$T(\vec{0}_n) + T(\vec{0}_n) = T(\vec{0}_n + \vec{0}_n) = T(\vec{0}_n) = T(\vec{0}_n) + \vec{0}_m.$$

Por lo tanto, usando la ley de la cancelación concluimos que $T(\vec{0}_n) = \vec{0}_m$. ■

Rango

Definición 5.17 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal. Definimos el rango de T, denotado por Rango(T), como el conjunto de todos los vectores $\vec{b} \in \mathbb{R}^m$ tales que existe $\vec{x} \in \mathbb{R}^n$ con la propiedad de que $T(\vec{x}) = \vec{b}$. Esto es, el rango de T es el conjunto de las imágenes, bajo de T, de los vectores de $\mathbb{R}^n$. En símbolos,*

$$Rango(T) = \{\vec{b} \mid T(\vec{x}) = \vec{b} \text{ para algún } x \in \mathbb{R}^n\}$$

o alternativamente,

$$Rango(T) = \{T(\vec{x}) \mid \vec{x} \in \mathbb{R}^n\}.$$

Observación 5.18 *Gracias a la Proposición (5.16) es claro que si*

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

es una transformación lineal, entonces $\vec{0}_m \in Rango(T)$.

Ejemplo 5.19 *El rango de la transformación identidad*

$$I : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

es igual al conjunto

$$\{I(\vec{x}) \mid \vec{x} \in \mathbb{R}^n\} = \{\vec{x} \mid \vec{x} \in \mathbb{R}^n\} = \mathbb{R}^n.$$

Ejemplo 5.20 *El rango de la transformación constante cero*

$$\Theta : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

es igual al conjunto

$$\{\Theta(\vec{x}) \mid \vec{x} \in \mathbb{R}^n\} = \{\vec{0}_m\}.$$

Ejemplo 5.21 *Consideremos la transformación lineal* $T_3 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ *dada por*

$$T_3\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Como $T_3\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$*, es claro que* $\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in Rango(T_3)$*. Supongamos ahora que deseamos describir todo el rango de* T_3*. Como la imagen de cualquier vector bajo* T_3 *es tal que su segunda, tercera y cuarta componente son iguales a cero, es claro que*

$$Rango(T_3) \subset \left\{ \begin{bmatrix} a \\ 0 \\ 0 \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\}.$$

Afirmamos que, de hecho,

$$Rango(T_3) = \left\{ \begin{bmatrix} a \\ 0 \\ 0 \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\}.$$

En efecto,

$$T_3\left(\begin{bmatrix} a \\ 0 \end{bmatrix}\right) = \begin{bmatrix} a+0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \in Rango(T_3).$$

Como el lector se podrá imaginar, no siempre es obvio determinar a simple vista el rango de una transformación lineal. Por ello, la siguiente observación (cuya demostración está autocontenida en el enunciado de la misma) es particularmente útil.

Observación 5.22 *Sean* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *una transformación lineal,* $A \in M_{m\times n}$ *la matriz estándar que representa a* T *y* $\vec{b} \in \mathbb{R}^m$*. Entonces las siguientes afirmaciones son equivalentes.*

1. $\vec{b} \in Rango(T)$.
2. *Existe* $\vec{x} \in \mathbb{R}^n$ *con la propiedad de que* $T(\vec{x}) = \vec{b}$.
3. *Existe* $\vec{x} \in \mathbb{R}^n$ *con la propiedad de que* $A\vec{x} = \vec{b}$.

4. *El sistema lineal cuya matriz aumentada es* $(A|\vec{b})$ *es consistente.*

5. $\vec{b} \in span\{A^1, A^2, \dots A^n\}$.

En particular, $Rango(T) = span\{A^1, A^2, \dots A^n\}$.

Ejemplo 5.23 *Sea* $T_4 : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ *la transformación lineal dada por*

$$T_4\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x \\ 0 \end{bmatrix}$$

y supongamos que deseamos, primero, encontrar la matriz estándar que representa a T_4 *y después, encontrar el rango de* T_4*. En el primer caso recordamos que la matriz buscada es* $[T_4(e_1)\ T_4(e_2)]$*. Utilizando la definición de* T_4 *se trata de la matriz*

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

Nótese que el rango de esta transformación es igual al conjunto de todos los vectores de $\mathbb{R}^3$ *cuya tercera componente es igual a cero. Este resultado también puede comprobarse notando que el rango de* T_4 *es el conjunto de todos los vectores* $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ *tales que la matriz aumentada*

$$\left[\begin{array}{cc|c} 0 & 1 & a \\ 1 & 0 & b \\ 0 & 0 & c \end{array}\right]$$

representa un sistema consistente. Obviamente, tal consistencia se obtiene syss $c = 0$.

Ejemplo 5.24 *Sea* $T_5 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ *la transformación lineal que satisface las siguientes igualdades:*

$$T_5\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 4 \end{bmatrix}\ ,\ T_5\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\ \text{ y }\ T_5\left(\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

y supongamos que deseamos, primero, encontrar la matriz estándar que representa a T_4 *y después encontrar el rango de* T_4.

En el primer caso recordamos que la matriz que representa a T_5 *es la matriz* $A = [T_5(e_1)\ T_5(e_2)\ T_5(e_3)]$*. Ahora bien, para calcular* $T_5(\vec{e}_1)$ *notamos que*

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

Por lo tanto,

$$T_5\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = T_5\left(\begin{bmatrix}1\\2\\0\end{bmatrix} - 2\begin{bmatrix}0\\1\\0\end{bmatrix}\right).$$

Pero T_5 preserva sumas y multiplicación escalar, por lo que

$$\begin{aligned} T_5\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) &= T_5\left(\begin{bmatrix}1\\2\\0\end{bmatrix}\right) - 2T_5\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) \\ &= \begin{bmatrix}0\\4\end{bmatrix} - 2\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}-2\\4\end{bmatrix}. \end{aligned}$$

De manera análoga y con el fin de calcular $T_5(\vec{e}_3)$ notamos que

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

Por lo tanto,

$$T_5\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = T_5\left(\frac{1}{2}\begin{bmatrix}0\\0\\2\end{bmatrix}\right) = \frac{1}{2}T_5\left(\begin{bmatrix}0\\0\\2\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}6\\0\end{bmatrix} = \begin{bmatrix}3\\0\end{bmatrix}$$

(la segunda igualdad se sigue porque T_5 preserva multiplicación escalar). En conclusión, la matriz que representa a la transformación T_5 es igual a la matriz

$$A = \begin{bmatrix}-2 & 1 & 3\\ 4 & 0 & 0\end{bmatrix} \tag{5.1}$$

Finalmente, el rango de T_5 es el conjunto de todos los vectores $\begin{bmatrix}a\\b\end{bmatrix}$ tales que la matriz

$$\left[\begin{array}{ccc|c} -2 & 1 & 3 & a\\ 4 & 0 & 0 & b \end{array}\right]$$

representa a un sistema consistente. Otra forma más práctica es simplemente notar que

$$\begin{aligned} Rango(T_5) &= span\left\{\begin{bmatrix}-2\\4\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}3\\0\end{bmatrix}\right\} \\ &= span\left\{\begin{bmatrix}-2\\4\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}\right\} = \mathbb{R}^2. \end{aligned}$$

Esta última igualdad se obtiene recordando que, por el corolario (4.31), un conjunto LI con n vectores en $\mathbb{R}^n$, genera a todo $\mathbb{R}^n$.

Ejemplo 5.25 *Consideremos a la transformación lineal T_5 del ejemplo anterior y supongamos que deseamos encontrar el conjunto de todos los vectores*

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

tales que $T_5(\vec{x}) = \begin{bmatrix} 4 \\ 12 \end{bmatrix}$. Para ello, simplemente recordamos que para la matriz A dada por (5.1), se tiene que $A\vec{x} = T_5(\vec{x})$; por lo tanto, nuestro problema se reduce a resolver el sistema $A\vec{x} = \begin{bmatrix} 4 \\ 12 \end{bmatrix}$ cuya matriz aumentada es

$$\left[\begin{array}{ccc|c} -2 & 1 & 3 & 4 \\ 4 & 0 & 0 & 12 \end{array}\right]$$

y que tiene como FER a la matriz

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 3 & 10 \end{array}\right].$$

Por lo tanto, el conjunto buscado es

$$\left\{ \begin{bmatrix} 3 \\ 10 - 3x_3 \\ x_3 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 3 \\ 10 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\}.$$

Definición 5.26 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal. Decimos que T es suprayectiva (o "sobre") syss Rango$(T) = \mathbb{R}^m$.*

Como ya hemos visto, en el caso de las transformaciones lineales

$$T_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2 \text{ y } T_5 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2,$$

hay transformaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$ con $n > m$ que pueden no ser suprayectivas (este es el caso de T_1 y T_5). En el caso de las transformaciones T_2, T_3 y T_4 tenemos que éstas van de $\mathbb{R}^n$ en $\mathbb{R}^m$ con $n < m$, por lo que -utilizando el siguiente resultado- podemos afirmar que ninguna de las tres es suprayectiva (¡y tal afirmación puede realizarse incluso sin conocer el comportamiento específico de cada una de ellas!).

Proposición 5.27 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal. Si $n < m$, entonces T no es suprayectiva.*

Demostración. Sea A la matriz estándar que representa a T. Gracias a la Observación 5.22 sabemos que $Rango(T) = span\{A^1, \ldots, A^n\}$. Este conjunto es, por supuesto, un subconjunto de $\mathbb{R}^m$. Como $n < m$ y gracias a la máxima de “pecado por defecto” al final del Corolario 4.28, podemos asegurar que

$$Rango(T) = span\{A^1, \ldots, A^n\} \subsetneq \mathbb{R}^m,$$

por lo que T no es suprayectiva. ■

En el caso de que el dominio y el codominio de una transformación lineal coincidan, tenemos el siguiente resultado.

Proposición 5.28 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una transformación lineal y supongamos que A es la matriz estándar que representa a T. Entonces T es suprayectiva syss el conjunto $\{A^1, \ldots A^n\}$ es LI (o lo que es lo mismo, syss la FER de A es la matriz identidad).*

Demostración. Gracias a la observación (5.22) sabemos que

$$Rango(T) = span\{A^1, \ldots, A^n\}.$$

Este conjunto es, por supuesto, un subconjunto de $\mathbb{R}^n$. Ahora, por el corolario (4.31), sabemos que el conjunto generado por n vectores de $\mathbb{R}^n$ es igual a todo $\mathbb{R}^n$ syss esos n vectores forman un conjunto LI. Por lo tanto, T es suprayectiva (i.e., $Rango(T) = \mathbb{R}^n$) syss el conjunto $\{A^1, \ldots A^n\}$ es LI. ■

Ejemplo 5.29 *Sea $T_6 : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ la transformación lineal dada por*

$$T_6\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ x - y \\ z \end{bmatrix}$$

y supongamos que deseamos determinar si T_6 es suprayectiva. Para ello notamos, primero, que la matriz estándar que representa a T_6 es la matriz

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

cuya forma escalonada reducida es la matriz identidad de 3×3. Por la proposición anterior concluimos que T_6 es suprayectiva.

Inyectividad

Definición 5.30 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal. Decimos que T es* ***inyectiva*** *(o* ***uno-a-uno****) syss para cualesquiera dos vectores $\vec{x}, \vec{y} \in \mathbb{R}^n$, se cumple que si*

$$\vec{x} \neq \vec{y}, \text{ entonces } T(\vec{x}) \neq T(\vec{y})$$

(lo que se lee como que ***dos vectores distintos no pueden tener la misma imagen****). Equivalentemente, T es inyectiva syss para cualesquiera dos vectores $\vec{x}, \vec{y} \in \mathbb{R}^n$, se cumple que si*

$$T(\vec{x}) = T(\vec{y}), \text{ entonces } \vec{x} = \vec{y}.$$

Ejemplo 5.31 *Consideremos la transformación lineal $T_5 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ cuya matriz estándar está dada por*

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

considerada previamente en el ejemplo 5.24.

Notamos que

$$T_5\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} -2 & 1 & 3 \\ 4 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} =$$

$$x\begin{bmatrix} -2 \\ 4 \end{bmatrix} + y\begin{bmatrix} 1 \\ 0 \end{bmatrix} + z\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -2x + y + 3z \\ 4x \end{bmatrix},$$

por lo que

$$T_5\left(\begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = T_5\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right).$$

Dado que T_5 envia dos vectores distintos a la misma imagen, concluimos que T_5 no es inyectiva.

Ejemplo 5.32 *Consideremos la transformación lineal $T_3 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ dada por*

$$T_3\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x + y \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

que introducimos previamente en el ejemplo 5.21 como

$$T_3\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} = T_3\left(\begin{bmatrix} 2 \\ 0 \end{bmatrix}\right),$$

T_3 no es inyectiva.

Ejemplo 5.33 *Consideremos la transformación lineal* $T_7 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ *dada por*

$$T_7\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ 2x \\ 0 \\ 0 \end{bmatrix}$$

y supongamos que deseamos determinar si T_7 *es o no inyectiva. Supongamos que*

$$T_7\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = T_7\left(\begin{bmatrix} c \\ d \end{bmatrix}\right),$$

entonces debe comprobarse que $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$. *A partir de la definición de* T_7, *tenemos que*

$$\begin{bmatrix} a+b \\ 2a \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} c+d \\ 2c \\ 0 \\ 0 \end{bmatrix}.$$

De aquí se siguen las igualdades

$$a+b = c+d,$$
$$2a = 2c,$$

con lo cual se obtiene que $a = c$ *y* $b = d$. *De lo anterior se concluye que* $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$ *y* T_7 *es inyectiva.*

En lo que sigue proporcionaremos un método sistemático para decidir si una transformación lineal es o no inyectiva. Para ello, necesitamos la siguiente equivalencia que, a su vez, hace referencia a la Observación 5.16.

Proposición 5.34 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *es una transformación lineal, entonces las siguientes afirmaciones son equivalentes:*

(a) T *es inyectiva.*

(b) Si $\vec{x}$ *es tal que* $T(\vec{x}) = \vec{0}_m$, *entonces* $\vec{x} = \vec{0}_n$.

Demostración. Mostremos primero que (a) implica (b). Supongamos que T es inyectiva y que $T(\vec{x}) = \vec{0}_m$. Debemos verificar que $\vec{x}$ está forzado a ser el vector cero de $\mathbb{R}^n$. Ahora bien, como $T(\vec{0}_n) = \vec{0}_m = T(\vec{x})$ y T es inyectiva, debe cumplirse que $\vec{x} = \vec{0}_n$.

Supongamos ahora que T satisface (b) y que $\vec{a}$ y $\vec{b}$ son tales que

$$T(\vec{a}) = T(\vec{b}) \quad (\bigstar);$$

debemos probar que $\vec{a} = \vec{b}$. Por linealidad de T y gracias a ($\bigstar$), se tiene que

$$T(\vec{a} - \vec{b}) = T(\vec{a}) - T(\vec{b}) = \vec{0}_m.$$

Finalmente, utilizando (b) concluimos que $\vec{a} - \vec{b} = \vec{0}_n$. Esto es, $\vec{a} = \vec{b}$. ■

Corolario 5.35 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal y supongamos que A es la matriz estándar que representa a T. Entonces T es inyectiva syss el sistema lineal $A\vec{x} = \vec{0}_m$ tiene solución única. Esto es, T es inyectiva syss el conjunto $\{A^1, \ldots A^n\}$ es LI.*

Demostración. Gracias a la proposición anterior sabemos que T es inyectiva syss el sistema lineal $T(\vec{x}) = \vec{0}_m$ tiene solución única (la solución trivial). Ahora bien, como $T(\vec{x}) = A\vec{x}$, tenemos que T es inyectiva syss el sistema lineal $A\vec{x} = \vec{0}_m$ tiene solución única. ■

Ejemplo 5.36 *En el ejemplo 5.21 ya verificamos que la transformación lineal $T_3 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$, dada por*

$$T_3\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

no es inyectiva. Esta conclusión puede obtenerse de otra forma, utilizando el Corolario 5.35. Para ello, primero notamos que la matriz estándar que representa a T_3 es la matriz

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

Como el sistema homogéneo

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

tiene una infinidad de soluciones (nótese que la segunda columna de la matriz de coeficientes es multiplo de la primera), concluimos que T_3 no es inyectiva.

Ejemplo 5.37 *En el ejemplo 5.33 ya verificamos que la transformación lineal* $T_7 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$*, dada por*

$$T_7\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ 2x \\ 0 \\ 0 \end{bmatrix},$$

es inyectiva. Utilizando el Corolario 5.35 llegaremos a la misma conclusión. Para ello, primero notamos que la matriz estándar que representa a T_7 *es la matriz*

$$\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Como el sistema homogéneo

$$\begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

tiene solución única (nótese que la segunda columna de la matriz de coeficientes no es multiplo de la primera), concluimos (de nuevo) que T_7 *es inyectiva.*

Ejemplo 5.38 *En el ejemplo 5.24 ya verificamos que la transformación lineal* $T_5 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ *cuya matriz estándar es*

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

no es inyectiva. Utilizando el corolario 5.35 llegaremos a la misma conclusión. Como el sistema homogéneo

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

tiene una infinidad de soluciones (nótese que hay más variables que ecuaciones), concluimos que T_5 *no es inyectiva.*

El ejemplo anterior puede generalizarse como sigue.

Proposición 5.39 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *es una transformación lineal y* $n > m$*, entonces* T *no es inyectiva.*

Demostración. Supongamos que $A \in M_{m\times n}$ es la matriz estándar que representa a T y consideremos el sistema homogéneo $A\vec{x} = \vec{0}_m$ (esto es, $T(\vec{x}) = \vec{0}_m$). Como este sistema tiene una infinidad de soluciones (pues tiene más variables que ecuaciones), concluimos que T no es inyectiva. De hecho, esta infinidad de soluciones nos dice que existe un número infinito vectores en $\mathbb{R}^n$ que, bajo T, tienen al vector $\vec{0}_m$ como imagen. ■

El siguiente resultado (basado esencialmente en la Proposición 5.28, la discusión que precede a los corolarios 4.31 y 4.32 y el corolario 5.35), nos dice que, en el caso de transformaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^n$, la inyectividad es sinónimo de otras muchas propiedades.

Teorema 5.40 *Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es una transformación lineal y $A \in M_{n\times n}$ es la matriz estándar que representa a T, entonces las siguientes afirmaciones son equivalentes.*

1. *$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es suprayectiva.*
2. *Para cualquier vector $\vec{b} \in \mathbb{R}^n$, el sistema $A\vec{x} = b$ tiene solución (única).*
3. *La FER de A es la matriz identidad de $n \times n$.*
4. *Las columnas de A forman un conjunto linealmente independiente.*
5. *El sistema homogéneo $A\vec{x} = \vec{0}_n$ tiene exactamente una solución (a saber, la trivial).*
6. *T es inyectiva.*

Ejemplo 5.41 *Consideremos la transformación lineal $T_6 : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ dada por*

$$T_6\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ x-y \\ z \end{bmatrix},$$

introducida previamente en el Ejemplo 5.29. Ya hemos notado que la matriz estándar que representa a T_6 es la matriz

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

Cláramente, sus columnas forman un conjunto LI, por lo que T_6 y su matriz asociada satisfacen todas las propiedades enunciadas en el teorema anterior. En particular, T_6 no sólo es suprayectiva (como habíamos determinado anteriormente), sino que además es inyectiva.

Finalizamos esta sección mostrando que las transformaciones lineales inyectivas preservan independencia lineal.

Proposición 5.42 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal y sea $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subset \mathbb{R}^n$ cualquier conjunto LI de vectores. Entonces, T es inyectiva syss el conjunto $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_p)\}$ es LI.*

Demostración. Supongamos que T es inyectiva y sean $c_1, \ldots c_p$ escalares tales que

$$c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \ldots + c_pT(\vec{v}_p) = \vec{0}_m.$$

Por linealidad, tenemos que

$$\begin{aligned} c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \ldots + c_pT(\vec{v}_p) &= \\ T(c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_p\vec{v}_p) &= \vec{0}_m \end{aligned}$$

y, como T es inyectiva, se tiene que

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_p\vec{v}_p = \vec{0}_n.$$

Esta igualdad implica que

$$c_1 = c_2 = \ldots = c_p = 0,$$

puesto que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto LI. Concluimos que $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_p)\}$ también es LI.

Supongamos ahora que T preserva la independencia lineal de conjuntos de vectores y sean $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$, $\vec{v}_1 \neq \vec{v}_2$. Entonces, $\{\vec{v}_1 - \vec{v}_2\}$ es LI y, por hipótesis, el conjunto

$$\{T(\vec{v}_1 - \vec{v}_2)\} = \{T(\vec{v}_1) - T(\vec{v}_2)\}$$

también es LI. Entonces, necesariamente se cumple $T(\vec{v}_1) - T(\vec{v}_2) \neq \vec{0}_m$ o bien $T(\vec{v}_1) \neq T(\vec{v}_2)$. Concluimos así que T es inyectiva. ■

Como consecuencia inmediata de esta proposición tenemos que si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ una transformación lineal y $\vec{v}_1$ y $\vec{v}_2$ son dos vectores de $\mathbb{R}^n$ distintos del vector cero y tales que v_1 no es multiplo escalar de v_2 (i.e., $\{\vec{v}_1, \vec{v}_2\}$ es un conjunto L.I), entonces $T(\vec{v}_1)$ no es multiplo escalar de $T(\vec{v}_2)$, con lo que $T(\vec{v}_1)$ y $T(\vec{v}_2)$ no son vectores paralelos en el plano.

Geometría de transformaciones lineales

Como se mencionó anteriormente, las transformaciones lineales preservan las estructuras lineales. Los siguientes ejemplos ilustran visualmente este comportamiento.

Ejemplo 5.43 *Analicemos la geometría de la siguiente transformación lineal de $\mathbb{R}^2$ en $\mathbb{R}^2$ cuya matriz estándar está dada por*

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

Observemos que dado $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ *se tiene que*

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix},$$

con lo cual se trata de una reflexión sobre el eje horizontal, como se ilustra en la figura 4.1.

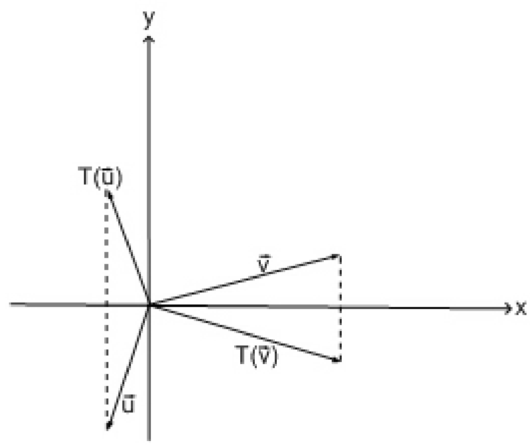

Figura 4.1

Ejemplo 5.44 *Si ahora deseamos una transformación lineal, digamos R, de $\mathbb{R}^2$ en $\mathbb{R}^2$, que rote a cada vector un ángulo θ, entonces podemos visualizar lo que les pasa a los vectores $\vec{e}_1$ y $\vec{e}_2$ como en la figura 4.2.*

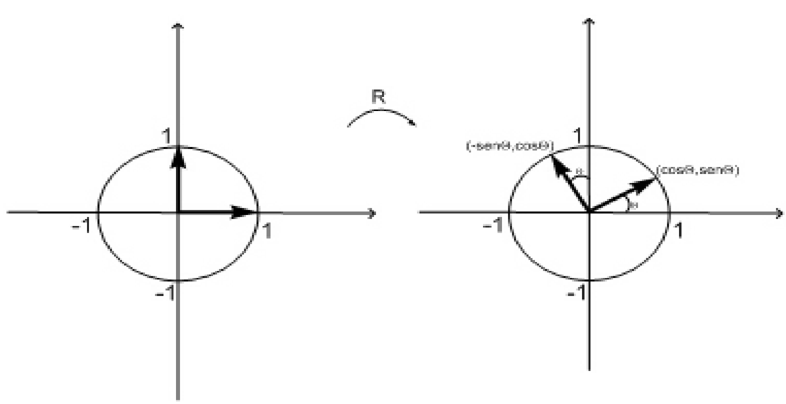

Figura 4.2

Por lo tanto, la matriz estándar asociada a esta transformación es simplemente

$$R=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.$$

Ejemplo 5.45 *Sean $\vec{u}$ y $\vec{v}$ dos vectores distintos en $\mathbb{R}^2$. Se quiere describir a los puntos P en el segmento que conecta a estos vectores[2]. Notemos, de la figura 4.3, que para algún $s\in[0,1]$ se cumple*

$$P=\vec{u}+s(\vec{v}-\vec{u}),$$

por lo tanto

$$P=(1-s)\vec{u}+s\vec{v}.$$

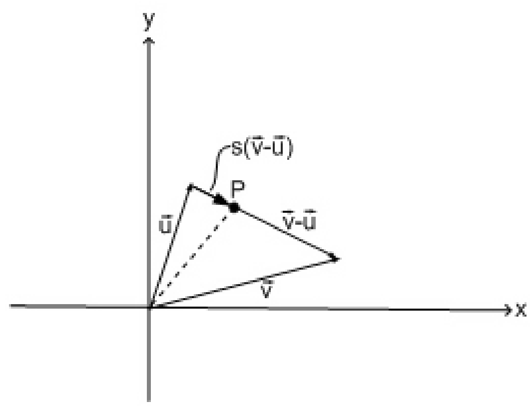

Figura 4.3

Ejemplo 5.46 *Si aplicamos una transformación lineal $T:\mathbb{R}^2\longrightarrow\mathbb{R}^2$ a los puntos del segmento que conecta a $\vec{u}$ y a $\vec{v}$ (como en el ejemplo anterior), se tiene que, por linealidad,*

$$T\left((1-s)\vec{u}+s\vec{v}\right)=(1-s)T\left(\vec{u}\right)+sT\left(\vec{v}\right).$$

Es decir, se obtienen los puntos que conectan a los vectores $T\left(\vec{u}\right)$ y $T\left(\vec{v}\right)$. Por ejemplo, si el conjunto C en $\mathbb{R}^2$es el cuadrado cuyos vértices están en los puntos $(0,0)$, $(1,0)$, $(0,1)$ y $(1,1)$, entonces para la transformación lineal dada por

$$T\left(\begin{bmatrix}x\\ y\end{bmatrix}\right)=\begin{bmatrix}3x\\ 2y\end{bmatrix},$$

la imagen del cuadrado, $T(C)$, se ilustra en la figura 4.4. Observamos que todos los puntos intermedios entre los vértices van a dar a puntos intermedios entre las imágenes de los mismos vértices.

[2] En realidad los vectores podrían pertenecer a $\mathbb{R}^n$, simplemente se toma $n=2$ por conveniencia.

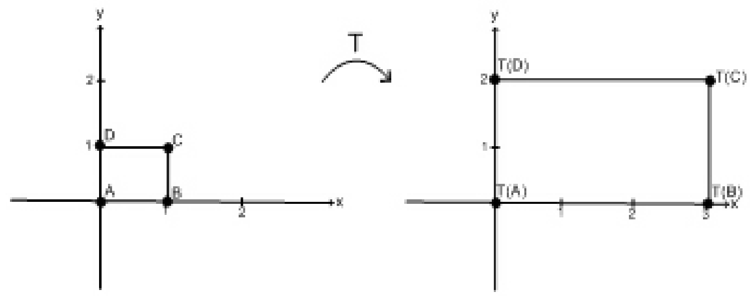

Figura 4.4

Capítulo 6

Álgebra Matricial

Introducción

En el capítulo anterior vimos que las transformaciones lineales,

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

pueden representarse como matrices de $m \times n$. Recordemos que en los cursos elementales de cálculo se estudian funciones (continuas o diferenciables, por ejemplo), cuyo dominio es algún subconjunto Ω de $\mathbb{R}^n$ y que toman valores en los números reales, es decir,

$$f : \Omega \longrightarrow \mathbb{R}.$$

Estas funciones pueden sumarse y multiplicarse por escalares (reales) simplemente definiendo, para toda pareja de funciones f, g y c un escalar real, la suma y el producto como sigue:

$$\begin{aligned}(f+g) : \Omega &\longrightarrow \mathbb{R},\\ (f+g)(x) &= f(x) + g(x),\\ (cf) : \Omega &\longrightarrow \mathbb{R},\\ (cf)(x) &= cf(x).\end{aligned}$$

Con estas definiciones, $f + g$ y cf siguen teniendo las mismas propiedades que las funciones originales. Por ejemplo, si f y g son continuas (o diferenciables o doblemente diferenciables, por ejemplo), $f + g$ y cf también lo serán. Asimismo, puede demostrarse que el conjunto de funciones

$$\{f : \Omega \longrightarrow \mathbb{R}\, | f \text{ es continua}\}\,,$$

en donde la palabra "continua" puede sustituirse por "diferenciable" o "doblemente diferenciable" o alguna característica similar, satisface todas las propiedades del espacio vectorial $\mathbb{R}^n$ dadas en la Proposición 3.6.

Una pregunta natural es si estas operaciones pueden definirse para las transformaciones lineales de manera que la suma y el producto por un escalar también sean transformaciones lineales. Lo más natural es utilizar la misma definición que se utiliza en cálculo. Concretamente, si

$$\begin{aligned} T : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \quad \text{y} \\ S : \mathbb{R}^n &\longrightarrow \mathbb{R}^m \end{aligned}$$

son dos transformaciones lineales y $c \in \mathbb{R}$, definimos

$$(T + S) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

como

$$(T + S)(\vec{x}) = f(\vec{x}) + g(\vec{x}).$$

Similarmente,

$$(cT) : \mathbb{R}^n \longrightarrow \mathbb{R}^m\ ,$$

está dada por

$$(cT)(\vec{x}) = cf(\vec{x}).$$

Es fácil verificar que las transformaciones así definidas también son lineales y se deja como un ejercicio al lector.

Ahora bien, si A y B son las matrices estándar asociadas a T y S, podemos denotar por $A + B$ a la matriz estándar asociada a $T + S$ y por cA a la matriz estándar asociada a cT. ¿Cómo son estas matrices? En este capítulo se estudian éstas y otras operaciones para matrices.

El conjunto de matrices $M_{m\times n}$ como espacio vectorial

Dada cualquier matriz $A \in M_{m\times n}$, denotaremos por A_{ij} a la entrada ij de dicha matriz.

Definición 6.1 *Sean* $A, B \in M_{m\times n}$ *y sea* c *un número real. Entonces definimos,*

(i) la ***suma de*** A ***y*** B *como la matriz* $A+B \in M_{m\times n}$*, cuyas entradas están dadas como*

$$(A+B)_{ij} = A_{ij} + B_{ij},$$

para toda i *y para toda* j*. En otras palabras, simplemente sumamos "entrada por entrada".*

(ii) El producto del escalar c *por la matriz* A *es la matriz* $cA \in M_{m\times n}$ *cuyas entradas están dadas por*

$$(cA)_{ij} = cA_{ij},$$

para toda i *y para toda* j*. Es decir, todas las entradas de la matriz* A *se multiplican por* c*.*

Notación 6.1 *En el contexto de* $M_{m\times n}$ *utilizaremos las siguientes convenciones.*

(a) Recordemos que la matriz cero (aquella cuyas componentes son todas cero) se denota por Θ*.*

(b) Si $A \in M_{m\times n}$*, entonces* $-A$ *denota a la matriz* $(-1)A$*.*

(c) Si $A, B \in M_{m\times n}$*, entonces* $A - B$ *significa* $A + (-B)$*.*

(d) Como siempre, la palabra escalar es sinónimo de número real.

Ejemplo 6.2 *Consideremos a las matrices*

$$A = \begin{bmatrix} 3 & 4 & 5 \\ -4 & 2 & 1 \end{bmatrix} \text{ y } B = \begin{bmatrix} 7 & -4 & 9 \\ 6 & 0 & -12 \end{bmatrix}.$$

Entonces

$$A+B = \begin{bmatrix} 10 & 0 & 14 \\ 2 & 2 & -11 \end{bmatrix}, 2A = \begin{bmatrix} 6 & 8 & 10 \\ -8 & 4 & 2 \end{bmatrix}, 2A-3B = \begin{bmatrix} -15 & 20 & -17 \\ -26 & 4 & 38 \end{bmatrix}.$$

Por un lado definimos la suma y el producto por escalares entre transformaciones lineales y, por otro, la suma y el producto por escalares para las matrices. Para que esto sea de utilidad debemos tener que las matrices estándar asociadas a las transformaciones lineales de suma y producto por un escalar correspondan a las operaciones definidas entre las matrices. Concretamente tenemos el siguiente resultado.

Proposición 6.3 *Considérense*

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^m \quad y$$
$$S : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

transformaciones lineales con matrices estándar asociadas A y B, respectivamente. Sea c un escalar, entonces $A + B$ es la matriz estándar asociada a $T + S$ y cA es la matriz estándar asociada a cT. La suma y el producto por un escalar para transformaciones y matrices son los definidos anteriormente.

Demostración. Recordemos, del capítulo anterior, que las matrices estándar que representan a las transformaciones dadas están dadas como

$$A = [T(\vec{e}_1)\ T(\vec{e}_2) \cdots T(\vec{e}_n)]\,,$$
$$B = [S(\vec{e}_1)\ S(\vec{e}_2) \cdots S(\vec{e}_n)]\,,$$

es decir, las matrices cuyas columnas son las imágenes de los vectores de la base estándar bajo la transformación dada. Por definición, la matriz estándar asociada a $T + S$ está dada por

$$\begin{aligned}&[(T+S)(\vec{e}_1)\ (T+S)(\vec{e}_1) \cdots (T+S)(\vec{e}_n)]\\ &= [T(\vec{e}_1) + S(\vec{e}_1)\ T(\vec{e}_2) + S(\vec{e}_2) \cdots T(\vec{e}_n) + S(\vec{e}_n)]\\ &= [T(\vec{e}_1)\ T(\vec{e}_2) \cdots T(\vec{e}_n)] + [S(\vec{e}_1)\ S(\vec{e}_2) \cdots S(\vec{e}_n)]\\ &= A + B.\end{aligned}$$

Similarmente, la matriz asociada a cT ese calcula como

$$\begin{aligned}[(cT)(\vec{e}_1)\ (cT)(\vec{e}_2) \cdots (cT)(\vec{e}_n)] &= [cT(\vec{e}_1)\ cT(\vec{e}_2) \cdots cT(\vec{e}_n)]\\ &= c\,[T(\vec{e}_1)\ T(\vec{e}_2) \cdots T(\vec{e}_n)] = cA.\end{aligned}$$

■

El siguiente resultado se sigue fácilmente de las propiedades algebraicas de las operaciones de suma y multiplicación por un escalar definidas en $M_{m\times n}$.

Proposición 6.4 *El conjunto $M_{m\times n}$, junto con las operaciones de suma entre matrices y de producto por un escalar que acabamos de definir, satisface las siguientes propiedades:*

V1 Si $A, B \in M_{m\times n}$, entonces $A + B = B + A$ (conmutatividad de la suma).

V2 Si $A, B, C \in M_{m\times n}$, entonces $(A + B) + C = A + (B + C)$ (asociatividad de la suma).

V3 Si $A \in M_{m\times n}$*, entonces* $A+\Theta = A$ *(existencia de neutro aditivo).*

V4 Si $A \in M_{m\times n}$*, entonces* $A - A = \Theta$ *(existencia de inversos aditivos).*

V5 Si $A \in M_{m\times n}$*, entonces* $1A = A$*.*

V6 Si $A \in M_{m\times n}$ *y* β *y* γ *son escalares, entonces* $(\beta\gamma)A = \beta(\gamma A)$*.*

V7 Si $A \in M_{m\times n}$ *y* β *y* γ *son escalares, entonces* $(\beta+\gamma)A = \beta A+\gamma A$ *(primera forma de distributividad).*

V8 Si $A, B \in M_{m\times n}$ *y* β *es un escalar, entonces* $\beta(A+B) = \beta A+\beta B$ *(segunda forma de distributividad).*

Demostración. A manera de ejemplo probamos $(V3)$ y $(V7)$. En el caso de $(V3)$ notamos que

$$(A+\Theta)_{ij} = A_{ij} + \Theta_{ij} = A_{ij} + 0 = A_{ij},$$

por lo que las matrices $A+\Theta$ y A son iguales (pues coinciden entrada por entrada). De forma similar, en el caso de $(V7)$ tenemos que

$$((\beta+\gamma)A)_{ij} = (\beta+\gamma)A_{ij} = \beta A_{ij}+\gamma A_{ij} = (\beta A)_{ij}+(\gamma A)_{ij} = (\beta A+\gamma A)_{ij},$$

por lo que las matrices $(\beta+\gamma)A$ y $\beta A+\gamma A$ son iguales (pues coinciden entrada por entrada). ■

Estas ocho propiedades son idénticas a las dadas en la proposición 3.6 para los espacios vectoriales $\mathbb{R}^n$ y se denominan **propiedades de espacio vectorial**. En general, cualquier conjunto cuyos objetos satisfacen V1 a V8 será un espacio vectorial. De esta forma, se generalizan las propiedades fundamentales de los espacios $\mathbb{R}^n$ a estructuras más abstractas, que nada tienen que ver con conjuntos de "flechas" en $\mathbb{R}^n$. Por ejemplo, el conjunto $M_{m\times n}$ o equivalentemente el conjunto

$$\{T : \mathbb{R}^n \longrightarrow \mathbb{R}^m \mid T \text{ es una transformación lineal}\}$$

tienen la estructura de espacios vectoriales[1].

A partir de estas propiedades fundamentales se deducen otras más complejas. Estos resultados son inherentes a cualquier espacio vectorial (como es el caso de $\mathbb{R}^n$, $M_{m\times n}$, etcétera) y, de hecho, siempre son demostrados utilizando los mismos argumentos. Se tiene así que las Proposiciones 3.7 y 3.9, junto con el Corolario 3.8, también son válidas

[1] Así como también conjuntos de funciones del tipo

$$\{f : \Omega \longrightarrow \mathbb{R} \,|f \text{ es continua}\}.$$

en el nuevo entorno de matrices (o transformaciones lineales). A continuación enunciamos estos resultados. Se proporciona la demostración para la ley de la cancelación para que el lector se de cuenta que es idéntica al caso de $\mathbb{R}^n$ dado en el Capítulo 3.

Proposición 6.5 (Ley de la cancelación) *Si* $A, B, C \in M_{m\times n}$, *y* $A+B=A+C$, *entonces* $B=C$.

Demostración. Supongamos $A+B=A+C$. Sumando $-A$ a ambos lados de esta igualdad y aplicando las propiedades de asociatividad, del inverso y del neutro aditivo, obtenemos

$$\begin{aligned} B &= \Theta + B = (-A+A)+B = -A+(A+B) = -A+(A+C) \\ &= (-A+A)+C = \Theta + C = C. \end{aligned}$$

■

Corolario 6.6 *Si* $A, B \in M_{m\times n}$ *y* $A+B=A$, *entonces* $B=\Theta$.

Proposición 6.7 *Si* $A \in M_{m\times n}$ *y* β *es un escalar, entonces* $\beta A = \Theta$ *syss* $\beta = 0$ *o* $A=\Theta$.

Observemos que en lugar de matrices A, B, C en $M_{m\times n}$ podríamos haber utilizado transformaciones lineales $R.S, T$ de $\mathbb{R}^n$ en $\mathbb{R}^m$.

Proposición 6.8 *Si* A, $B \in M_{m\times n}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$ *y* $r \in \mathbb{R}$, *se cumplen las siguientes:*

1. $A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y}$.
2. $(A+B)\vec{x} = A\vec{x} + B\vec{x}$.
3. $(rA)\vec{x} = r(A\vec{x})$.
4. $A\vec{x} = B\vec{x}$ *para todo* $\vec{x} \in \mathbb{R}^n$ *syss* $A = B$.

Demostración. La demostración de (1) se realizó en el Lema 4.37 del Capítulo 4. En el caso de (2) Tenemos que si $A = [A^1\ A^2 \cdots A^n]$ y $B = [B^1\ B^2 \cdots B^n]$, entonces

$$\begin{aligned} (A+B)\vec{x} &= x_1\left(A^1+B^1\right) + x_2\left(A^2+B^2\right) + \cdots + x_n\left(A^n+B^n\right) \\ &= x_1A^1 + x_1B^1 + x_2A^2 + x_2B^2 + \cdots + x_nA^n + x_nB^n \\ &= \left(x_1A^1 + x_2A^2 + \cdots + x_nA^n\right) + \left(x_1B^1 + x_2B^2 + \cdots + x_nB^n\right) \\ &= A\vec{x} + B\vec{x}. \end{aligned}$$

Para demostrar (3) notamos que se cumple

$$\begin{aligned}(rA)\vec{x} &= x_1\left(rA^1\right) + x_2\left(rA^2\right) + \cdots + x_n\left(rA^n\right)\\ &= r\left(x_1A^1\right) + r\left(x_2A^2\right) + \cdots + r\left(x_nA^n\right)\\ &= r\left(x_1A^1 + x_2A^2 + \cdots + x_nA^n\right) = r(A\vec{x}).\end{aligned}$$

Finalmente, para (4), si $A\vec{x} = B\vec{x}$ para todo $\vec{x} \in \mathbb{R}^n$, en particular, para todo vector $\vec{e}_i$, $i = 1, 2, \ldots, n$, de la base estándar, se cumple

$$A^i = A\vec{e}_i = B\vec{e}_i = B^i.$$

Por lo tanto, $A = B$ pues ambas matrices tienen las mismas columnas. La otra implicación es inmediata. ■

La transpuesta de una matriz

Definición 6.9 *Si $A \in M_{m\times n}$, entonces* ***la matriz transpuesta de** A, denotada por A^T, se define como la matriz $A^T \in M_{n\times m}$, donde*

$$(A^T)_{ij} = A_{ji},$$

para toda $i \leq n$ y para toda $j \leq m$. En otras palabras, la matriz transpuesta de A es la matriz que se obtiene a partir de A convirtiendo (en el orden natural) sus renglones en columnas (o lo que es lo mismo sus columnas en renglones).

Observemos que si $A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$ es una matriz de $m \times 1$ y $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \end{bmatrix}$ es una matriz de $1 \times n$, entonces

$$\begin{aligned}A^T &= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \end{bmatrix} \in M_{1\times m},\\ B^T &= \begin{bmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{1n} \end{bmatrix} \in M_{n\times 1}.\end{aligned}$$

En general, dada $A \in M_{m\times n}$ podemos identificar a cualquier columna A^j y a cualquier renglón A_i como matrices de $m\times 1$ y $1\times n$, respectivamente, de manera que $\left(A^j\right)^T$ es un renglón con las mismas entradas que la columna A^j y $\left(A_i\right)^T$ es una columna con las mismas entradas que el renglón A_i. Con esta notación podemos escribir a la matriz transpuesta como

$$A^T = \begin{bmatrix} (A^1)^T \\ (A^2)^T \\ \vdots \\ (A^n)^T \end{bmatrix} = \begin{bmatrix} (A_1)^T & (A_2)^T & \cdots & (A_m)^T \end{bmatrix}.$$

Notemos que, por la definición de la transpuesta, se cumplen las siguientes igualdades:

$$\left(A^T\right)_j = \left(A^j\right)^T \text{ y } \left(A^T\right)^i = (A_i)^T. \tag{6.1}$$

Si

$$A = \begin{bmatrix} 3 & 4 & 5 \\ -4 & 2 & 1 \end{bmatrix},$$

entonces

$$A^T = \begin{bmatrix} 3 & -4 \\ 4 & 2 \\ 5 & 1 \end{bmatrix}.$$

Obsérvese que

$$(A_1)^T = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \ (A_2)^T = \begin{bmatrix} -4 \\ 2 \\ 1 \end{bmatrix},$$
$$\left(A^1\right)^T = \begin{bmatrix} 3 & -4 \end{bmatrix}, \ \left(A^2\right)^T = \begin{bmatrix} 4 & 2 \end{bmatrix}, \ \left(A^3\right)^T = \begin{bmatrix} 5 & 1 \end{bmatrix}.$$

El siguiente resultado enlista algunas de las propiedades más utilizadas de la matriz transpuesta. La demostración de todas ellas es inmediata.

Proposición 6.10 *Las siguientes afirmaciones son verdaderas*

1. *Si $A \in M_{m\times n}$, entonces $(A^T)^T = A$.*
2. *Si $A, B \in M_{m\times n}$, entonces $(A+B)^T = A^T + B^T$.*
3. *Si $A \in M_{m\times n}$ y r es un escalar, entonces $(rA)^T = rA^T$.*

Definición 6.11 *Sea $A \in M_{n\times n}$, decimos que A es una* ***matriz simétrica*** *si $A = A^T$.*

Ejemplo 6.12 *Si $A \in M_{m\times n}$ la matriz $A+A^T$ es simétrica. En efecto, utilizando los incisos (1) y (2) de la proposición anterior tenemos que:*

$$\left(A+A^T\right)^T = A^T + (A^T)^T = A^T + A = A + A^T.$$

Ejemplo 6.13 *Observar que si* $A \in M_{n\times n}$ *es una matriz simétrica, para todo* $i, j = 1, \ldots, n$ *debe cumplirse* $A_{ij} = A_{ji}$*, es decir, la matriz tiene simetría con respecto a su diagonal. Las siguientes matrices son simétricas:*

$$\begin{bmatrix} 0 & -1 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 4 & -5 \\ 4 & -1 & 3 \\ -5 & 3 & 666 \end{bmatrix}.$$

Ejemplo 6.14 *En cálculo vectorial se utiliza la matriz hessiana asociada a ciertas funciones reales. Esta es la matriz simétrica de las segundas derivadas como se ve a continuación. Si*

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

es una función doblemente diferenciable y con segundas derivadas continuas. Entonces la matriz hessiana está definida como

$$H(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix},$$

por continuidad de las derivadas, tenemos que $f_{xy}(x,y) = f_{yx}(x,y)$ *por lo que la matriz es simétrica. En general, si*

$$f : \in \mathbb{R}^n \longrightarrow \mathbb{R}$$

es doblemente diferenciable con segundas derivadas continuas, la hessiana (evaluada en $\vec{x} \in \mathbb{R}^n$*) está dada por*

$$H(\vec{x}) = \begin{bmatrix} f_{x_1x_1}(\vec{x}) & \cdots & f_{x_1x_n}(\vec{x}) \\ \vdots & \ddots & \vdots \\ f_{x_nx_1}(\vec{x}) & \cdots & f_{x_nx_n}(\vec{x}) \end{bmatrix}.$$

Por ejemplo, si $f(x,y) = x^2 + xy + y^3$*, entonces*

$$H(x,y) = \begin{bmatrix} 2 & 1 \\ 1 & 6y \end{bmatrix}.$$

Finalizamos esta sección con la siguiente observación.

Observación 6.15

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \in M_{1\times n}$$

y

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in M_{n\times 1},$$

éstas matrices pueden pensarse, respectivamente, como vectores renglón y columna en $\mathbb{R}^n$. *En este caso, las columnas de* A *consisten de un único elemento,* a_j, *y el producto* AB, *definido como en 3.28 del capítulo 3, es un escalar dado por,*

$$\begin{aligned} AB &= b_1a_1 + b_2a_2 + \cdots + b_na_n \\ &= a_1b_1 + a_2b_2 + \cdots + a_nb_n \in \mathbb{R}. \end{aligned}$$

El lector observador notará la similitud con el producto punto (o escalar) entre vectores definido en el Capítulo 3. Esto no es casualidad puesto que se cumple

$$AB = a_1b_1 + a_2b_2 + \cdots + a_nb_n = A \cdot B,$$

en donde A *y* B *se piensan como vectores* [2] *en* $\mathbb{R}^n$.

Ejemplo 6.16 *Sean* $A = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$ *y* $B = \begin{bmatrix} 1 \\ 0 \\ 666 \\ 7 \end{bmatrix}$ *vectores en* $\mathbb{R}^4$. *El producto punto* $A \cdot B$ *puede también escribirse como un producto de matrices, en efecto,*

$$A \cdot B = (-1)(1) + (2)(0) + (0)(666) + (1)(7) = 8 = A^T B.$$

Producto de matrices

En esta sección se define una operación nueva en $M_{m\times n}$ que llamamos producto de matrices. Para este fin, utilizamos como punto de partida la operación de composición entre transformaciones lineales. Sean

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^p, \; S : \mathbb{R}^p \longrightarrow \mathbb{R}^m$$

dos transformaciones lineales con matrices estándar asociadas A y B, respectivamente, y sea

$$S \circ T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

su composición. Es inmediato verificar que la función $S \circ T$ es también una transformación lineal. Por ejemplo, $S \circ T$ preserva sumas ya que

[2] Independientemente de que se trate de columnas o renglones, A y B pueden identificarse con los vectores $(a_1, \ldots, a_n)$ y $(b_1, \ldots, b_n)$ en $\mathbb{R}^n$.

dados $\vec{x}$, $\vec{y}$ vectores en $\mathbb{R}^n$, se tiene que

$$\begin{aligned}(S\circ T)(\vec{x}+\vec{y}) &= S(T(\vec{x}+\vec{y}))\\ &= S(T(\vec{x})+T(\vec{y}))\\ &= S(T(\vec{x}))+S(T(\vec{y}))\\ &= (S\circ T)(\vec{x})+(S\circ T)(\vec{y})\,.\end{aligned}$$

Análogamente se demuestra que para cualquier escalar c se cumple

$$(S\circ T)(c\vec{x}) = c\,(S\circ T)(\vec{x})\,.$$

Sean $A\in M_{m\times p}$ y $B\in M_{p\times n}$ las matrices estándar asociadas a S y T, respectivamente. Las siguientes igualdades son inmediatas para todo $\vec{x}\in\mathbb{R}^n$:

$$(S\circ T)(\vec{x}) = S(T(\vec{x})) = S(B\vec{x}) = A\,(B\vec{x})\,.$$

De lo anterior, es claro que lo deseable es definir el producto de las matrices AB de tal forma que

$$(AB)\vec{x} = A\,(B\vec{x})\,,$$

de manera que la matriz estándar asociada a la composición $S\circ T$ sea, precisamente, la matriz producto AB. Se tiene así la siguiente definición.

Definición 6.17 *Sean $A\in M_{m\times p}$, $B\in M_{p\times n}$ matrices y*

$$T:\mathbb{R}^n\longrightarrow\mathbb{R}^p,\;\; S:\mathbb{R}^p\longrightarrow\mathbb{R}^m$$

las transformaciones asociadas definidas como

$$T(\vec{x}) = B\vec{x},\;\; S(\vec{y}) = A\vec{y},$$

para cualesquiera $\vec{x}\in\mathbb{R}^n$ y $\vec{y}\in\mathbb{R}^p$. Entonces se define el producto AB como la matriz estándar de $m\times n$ asociada a la composición $S\circ T$. Es decir, si $\{\vec{e}_1,\vec{e}_2,\ldots,\vec{e}_n\}$ es la base estándar de $\mathbb{R}^n$, por construcción

$$AB = \begin{bmatrix} S(T(\vec{e}_1)) & S(T(\vec{e}_2)) & \cdots & S(T(\vec{e}_n)) \end{bmatrix}. \qquad (6.2)$$

La definición anterior nos dice que es necesario evaluar $S(T(\vec{e}_j))$ para toda $j=1,2,\ldots,n$. Esto puede ser laborioso ya que hay que encontrar las transformaciones S, T y $S\circ T$. Lo ideal sería poder calcular el producto AB a partir de las matrices A y B. Para este efecto, supongamos que

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p}\\ a_{21} & a_{22} & \ldots & a_{2p}\\ \vdots & & & \\ a_{m1} & a_{m2} & \ldots & a_{mp}\end{bmatrix},\; B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n}\\ b_{21} & b_{22} & \ldots & b_{2n}\\ \vdots & & & \\ b_{p1} & b_{p2} & \ldots & b_{pn}\end{bmatrix}.$$

De acuerdo a la definición dada por (6.2), la $j-$*ésima* columna del producto AB está dada por

$$\begin{aligned}(AB)^j &= (S \circ T)(\vec{e}_j) = S(T(\vec{e}_j))\\ S(B\vec{e}_j) &= S\left(B^j\right) = AB^j.\end{aligned}$$

en donde la penúltima igualdad se cumple en virtud de que $B^j = B\vec{e}_j$. Hemos demostrado así la siguiente proposición, misma que proporciona una definición alternativa del producto AB, sin apelar a las transformaciones lineales asociadas.

Proposición 6.18 *Si $A \in M_{m\times p}$ y $B \in M_{p\times n}$, entonces el producto $AB \in M_{m\times p}$ satisface que, para cada $j \in \{1, \ldots n\}$, la $j-$ésima columna de AB es una combinación lineal de las columnas de A. Concretamente,*

$$(AB)^j = b_{1j}A^1 + b_{2j}A^2 + \cdots + b_{pj}A^p = AB^j.$$

Podemos refinar aún más el procedimiento anterior para obtener explícitamente cada entrada de la columna $(AB)^j$. En efecto, como

$$(AB)^j = AB^j = b_{1j}A^1 + b_{2j}A^2 + \cdots + b_{pj}A^p$$

y las columnas de A están dadas por

$$A^k = \begin{bmatrix} a_{1k}\\ a_{2k}\\ \vdots\\ a_{mk}\end{bmatrix},$$

tenemos que, utilizando las consideraciones de la Observación 6.15,

$$\begin{aligned}(AB)^j &= b_{1j}\begin{bmatrix} a_{11}\\ a_{21}\\ \vdots\\ a_{m1}\end{bmatrix} + b_{2j}\begin{bmatrix} a_{12}\\ a_{22}\\ \vdots\\ a_{m2}\end{bmatrix} + \cdots + b_{pj}\begin{bmatrix} a_{1p}\\ a_{2p}\\ \vdots\\ a_{mp}\end{bmatrix}\\ &= \begin{bmatrix} b_{1j}a_{11} + b_{2j}a_{12} + \cdots + b_{pj}a_{1p}\\ b_{1j}a_{21} + b_{2j}a_{22} + \cdots + b_{pj}a_{2p}\\ \vdots\\ b_{1j}a_{m1} + b_{2j}a_{m2} + \cdots + b_{pj}a_{mp}\end{bmatrix}\\ &= \begin{bmatrix} A_1B^j\\ A_2B^j\\ \vdots\\ A_mB^j\end{bmatrix} = \begin{bmatrix} A_1\cdot B^j\\ A_2\cdot B^j\\ \vdots\\ A_m\cdot B^j\end{bmatrix}.\end{aligned}$$

Observamos que la entrada en el $i-$*ésimo* renglón de la columna $(AB)^j$ es simplemente

$$(AB)_{ij} = A_i B^j = A_i \cdot B^j.$$

La siguiente proposición se sigue de la discusión anterior y nos proporciona una forma de calcular cada entrada de la matriz producto de dos matrices,

Proposición 6.19 *Si $A \in M_{m\times p}$ y $B \in M_{p\times n}$, las entradas de la matriz producto AB, están dadas como*

$$(AB)_{ij} = A_i B^j = A_i \cdot B^j,$$

para toda $i \leq m$ y para toda $j \leq n$.

Observación 6.20 *La entrada ij del producto AB se define como el producto del $i-$ésimo renglón de A (que identificamos con una matriz de $1 \times p$) con la columna $j-$ésima de B (que se identifica con una matriz de $p\times 1$). El producto $A_i B^j$ es un número real (equivalentemente, una matriz de 1×1) igual al producto punto $A_i \cdot B^j$. Por construcción, el producto AB que hemos definido generaliza el producto $A\vec{x}$ de una matriz por un vector*[3]*.*

Observación 6.21 *El producto matricial AB está definido syss el número de columnas de A es igual al número de renglones de B. En tal caso, el número de renglones de AB es igual al número de renglones de A, mientras que el número de columnas de AB es igual al número de columnas de B.*

La Proposición 6.18 nos proporciona una descripción de las columnas de la matriz producto. Una pregunta natural es si puede hacerse algo similar con los renglones de la matriz producto. La siguiente proposición responde a esta pregunta en afirmativo.

Proposición 6.22 *Si $A \in M_{m\times n}$ y $B \in M_{n\times p}$, entonces el producto $AB \in M_{m\times p}$ satisface que, para cada $i \in \{1, \dots m\}$, el $i-$ésimo renglón de AB está dado como*

$$(AB)_i = A_i B.$$

Demostración. Sea $i \in \{1, \dots m\}$. La Proposición 6.19 nos dice que la entrada en la $j-$*ésima* columna del renglón $(AB)_i$ está dada por

$$A_i B^j.$$

Entonces,

$$(AB)_i = \left[A_i B^1 \; A_i B^2 \cdots A_i B^p\right] = A_i B.$$

■

[3]Simplemente se piensa al vector $\vec{x}$ como una matriz con una sóla columna.

Ejemplo 6.23 *Consideremos a las matrices*

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 3 & 0 & -2 \end{bmatrix}, \; B = \begin{bmatrix} 2 & 0 \\ -1 & 0 \\ 4 & 1 \end{bmatrix}, \; C = \begin{bmatrix} 6 & 7 \\ -1 & 8 \end{bmatrix},$$

$$D = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \; E = \begin{bmatrix} 1 & -1 \end{bmatrix}.$$

y supongamos que deseamos calcular los productos AB, AC, AD y DE. En el primer caso notamos que $A \in M_{2\times 3}$ y $B \in M_{3\times 2}$, por lo que $AB \in M_{2\times 2}$. Además,

$$(AB)_{11} = A_1 B^1 = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = 4 - 1 - 4 = -1,$$

$$(AB)_{12} = A_1 B^2 = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = -1,$$

$$(AB)_{21} = A_2 B^1 = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix} = 6 - 8 = -2,$$

$$(AB)_{22} = A_2 B^2 = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = -2,$$

Por lo tanto,

$$AB = \begin{bmatrix} -1 & -1 \\ -2 & -2 \end{bmatrix}.$$

En el segundo caso, notamos que el número de columnas de A es distinto al número de renglones de C, por lo que el producto AC no está definido. En el tercer caso notamos que $A \in M_{2\times 3}$ y que $D \in M_{3\times 1}$, por lo que $AD \in M_{2\times 1}$. Además,

$$(AD)_{11} = \begin{bmatrix} 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 8 - 1 = 7,$$

$$(AD)_{21} = \begin{bmatrix} 3 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 12 - 2 = 10.$$

Por lo tanto,

$$AD = \begin{bmatrix} 7 \\ 10 \end{bmatrix}.$$

Alternativamente, como D es una matriz de 3×1 y puede pensarse como un vector en $\mathbb{R}^3$, AD también puede calcularse como

$$AD = 4\begin{bmatrix}2\\3\end{bmatrix} + 0\begin{bmatrix}1\\0\end{bmatrix} + 1\begin{bmatrix}-1\\-2\end{bmatrix} = \begin{bmatrix}7\\10\end{bmatrix}.$$

Finalmente, el producto DE está definido pues $D \in M_{3\times 1}$ y $E \in M_{1\times 2}$. Obsérvese que los renglones de D y las columnas de E son simplemente números reales. Las entradas del producto $DE \in M_{3\times 2}$ se calculan como sigue:

$$\begin{aligned}
&(DE)_{11} = D_1E^1 = 4 \times 1 = 4,\ (DE)_{12} = D_1E^2 = 4 \times (-1) = -4,\\
&(DE)_{21} = D_2E^1 = 0 \times 1 = 0,\ (DE)_{22} = D_2E^2 = 0 \times (-1) = 0,\\
&(DE)_{31} = D_3E^1 = 1 \times 1 = 1,\ (DE)_{32} = D_3E^2 = 1 \times (-1) = -1,
\end{aligned}$$

de manera que

$$DE = \begin{bmatrix}4\\0\\1\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} = \begin{bmatrix}4 & -4\\0 & 0\\1 & -1\end{bmatrix}.$$

Como $C \in M_{2\times 2}$ y $A \in M_{2\times 3}$ tendremos que $CA \in M_{2\times 3}$, en efecto,

$$\begin{aligned}
CA &= \begin{bmatrix}6 & 7\\-1 & 8\end{bmatrix}\begin{bmatrix}2 & 1 & -1\\3 & 0 & -2\end{bmatrix}\\
&= \begin{bmatrix}(6)(2)+(7)(3) & (6)(1)+(7)(0) & (6)(-1)+(7)(-2)\\(-1)(2)+(8)(3) & (-1)(1)+(8)(0) & (-1)(-1)+(8)(-2)\end{bmatrix}\\
&= \begin{bmatrix}33 & 6 & -20\\22 & -1 & -15\end{bmatrix}.
\end{aligned}$$

Una vez calculado el producto CA podríamos preguntarnos acerca del producto AC. En este caso, ¡AC ni siquiera está definido! La razón es que el número de columnas de A (3) no coincide con el número de renglones de C (2).

Ejemplo 6.24 *Cuando A y B son matrices cuadradas de $n\times n$, ambos productos, AB y BA, están definidos y son matrices de $n \times n$. Por ejemplo, sean*

$$A = \begin{bmatrix}-2 & -1\\1 & 0\end{bmatrix} \text{ y } B = \begin{bmatrix}3 & -1\\-1 & 3\end{bmatrix},$$

entonces

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

$$\begin{bmatrix} (-2)(3) + (-1)(-1) & (-2)(-1) + (-1)(3) \\ (1)(3) + (0)(-1) & (1)(-1) + (0)(3) \end{bmatrix} = \begin{bmatrix} -5 & -1 \\ 3 & -1 \end{bmatrix},$$

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

$$\begin{bmatrix} (3)(-2) + (-1)(1) & (3)(-1) + (-1)(0) \\ (-1)(-2) + (3)(1) & (-1)(-1) + (3)(0) \end{bmatrix} = \begin{bmatrix} -7 & -3 \\ 5 & 1 \end{bmatrix}.$$

Observemos que $AB \neq BA$, de manera que el producto de matrices no es conmutativo, aún si las matrices son cuadradas y del mismo tamaño.

Ejemplo 6.25 *Consideremos a las matrices*

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

y supongamos que deseamos calcular la tercera columna del producto AB. Una forma eficiente de hacerlo (en lugar de calcular toda la matriz AB) es simplemente utilizando la Proposición 6.18. De esta forma,

$$(AB)^3 = AB^3 = 3 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 6 \begin{bmatrix} 3 \\ 6 \end{bmatrix} + 9 \begin{bmatrix} 4 \\ 7 \end{bmatrix}.$$

El siguiente resultado enlista algunas propiedades del producto de matrices.

Ejemplo 6.26 *Considérense las transformaciones S y T dadas como sigue.*

$$\mathbb{R}^3 \xrightarrow{T} \mathbb{R}^2 \xrightarrow{S} \mathbb{R}^2,$$

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x + y - z \\ 2x + z \end{bmatrix},$$

$$S\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x - y \\ x + 2y \end{bmatrix}.$$

Las matrices estándar asociadas son, respectivamente,

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \quad y \quad B = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}.$$

Podemos calcular la composición

$$S \circ T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$$

de dos formas: la primera es directamente, sin utilizar las matrices. En este caso se tiene

$$\begin{aligned}(S \circ T)\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) &= S\left(T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)\right) = S\left(\begin{bmatrix} x+y-z \\ 2x+z \end{bmatrix}\right) \\ &= \begin{bmatrix} 2(x+y-z)-(2x+z) \\ (x+y-z)+2(2x+z) \end{bmatrix} = \begin{bmatrix} 2y-3z \\ 5x+y+z \end{bmatrix}.\end{aligned}$$

La segunda forma de encontrar $S \circ T$ es utilizando el producto AB como sigue.

$$\begin{aligned}(S \circ T)\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) &= \underbrace{\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}}_{B}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= \begin{bmatrix} 0 & 2 & -3 \\ 5 & 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2y-3z \\ 5x+y+z \end{bmatrix}.\end{aligned}$$

Ejemplo 6.27 *Sean $T_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ y $T_2 : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ las transformaciones dadas por*

$$T_1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-z \\ y+z \end{bmatrix}, \quad T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 2x \\ 0 \end{bmatrix}.$$

Entonces $T_2 \circ T_1 : \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ está dada por

$$(T_2 \circ T_1)\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = T_2\left(T_1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right)\right) = T_2\left(\begin{bmatrix} x-z \\ y+z \end{bmatrix}\right) = \begin{bmatrix} x-z \\ y+z \\ 2(x-z) \\ 0 \end{bmatrix}.$$

Por lo tanto, la matriz estándar que representa a $T_2 \circ T_1$ es la matriz de 4×3 cuya primera, segunda y tercera columna son

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

Proposición 6.28 *Las siguientes afirmaciones son verdaderas*

1. *El producto de matrices no es conmutativo.*
2. *El producto de matrices es asociativo. Esto es, si* $D \in M_{m\times n}$, $E \in M_{n\times p}$ *y* $F \in M_{p\times q}$, *entonces* $D(EF) = (DE)F$.
3. *Si* $A, B \in M_{m\times n}$ *y* $C \in M_{n\times p}$, *entonces* $(A+B)C = AC + BC$.
4. *Si* $A \in M_{m\times n}$ *y* $B, C \in M_{n\times p}$, *entonces* $A(B+C) = AB + AC$.
5. *Si* $A \in M_{m\times n}$, I_m *denota a la matriz identidad de* $m\times m$, *entonces* $I_m A = A$.
6. *Si* $A \in M_{m\times n}$, I_n *denota a la matriz identidad de* $n \times n$, *entonces* $AI_n = A$.
7. *Si* $A \in M_{m\times n}$ *y* $B \in M_{n\times p}$, *entonces* $(AB)^T = B^T A^T$.

Demostración. La demostración de (1) se sigue del Ejemplo 6.24 las propiedades (2). (5) y (6) son inmediatas recordando que el producto de las matrices correponde a la composición de las transformaciones lineales correspondientes. Las propiedades (3) y (4) se siguen de los incisos (2) y (4) de la Proposición 6.8 y de la definición del producto de matrices. Para ilustrar, demostramos la propiedad (3): dado $\vec{x} \in \mathbb{R}^p$ se cumple

$$\begin{aligned}(A+B)C\vec{x} &= A(C\vec{x}) + B(C\vec{x})\\ &= (AC)\,\vec{x} + (BC)\,\vec{x}\\ &= (AC+BC)\,\vec{x},\end{aligned}$$

por lo que $(A+B)C = AC + BC$.

En el caso de (7), utilizamos las propiedades de matrices transpuestas, la Proposición 6.19 y la conmutatividad del producto punto para obtener:

$$\begin{aligned}((AB)^T)_{ij} &= (AB)_{ji} = A_j \cdot B^i = B^i \cdot A_j\\ &= \left(B^T\right)_i \cdot (A^T)^j = (B^T A^T)_{ij},\end{aligned}$$

por lo que las matrices $(AB)^T$ y $B^T A^T$ son iguales (pues coinciden entrada por entrada). ■

Ejemplo 6.29 *Sean*

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 5 & -2 \\ 1 & 1 \end{bmatrix},$$

entonces

$$AB = \begin{bmatrix} 4 & -3 \\ 10 & -4 \\ 16 & -5 \end{bmatrix}.$$

Se tiene así que

$$\begin{aligned} (AB)^T &= \begin{bmatrix} 4 & 10 & 16 \\ -3 & -4 & -5 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \end{bmatrix} \\ &= B^T A^T. \end{aligned}$$

Observemos que el producto $A^T B^T$ ni siquiera está definido.

Otras matrices de interés

Matrices diagonales

Consideremos el conjunto de matrices cuadradas $M_{n\times n}$. Un subconjunto que se utiliza recurrentemente es el conjunto de las matrices diagonales definido como sigue.

Definición 6.30 *Una matriz $D \in M_{n\times n}$ es* ***diagonal*** *syss $D_{ij} = 0$ para toda $i \neq j$. Es decir, todos las entradas fuera de la diagonal son nulas.*

Ejemplo 6.31 *La matriz identidad I_n es diagonal así como también lo es la matriz cero Θ. Las matrices*

$$\begin{bmatrix} 666 & 0 \\ 0 & \pi \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

son diagonales.

Observación 6.32 *Las matrices diagonales son simétricas.*

Observación 6.33 *Si $C, D \in M_{n\times n}$ son diagonales, entonces CD es diagonal y $CD = DC$, es decir, el producto entre matrices diagonales es conmutativo. Esto es simple de verificar ya que*

$$(CD)_{ij} = \begin{cases} C_{ii}D_{jj} \text{ si } i = j, \\ 0 \text{ si } i \neq j. \end{cases}$$

Observación 6.34 *Evidentemente, si $C, D \in M_{n\times n}$ son diagonales, $C+D$ es diagonal. Adicionalmente, si C es diagonal y todas sus entradas en la diagonal son no nulas, la matriz D cuyas entradas diagonales son $D_{ii} = \frac{1}{C_{ii}}$ será un inverso multiplicativo de C. Veamos un ejemplo.*

Ejemplo 6.35 *Si*

$$C = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{bmatrix},$$

entonces la matriz

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

es tal que

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

La matriz D es, de facto, inversa multiplicativa de C. Este ejemplo nos da la pauta para construir inversos multiplicativos de una matriz diagonal para la cual $C_{ii} \neq 0$ para toda $i = 1, \ldots, n$. Observar que esto es equivalente a que sus columnas sean LI.

Las observaciones anteriores implican que el conjunto de matrices diagonales en $M_{n\times n}$ es cerrado bajo la suma, el producto (que es conmutativo), contiene neutros aditivos y multiplicativos e inversos aditivos. Además, si sus columnas son LI, también posee inversos multiplicativos Ahora bien, dada una matriz $A \in M_{n\times n}$, no necesariamente diagonal, ¿cuándo tendrá ésta un inverso multiplicativo? En el siguiente capítulo abordaremos precisamente este tema.

Matrices elementales

Dentro de las matrices de $n \times n$ existe una clase destacada cuyos elementos se conocen como matrices elementales. Estas matrices son útiles para representar las operaciones elementales entre renglones como productos de matrices como veremos más adelante. Su definición es sumamente simple y es la siguiente.

Definición 6.36 *Una matriz $E \in M_{n\times n}$ es una* ***matriz elemental*** *si E puede obtenerse a partir de la matriz identidad I realizando una (y* *sólo una) operación elemental entre renglones (de los tres tipos descritos en la Definición 2.15).*

Ejemplo 6.37 *La matriz*

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

es una matriz elemental de 2×2 *puesto que se obtiene a partir de la identidad intercambiando los renglones.*

Ejemplo 6.38 *La matriz*

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

es una matriz elemental de 3×3*, ya que se obtiene de la identidad multiplicando el segundo renglón por* π.

Ejemplo 6.39 *La matriz*

$$E = \begin{bmatrix} 1 & 0 & 666 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

es una matriz elemental de 3×3 *pues se obtiene de la identidad sumando al primer renglón* 666 *veces el tercero.*

Ejemplo 6.40 *La matriz*

$$F = \begin{bmatrix} 0 & -7 \\ 1 & 0 \end{bmatrix}$$

no es una matriz elemental puesto que no puede obtenerse de la identidad utilizando una sóla operación elemental entre sus renglones.

Consideremos a la matriz

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

y a las matrices elementales

$$E(1) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ E(2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 666 \end{bmatrix},$$

$$E(3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 666 \\ 0 & 0 & 1 \end{bmatrix},$$

en donde $E(1)$ se obtiene de la identidad intercambiando los renglones 1 y 2, $E(2)$ se obtiene de la identidad multiplicando el tercer renglón por 666 y $E(3)$ se obtiene de la identidad sumando 666 veces el tercer renglón al segundo. Observemos los siguientes productos

$$\begin{aligned}
E(1)A &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}, \\
E(2)A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 666 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -1 \\ 1 & 2 \\ 666 & 666 \end{bmatrix}, \\
E(3)A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 666 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}. \\
&= \begin{bmatrix} 0 & -1 \\ 667 & 668 \\ 1 & 1 \end{bmatrix}
\end{aligned}$$

El Ejemplo 6 muestra que multiplicar, por la izquierda, a la matriz A por la matriz elemental $E(i)$, $i = 1, 2, 3$, es equivalente a realizar la operación elemental utilizada para obtener $E(i)$ a partir de I, sobre los renglones de A. Esta propiedad de las matrices elementales puede enunciarse como sigue.

Proposición 6.41 *Sean $A \in M_{m\times n}$, $E \in M_{m\times m}$ tales que E es una matriz elemental. Entonces EA es una matriz que se obtiene de A realizando la misma operación elemental sobre los renglones de I que se utilizó para obtener E.*

Demostración. Observemos que la matriz identidad de $m \times m$ tiene como renglones a los vectores $\vec{e}_1^{\,T}, \vec{e}_2^{\,T}, \ldots, \vec{e}_m^{\,T}$ de la base estándar de $\mathbb{R}^m$. Adicionalmente, si E es una matriz elemental de $m \times m$, se tiene que, por la Proposición 6.22,

$$E_i A = A_i. \qquad (\spadesuit)$$

Para el primer tipo de operación elemental, supongamos que $E(1)$ se obtiene de I intercambiando los renglones $\vec{e}_k$ y $\vec{e}_l$. Entonces, utilizando

(♠), tenemos que los renglones del producto $E(1)A$ están dados como

$$(E(1)A)_i = \begin{cases} \vec{e}_i^T A = A_i \text{ si } i \neq k, l, \\ \vec{e}_l^T A = A_l \text{ si } i = k, \\ \vec{e}_k^T A = A_k \text{ si } i = l. \end{cases}$$

Así, la matriz $E(1)A$ se obtiene de A intercambiando los renglones k y l. Para el segundo tipo de operación elemental, sea $E(2)$ la matriz elemental que se obtiene de I multiplicando el renglón k por el escalar $c \neq 0$. Como antes, los renglones del producto $E(2)A$ son,

$$(E(2)A)_i = \begin{cases} \vec{e}_i^T A = A_i \text{ si } i \neq k, \\ c\vec{e}_k^T A = cA_k \text{ si } i = k, \end{cases}$$

y $E(2)A$ se obtiene de A multiplicando el $k - \textit{ésimo}$ renglón por el escalar c. Finalmente, para el tercer tipo de operación elemental, sea $E(3)$ la matriz elemntal que se obtiene de I sumando c veces el renglón l al renglón k. Los renglones del producto $E(3)A$ están dados por

$$(E(3)A)_i = \begin{cases} \vec{e}_i^T A = A_i \text{ si } i \neq k, \\ \left(c\vec{e}_l^T + \vec{e}_k^T\right) A = cA_l + A_k \text{ si } i = k, \end{cases}$$

es decir, $E(3)A$ se obtiene de A sumando c veces el renglón l al renglón k. ■

Corolario 6.42 *Sea $A \in M_{m\times n}$ y $\tilde{A}$ su FER, entonces existen matrices elementales $E(1), E(2), \ldots, E(k) \in M_{m\times m}$ tales que*

$$\tilde{A} = E(k)E(k-1)\ldots E(1)A.$$

Ejercicio 6.1 *Sea*

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

podemos obtener su FER como

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & -1 & 3 \\ 2 & -1 & 1 \end{bmatrix} \overset{-R_1+R_2}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 2 & -1 & 1 \end{bmatrix} \overset{-2R_1+R_3}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix} \overset{3R_2+R_3}{\sim}$$

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \overset{-2R_2+R_1}{\sim} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \overset{R_3+R_1}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \overset{R_2\leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Alternativamente, podemos multiplicar secuencialmente por las matrices elementales correspondientes a cada operación entre los renglones, *es decir,*

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{-R_1+R_2} \begin{bmatrix} 1 & -1 & 2 \\ 1 & -1 & 3 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 2 & -1 & 1 \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}}_{-2R_1+R_3} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}}_{3R_2+R_3} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{-2R_2+R_1} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{R_3+R_1} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{R_2\leftrightarrow R_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Concluimos esta sección analizando las matrices elementales transpuestas. Supongamos que la matriz elmental $E(1)$ se obtiene de la identidad intercambiando los renglones k y l, en donde $k < l$, es decir

$$I \overset{R_k\leftrightarrow R_l}{\sim} E(1).$$

En forma análoga, obtenemos $E(2)$ y $E(3)$ a partir de I como sigue

$$\begin{aligned} I &\overset{aR_k}{\sim} E(2), \\ I &\overset{cR_k+R_l}{\sim} E(3), \end{aligned}$$

en donde $c \in \mathbb{R}$, $c \neq 0$. Como los renglones y las columnas de I son los vectores de la base canónica $\vec{e}_1, \ldots, \vec{e}_n$, es un ejercicio sencillo demostrar que

$$\begin{aligned} E(1)^T &= E(1), \\ E(2)^T &= E(2). \end{aligned}$$

En el caso de $E(3)$, esta matriz es idéntica a la identidad excepto que $E(3)_{lk} = b$. Así, se tiene que su transpuesta es idéntica a la identidad excepto por la entrada kl que está dada por $E(3)_{kl} = b$. Podemos expresar esto como

$$I \overset{bR_l+R_k}{\sim} E(3)^T,$$

es decir, $E(3)^T$ es una matriz elemental del mismo tipo que $E(3)$.

Ejemplo 6.43 *Observemos las transpuestas de las siguientes matrices elementales.*

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Capítulo 7

Inversas

Introducción

Como vimos en el capítulo anterior, las operaciones entre matrices corresponden a operaciones análogas entre sus transformaciones lineales asociadas. Recordemos que dados cualesquiera conjuntos $\mathcal{A}$ y $\mathcal{B}$ y una función $f : \mathcal{A} \longrightarrow \mathcal{B}$, bajo ciertas condiciones (si f es inyectiva y suprayectiva), existe la llamada función inversa denotada por f^{-1}. Ésta última satisface

$$\begin{gathered} f^{-1} : \mathcal{B} \longrightarrow \mathcal{A}, \\ \left(f^{-1} \circ f\right)(x) = x, \quad \left(f \circ f^{-1}\right)(y) = y, \end{gathered}$$

de manera que $f^{-1} \circ f$ y $f \circ f^{-1}$ corresponden a las funciones identidad

$$\begin{gathered} I_{\mathcal{A}} : \mathcal{A} \longrightarrow \mathcal{A}, \\ I_{\mathcal{B}} : \mathcal{B} \longrightarrow \mathcal{B}. \end{gathered}$$

Una pregunta natural es si las transformaciones lineales, o equivalentemente sus matrices asociadas, pueden tener inversas y, de ser así, bajo qué circunstancias existen estas inversas. Comenzaremos con examinar la existencia de inversos multiplicativos para las matrices para relacionarlo, posteriormente, con los inversos de sus transformaciones lineales correspondientes.

Inversos multiplicativos para matrices

Si deseamos que una matriz A tenga un inverso multiplicativo, digamos B, de manera que tanto AB como BA coincidan con la matriz identidad, ambas matrices deben ser cuadradas y del mismo tamaño para que los productos estén bien definidos. Tenemos así la siguiente definición.

Definición 7.1 *Sea $A \in M_{n\times n}$ una matriz cuadrada. Decimos que A es **invertible** syss existe una matriz $B \in M_{n\times n}$ tal que*

$$AB = I = BA.$$

Observación 7.2 *Cuando no exista confusión, denotaremos a la matriz identidad simplemente por I, al igual que su transformación lineal asociada.*

Ejemplo 7.3 *Como se vio en el capítulo anterior, dada la matriz diagonal*

$$A = \begin{bmatrix} \pi & 0 \\ 0 & 666 \end{bmatrix},$$

se tiene que

$$B = \begin{bmatrix} \pi^{-1} & 0 \\ 0 & 666^{-1} \end{bmatrix}$$

satisface $AB = I = BA$.

Ejemplo 7.4 *Consideremos a la matriz*

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

Esta matriz es invertible pues

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

Ejemplo 7.5 *Consideremos a la matriz*

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Entonces A es invertible pues si hacemos

$$B = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix},$$

es sencillo verificar que $AB = I = BA$.

Ejemplo 7.6 *Sea*

$$A = \begin{bmatrix} 1 & 0 & -1 \\ \pi & 666 & 7 \end{bmatrix},$$

esta matriz no es invertible puesto que $A \in M_{2\times 3}$, es decir, no es cuadrada.

El siguiente resultado nos dice que si la matriz inversa B existe, ésta será un verdadero inverso multiplicativo (se cumple la unicidad de la matriz B).

Proposición 7.7 *Si A es una matriz invertible, entonces existe una* ***única*** *matriz B tal que $AB = I = BA$. Tal matriz B es conocida como* ***la inversa de*** *A y* ***es denotada por*** *A^{-1}.*

Demostración. Probaremos algo más general. De hecho, veremos que si A tiene una inversa por la derecha B (esto es, $AB = I$) y una inversa por la izquierda C (esto es, $CA = I$), entonces $C = B$. En efecto,

$$B = IB = (CA)B = C(AB) = CI = C.$$

Por lo tanto, si existe una matriz A^{-1} que es simultáneamente inversa derecha e izquierda, ésta debe ser única. ■

A continuación veremos que si $A \in M_{n\times n}$ tiene una inversa por la izquierda, entonces tendrá una inversa por la derecha y, como acabamos de ver, ambas coinciden y son iguales a A^{-1}. En otras palabras, A es invertible syss A tiene un inverso izquierdo (más adelante veremos que podemos sustituir "izquierdo" por "derecho").

Lema 7.8 *Si $A \in M_{n\times n}$ tiene una inversa por la izquierda, es decir, existe $B \in M_{n\times n}$ tal que $BA = I$, entonces las columnas de A forman un conjunto LI (o lo que es lo mismo, la FER de A es la matriz identidad).*

Demostración. Consideremos el sistema lineal

$$A\vec{x} = \vec{0}$$

Debemos probar que

$$\vec{x} = \vec{0}.$$

Ahora bien, multiplicando por la izquierda ambos lados de $A\vec{x} = \vec{0}$ por la matriz B, obtenemos que

$$\vec{x} = I\vec{x} = (BA)\vec{x} = B(A\vec{x}) = B\vec{0} = \vec{0}.$$

■

Ejemplo 7.9 *Sea*

$$A = \begin{bmatrix} 1 & -3 \\ -1 & 3 \end{bmatrix},$$

Esta es una matriz cuadrada de 2×2; *sin embargo, como sus columnas son LD, no puede tener inversa por la izquierda y, por lo tanto, no es invertible.*

Proposición 7.10 *Si* $A \in M_{n\times n}$ *tiene una inversa por la izquierda, entonces* A *tendrá una inversa por la derecha y ambas coinciden. Es decir,* A *es invertible.*

Demostración. Por el lema anterior tenemos que el conjunto de columnas de A, $\{A^1, A^2, \ldots, A^n\}$, es linealmente independiente. Por el teorema 5.40, A representa una transformación lineal suprayectiva. Así, dado cualquier $\vec{b} \in \mathbb{R}^n$, el sistema $A\vec{x} = \vec{b}$ tiene solución. En particular, si tomamos los vectores $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ de la base estándar en $\mathbb{R}^n$, tendremos que existen $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ en $\mathbb{R}^n$ de tal suerte que

$$A\vec{x}_1 = \vec{e}_1,\ A\vec{x}_2 = \vec{e}_2, \ldots, A\vec{x}_n = \vec{e}_n.$$

Sea C la matriz cuyas columnas son $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$; entonces, utilizando el resultado del capítulo anterior que nos dice que $(AC)^j = AC^j$, tenemos que, como $C^j = \vec{x}_j$ y $A\vec{x}_j = \vec{e}_j$,

$$AC = A\,[\vec{x}_1\ \vec{x}_2\ \cdots\ \vec{x}_n] = [\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n] = I,$$

de manera que C es inverso derecho de A. En particular, la demostración de la proposición 7.7 nos dice que si B es la inversa izquierda de A, entoncas $B = C = A^{-1}$. ■

Utilizando el mismo argumento que el empleado en la proposición 7.8 se obtiene un método, conceptualmente sencillo, para resolver sistemas de ecuaciones de $n\times n$ (para el caso en el cual la correspondiente matriz de coeficientes sea invertible).

Proposición 7.11 *Sea* $A \in M_{n\times n}$ *una matriz invertible,* $\vec{b} \in \mathbb{R}^n$ *y consideremos el sistema lineal* $A\vec{x} = \vec{b}$. *Entonces este sistema tiene solución única y más aún,*

$$\vec{x} = A^{-1}\vec{b}.$$

.

Demostración. La demostración es inmediata, multiplicando ambos lados de la expresión $A\vec{x} = \vec{b}$ por A^{-1}, como se muestra a continuación

$$A^{-1}A\vec{x} = (A^{-1}A)\vec{x} = Ix = x = A^{-1}\vec{b}.$$

■

Ejemplo 7.12 *Sea*

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

y supongamos que queremos resolver los sistemas

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad A \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

utilizando que

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

En el primer caso multiplicamos (por la izquierda) ambos lados de la ecuación matricial en cuestión por A^{-1}, obteniendo que

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = 1 \begin{bmatrix} -2 \\ \frac{3}{2} \end{bmatrix} + 4 \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ -\frac{1}{2} \end{bmatrix}.$$

Análogamente, en el segundo caso multiplicamos (por la izquierda) ambos lados de la ecuación matricial en cuestión por A^{-1} obteniendo que

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 5 \end{bmatrix} = 0 \begin{bmatrix} -2 \\ \frac{3}{2} \end{bmatrix} + 5 \begin{bmatrix} 1 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 5 \\ -\frac{5}{2} \end{bmatrix}.$$

Observación 7.13 *Los ejemplos anteriores muestran que el método de multiplicar por la matriz inversa es una forma "simple" para resolver algunos sistemas de ecuaciones de $n \times n$. Sin embargo, ¡nada es gratis en esta vida! El detalle está en que tenemos que conocer de antemano la matriz inversa y esto requiere, como siempre, de algo de esfuerzo. Más adelante veremos un método para encontrar esta matriz.*

Las demostraciones de los siguientes resultados apelan fuertemente al hecho de que si la inversa existe, ésta debe ser única.

Proposición 7.14 *Si $A \in M_{n\times n}$ es invertible, entonces A^{-1} también lo es y $(A^{-1})^{-1} = A$.*

Demostración. La demostración es muy simple notando que

$$AA^{-1} = I = A^{-1}A$$

y, por lo tanto, A es la inversa de A^{-1}, es decir, $A = (A^{-1})^{-1}$. ■

Proposición 7.15 *Si $A, B \in M_{n\times n}$ son invertibles, entonces el producto AB también lo es y $(AB)^{-1} = B^{-1}A^{-1}$.*

Demostración. Por unicidad del inverso, basta probar que

$$(B^{-1}A^{-1})(AB) = I.$$

Pero esto es claro ya que por asociatividad del producto,

$$(B^{-1}A^{-1})(AB) = B^{-1}\left(A^{-1}A\right)B = B^{-1}IB = B^{-1}B = I.$$

■

Ejemplo 7.16 *Si*

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$
$$A^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix},$$

entonces

$$AB = \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}$$

y

$$B^{-1}A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ \frac{5}{2} & -\frac{1}{2} \end{bmatrix}.$$

Puede verificarse fácilmente que

$$\begin{bmatrix} -4 & 1 \\ \frac{5}{2} & -\frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

de manera que $B^{-1}A^{-1}$ es un inverso izquierdo de AB y, por la proposición 7.10 $B^{-1}A^{-1} = (AB)^{-1}$.

El siguiente corolario es inmediato.

Corolario 7.17 *Si $A(1), A(2), \ldots, A(k) \in M_{n\times n}$ son invertibles, entonces el producto $A(1)A(2)\ldots A(k)$ también lo es y, de hecho,*

$$(A(1)A(2)\ldots A(k))^{-1} = A(k)^{-1}A(k-1)^{-1}\ldots A(1)^{-1}.$$

Observación 7.18 *Observemos que $I^{-1} = I$ ya que $II = I$.*

Si $A \in M_{n\times n}$, entonces la matriz transpuesta A^T también es una matriz de $n \times n$. Ahora bien, si A es invertible, ¿lo será también A^T? La siguiente proposición responde a esta pregunta.

Proposición 7.19 *Si $A \in M_{n\times n}$ es invertible, entonces A^T también lo es y $(A^T)^{-1} = (A^{-1})^T$.*

Demostración. Como $AA^{-1} = I = A^{-1}A$, tomando la transpuesta de ambos lados y utilizando el punto (4) de la Proposición 6.10 tenemos que

$$\begin{aligned}\left(AA^{-1}\right)^T &= (A^{-1})^T A^T = I^T = I,\\ \left(A^{-1}A\right)^T &= A^T (A^{-1})^T = I^T = I;\end{aligned}$$

por lo tanto, dado que la inversa es única, $(A^T)^{-1} = (A^{-1})^T$. ■

Corolario 7.20 *Si $A \in M_{n\times n}$ y B es una inversa derecha de A, entonces B^T es una inversa izquierda de A^T. Si $A \in M_{n\times n}$ y C es un inversa izquierda de A, entonces C^T es un inversa derecha de A^T.*

Demostración. En efecto, por el resultado que acabamos de demostrar

$$\begin{aligned}B^T A^T &= (AB)^T = I^T = I,\\ A^T C^T &= (CA)^T = I^T = I.\end{aligned}$$

■

Corolario 7.21 *Si $A \in M_{n\times n}$ tiene una inversa por la derecha, es decir, existe $C \in M_{n\times n}$ tal que $AC = I$, entonces los renglones de A forman un conjunto LI.*

Demostración. Dada $A \in M_{n\times n}$, sea C la inversa derecha de A. Por el Corolario 7.20, C^T es inversa izquierda de A^T. Utilizando el Lema 7.8 para A^T se tiene que sus columnas son LI. Dado que las columnas de A^T coinciden con los renglones de A, se concluye que los renglones de A son LI. ■

Ejemplo 7.22 *Sean*

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

entonces

$$A^T = \begin{bmatrix} 1 & 2\\ 0 & 1\end{bmatrix}, \quad \left(A^{-1}\right)^T = \begin{bmatrix} 1 & -2\\ 0 & 1\end{bmatrix}.$$

Puede verificarse fácilmente que

$$\left(A^{-1}\right)^T A^T = \begin{bmatrix} 1 & -2\\ 0 & 1\end{bmatrix}\begin{bmatrix} 1 & 2\\ 0 & 1\end{bmatrix} = \begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix},$$

de manera que $(A^{-1})^T = (A^T)^{-1}$.

El siguiente resultado nos dice que la existencia de una inversa derecha para la matriz $A \in M_{n\times n}$ garantiza que ésta sea invertible.

Proposición 7.23 *Si $A \in M_{n\times n}$ tiene una inversa por la derecha, entonces A tendrá una inversa por la izquierda y ambas coinciden.*

Demostración. Sea C la inversa derecha de A. Por el Corolario 7.20 C^T es inversa izquierda de A^T y aplicando la proposición 7.10 a la matriz A^T, concluimos que ésta tiene una inversa por la derecha, digamos D. Una vez más, por el Corolario 7.20 D^T es inversa izquierda de $\left(A^T\right)^T = A$. Finalmente, por la Proposición 7.7 $D^T = C = A^{-1}$. ∎

El siguiente corolario se sigue del lema 7.8, del Corolario 7.21 y de las Proposiciones 7.10 y 7.23.

Corolario 7.24 *Dada $A \in M_{n\times n}$, las siguientes son equivalentes:*

1. *A tiene una inversa por la izquierda.*
2. *Las columnas de A son LI.*
3. *A tiene una inversa por la derecha.*
4. *Los renglones de A son LI.*
5. *A es invertible.*

La Proposición 7.15 puede extenderse como sigue.

Proposición 7.25 *Sean $A, B \in M_{n\times n}$, entonces AB es invertible syss A y B son invertibles.*

Demostración. Si A y B son invertibles, por la Proposición 7.15 el producto AB también lo es. Ahora bien, si AB es invertible, se tiene que existe $(AB)^{-1}$ tal que

$$(AB)(AB)^{-1} = I.$$

Por asociatividad del producto reescribimos esta igualdad como

$$A\left[B(AB)^{-1}\right] = I.$$

Por lo tanto, A tiene una inversa derecha y por el Corolario 7.24 A es invertible y

$$A^{-1} = B(AB)^{-1}.$$

En forma análoga, como

$$(AB)^{-1}(AB) = I,$$

se tiene que

$$\left[(AB)^{-1}A\right]B = I,$$

por lo que B es invertible y

$$B^{-1} = (AB)^{-1}A.$$

■

En ocasiones se dice que una matriz es **no singular** cuando sus columnas son LI. El corolario anterior nos proporciona varias equivalencias de esta propiedad. En particular, como las matrices invertibles coinciden con las no singulares estos términos se utilizan indistintamente y, por lo tanto, si una matriz no es invertible, diremos que ésta es singular. En lo que sigue encontraremos la inversa de una matriz (invertible) utilizando el método de Gauss-Jordan estudiado en el Capítulo 2. Por lo visto anteriormente, será suficiente encontrar la inversa derecha o izquierda.

Observación 7.26 *Dada una matriz invertible A, las columnas de su matriz inversa pueden encontrarse utilizando la demostración de la Proposición 7.10. Concretamente, si $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ son las columnas de la matriz inversa, éstas satisfacen los siguientes sistemas de ecuaciones:*

$$A\vec{x}_1 = \vec{e}_1,\ A\vec{x}_2 = \vec{e}_2, \ldots, A\vec{x}_n = \vec{e}_n.$$

*Es decir, se tienen n sistemas de ecuaciones (uno por cada columna j). Afortunadamente, como comparten la misma matriz de coeficientes, todos estos **sistemas pueden condensarse en uno sólo** como veremos a continuación.*

Ejemplo 7.27 *Sea*

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

y supongamos que queremos encontrar $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = A^{-1}$. Para este efecto resolvemos los siguientes sistemas de ecuaciones:

$$\begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} b_{12} \\ b_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Como los sistemas poseen la misma matriz de coeficientes, podemos considerarlos como un sólo sistema con dos columnas de términos independientes obteniendo la FER como sigue:

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 0 & 4 & 0 & 1 \end{array}\right] \overset{\frac{1}{2}R_1}{\sim} \left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 4 & 0 & 1 \end{array}\right] \overset{\frac{1}{4}R_2}{\sim}$$
$$\left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & \frac{1}{4} \end{array}\right] \overset{-\frac{1}{2}R_2+R_1}{\sim} \left[\begin{array}{cc|cc} 1 & 0 & \frac{1}{2} & -\frac{1}{8} \\ 0 & 1 & 0 & \frac{1}{4} \end{array}\right].$$

Por lo tanto, la matriz

$$B = \begin{bmatrix} \frac{1}{2} & -\frac{1}{8} \\ 0 & \frac{1}{4} \end{bmatrix}$$

es una inversa por la derecha para A ($AB = I$). Por el Corolario 7.24, la matriz A será invertible, de manera que $B = A^{-1}$.

Ejemplo 7.28 *Sea*

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

y supongamos que deseamos encontrar A^{-1}. El lector paciente puede comprobar que

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \\ 0 & 0 & 1 & -\frac{1}{2} & 0 & \frac{1}{2} \end{array}\right],$$

por lo que

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

Ejemplo 7.29 *Si tomamos a la matriz*

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

del ejemplo 7.9, el procedimiento anterior nos proporciona la FER como

$$\left[\begin{array}{cc|cc} 1 & -3 & 1 & 0 \\ -1 & 3 & 0 & 1 \end{array}\right] \overset{R_1+R_2}{\sim} \left[\begin{array}{cc|cc} 1 & -3 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

Claramente el sistema (o más bien los dos sistemas compactados en uno) es inconsistente y no existe la matriz inversa de A.

Observación 7.30 *El procedimiento anterior queda resumido como sigue: si A es invertible y queremos encontrar su inversa utilizando el método de Gauss Jordan, simplemente se comienza con la matriz aumentada*

$$[A \mid I]$$

y se aplican las operaciones usuales para los renglones para llegar a la FER dada por

$$\left[I \mid A^{-1}\right].$$

Si interpretamos a las operaciones entre renglones en términos de multiplicación por matrices elementales, la observación 7.30 junto con el corolario 6.42 nos llevan al siguiente resultado.

Proposición 7.31 *$A \in M_{n\times n}$ es invertible syss existen matrices elementales $E(1), E(2), \ldots, E(k)$ tales que*

$$E(k)E(k-1)\ldots E(1)A = I.$$

En particular dado que, por definición, las matrices elementales se obtienen de la identidad mediante una operación elemental entre renglones, cualquier matriz elemental es invertible. Más aún, el inverso de una matriz elemental es otra matriz elemental ya que el inverso de cualquier operación elemental entre renglones es otra operación elemental del mismo tipo.

Ejemplo 7.32 *La matriz elemental*

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

que se obtiene a partir de la identidad intercambiando los renglones es su propia inversa (simplemente se vuelven a intercambiar los renglones). En efecto,

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

Ejemplo 7.33 *La matriz*

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

que se obtiene de la identidad multiplicando el segundo renglón por π tiene como inversa la matriz que se obtiene de la identidad multiplicando el segundo renglón por $\frac{1}{\pi}$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\pi} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ejemplo 7.34 *La matriz*

$$E = \begin{bmatrix} 1 & 0 & 666 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

que se obtiene de la identidad sumando al primer renglón 666 *veces el tercero tiene como inversa la matriz que se obtiene de la identidad sumando al primer renglón* -666 *veces el tercero. En efecto,*

$$\begin{bmatrix} 1 & 0 & -666 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 666 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Ejemplo 7.35 *En el ejemplo 7.27 podemos representarla sucesión de operaciones elementales sobre los renglones como productos por matrices elementales como sigue:*

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

de manera que

$$\left(\begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{8} \\ 0 & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}^{-1}.$$

Las consideraciones anteriores nos conducen al siguiente corolario.

Corolario 7.36 $A \in M_{n\times n}$ *es invertible syss existen matrices elementales* $\mathcal{E}(1), \mathcal{E}(2), \ldots, \mathcal{E}(k)$ *tales que*

$$A = \mathcal{E}(1)\mathcal{E}(2)\ldots\mathcal{E}(k).$$

Demostración. Utilizando la proposición 7.31, si A es invertible, se tiene que $E(k)E(k-1)\ldots E(1)A = I$. Entonces definiendo

$$\begin{aligned} \mathcal{E}(k) &= (E(k))^{-1}, \\ \mathcal{E}(k-1) &= (E(k-1))^{-1}, \\ &\vdots \\ \mathcal{E}(1) &= (E(1))^{-1}, \end{aligned}$$

se tiene que

$$\begin{aligned} A &= (\mathcal{E}(1)\mathcal{E}(2)\ldots\mathcal{E}(k))\,(E(k)E(k-1)\ldots E(1))\,A \\ &= \mathcal{E}(1)\mathcal{E}(2)\ldots\mathcal{E}(k)I = \mathcal{E}(1)\mathcal{E}(2)\ldots\mathcal{E}(k). \end{aligned}$$

Asimismo, como las matrices elementales son invertibles, si A es un producto de matrices elementales, entonces, por el corolario 7.17, A es invertible. ■

Ejemplo 7.37 *Si*

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix},$$

como se vio en el Ejemplo 7.35, las matrices elementales utilizadas para llegar a la FER de A son,

$$\begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}.$$

Sus inversas son las matrices elementales dadas como,

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

$$\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix},$$

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

y por el Corolario 7.36:

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

Transformaciones inversas

Cuando se tiene cualquier función entre dos conjuntos decimos que ésta es **biyectiva** cuando es inyectiva y suprayectiva. Es natural utilizar la misma definición para las funciones que son transformaciones lineales entre dos espacios vectoriales, de manera que así lo haremos. La siguiente proposición nos dice que el dominio y el contradominio de una transformación lineal biyectiva de $\mathbb{R}^n$ en $\mathbb{R}^n$ deben coincidir.

Proposición 7.38 *Si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal biyectiva, entonces $n = m$.*

Demostración. Basta recordar, por las Proposiciones 5.27 y 5.39, que si $n < m$, entonces $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ no puede ser suprayectiva y que si $n > m$, entonces $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ no puede ser inyectiva. ■

Observación 7.39 *Supongamos que se tiene una transformación lineal inyectiva*

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n.$$

Por el Teorema 5.40, sabemos que T también será suprayectiva y, por lo tanto biyectiva. Equivalentemente, si asumimos inicialmente que T es suprayectiva, por el mismo teorema concluimos que debe ser inyectiva y, en consecuencia biyectiva. En resumen, para transformaciónes lineales de $\mathbb{R}^n$ en $\mathbb{R}^n$, la inyectividad o suprayectividad equivalen a la biyectividad.

Dados $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una transformación lineal biyectiva y cualquier vector $\vec{y} \in \mathbb{R}^n$, por suprayectividad, tenemos que existe $\vec{x} \in \mathbb{R}^n$ tal que $T(\vec{x}) = \vec{y}$. Asimismo, como T es inyectiva, $\vec{x}$ es el único vector con esta propiedad. Podemos definir así la función (aún no sabemos si resultará ser una transformación lineal)

$$T^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

como $T^{-1}(\vec{y}) = \vec{x}$, en donde $\vec{x}$ es el único elemento que cumple $T(\vec{x}) = \vec{y}$. Si I es la transformación identidad, es claro que se cumplen

$$\begin{aligned}(T^{-1} \circ T)(\vec{x}) &= T^{-1}(T(\vec{x})) = T^{-1}(\vec{y}) = \vec{x} = I(\vec{x}),\\ (T \circ T^{-1})(\vec{y}) &= T(T^{-1}(\vec{y})) = T(\vec{x}) = \vec{y} = I(\vec{y}).\end{aligned}$$

Lo anterior nos dice que $T^{-1} \circ T = I = T \circ T^{-1}$.

Utilizando, una vez más el Teorema 5.40, si $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es biyectiva y $A \in M_{n\times n}$ es su matriz estándar asociada, las columnas de A son linealmente independientes y, por el Corolario 7.24, A será invertible. Sea A^{-1} el inverso de A y sea

$$S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

la transformación lineal asociada, es decir, $S(\vec{y}) = A^{-1}\vec{y}$ para $\vec{y} \in \mathbb{R}^n$. Si I_n representa la matriz identidad de $n \times n$, se tiene que para todo $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$\begin{aligned}(S \circ T)(\vec{x}) &= S(T(\vec{x})) = A^{-1}A\vec{x} = I_n\vec{x} = I(\vec{x}) = \vec{x},\\ (T \circ S)(\vec{y}) &= T(S(\vec{y})) = AA^{-1}\vec{y} = I_n\vec{y} = I(\vec{y}) = \vec{y}.\end{aligned}$$

Notemos que dado cualquier vector $\vec{y} \in \mathbb{R}^n$, la transformación S queda definida como $S(\vec{y}) = \vec{x}$, en donde $\vec{x} \in \mathbb{R}^n$ es el único vector tal que $T(\vec{x}) = \vec{y}$. Por lo tanto, S coincide con la definición de T^{-1} (como función entre conjuntos) y $S = T^{-1}$. De esta forma, T^{-1} es una transformación lineal biyectiva (no sólo una función) cuya matriz estándar asociada es A^{-1}.

Definición 7.40 *Al igual que en el caso de las matrices, decimos que una transformación lineal $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ es invertible si existe una transformación lineal $T^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, tal que*

$$T^{-1} \circ T = I = T \circ T^{-1}.$$

Los resultados correspondientes a las Proposiciones 7.14, 7.15, 7.25 y el Corolario 7.17 en el contexto de transformaciones se enuncian a continuación. Las demostraciones son inmediatas y se dejan como ejercicio al lector.

Proposición 7.41 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *es una transformación lineal invertible se cumple* $(T^{-1})^{-1} = T$.

Proposición 7.42 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *y* $S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *son transformaciones lineales invertibles, se cumple* $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$.

Proposición 7.43 *Si* $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *y* $S : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *son transformaciones lineales, entonces* S *y* T *son invertibles syss* $S \circ T$ *es invertible.*

Corolario 7.44 *Si* $T_i : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $i = 1, 2, \ldots, k$ *son transformaciones lineales invertibles, se cumple*

$$(T_1 \circ T_2 \circ \cdots \circ T_k)^{-1} = T_k^{-1} \circ T_{k-1}^{-1} \circ \cdots \circ T_1^{-1}.$$

Ejemplo 7.45 *Sea* $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ *dada por*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x \end{bmatrix}$$

y supongamos que deseamos calcular la matriz estándar que representa a T^{-1} *y* $T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$. *Para ello, notamos, primero, que si* A *es la matriz estándar que representa a* T, *entonces*

$$A = [T(\vec{e}_1)\ T(\vec{e}_2)] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Gracias a la discusión anterior, sabemos que A^{-1} *es la matriz que representa a* T^{-1}. *Para calcular esta matriz inversa notamos que*

$$\left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] \overset{R_1 \leftrightarrow R_2}{\sim} \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right],$$

por lo que

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

Finalmente,

$$T^{-1}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = A^{-1}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}.$$

Esto es, $T^{-1} = T$ *y* $A^{-1} = A$, *lo cual tiene mucho sentido pues es obvio que la composición* $T \circ T$ *es igual a la transformación identidad y que* $AA = I_2$.

Capítulo 8

Determinantes

Introducción

En este capítulo introduciremos un criterio numérico para decidir si una matriz cuadrada es o no invertible. Para ello comenzamos notando que dada la matriz de 2×2

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

si $ad - bc \neq 0$, necesariamente $ad \neq 0$ o bien $bc \neq 0$. Sin pérdida de generalidad, supongamos que $ad \neq 0$ (el caso $bc \neq 0$ es análogo) de manera que $a \neq 0 \neq d$. La FER de la matriz aumentada

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right]$$

se encuentra como sigue

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \overset{-\frac{c}{a}R_1+R_2}{\sim} \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d-\frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right] \overset{\frac{1}{a}R_1,\ \frac{a}{ad-bc}R_2}{\sim}$$
$$\left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right] \overset{-\frac{b}{a}R_2+R_1}{\sim} \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right].$$

Por lo tanto

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

lo cual se verifica fácilmente ya que,

$$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Lo anterior implica que si $ad-bc \neq 0$, entonces A es invertible.

Ahora veremos que si

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

es invertible entonces $ad-bc \neq 0$. Si A es invertible, entonces sus columnas son LI y por lo tanto, $a \neq 0$ o $b \neq 0$ (Si $a = b = 0$, las columnas son LD). Sin pérdida de generalidad supondremos que $a \neq 0$ (el lector puede verificar que el caso cuando $b \neq 0$ es completamente análogo). Al multiplicar el primer renglón de A por $-\frac{c}{a}$ y sumárselo al segundo obtenemos la matriz

$$\begin{bmatrix} a & b \\ 0 & d-\frac{bc}{a} \end{bmatrix}.$$

Notamos ahora que $d - \frac{bc}{a} \neq 0$, porque en caso contrario la matriz no sería invertible. Por lo tanto,

$$ad - bc \neq 0,$$

tal y como queríamos demostrar.

La función determinante y sus propiedades

El razonamiento de la sección anterior nos conduce directamente a la siguiente definición.

Definición 8.1 *Si A es una matriz de 2×2 de la forma*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

entonces se define el ***determinante*** *de A, denotado por $\det(A)$, como*

$$\det(A) = ad - bc.$$

Utilizando esta terminología tenemos el siguiente resultado.

Proposición 8.2 *Si A es una matriz de 2×2 de la forma*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

entonces A es invertible syss $\det(A) \neq 0$. *En cuyo caso,*

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Ejemplo 8.3 *Si*

$$A = \begin{bmatrix} 3 & -2 \\ 6 & -4 \end{bmatrix} \text{ y } B = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix},$$

entonces A no es invertible (pues $\det(A) = 3(-4) - (-2)6 = 0$*) y B sí es invertible (pues $det(B) = 1(4) - (2)(-3) = 10$). Más aún,*

$$B^{-1} = \frac{1}{10} \begin{bmatrix} 4 & -2 \\ 3 & 1 \end{bmatrix}.$$

Observación 8.4 *Dada*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

la matriz

$$\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

*se denomina la **matriz adjunta** de A y se denota por $adj(A)$. Esta matriz tiene la propiedad de que si A es invertible, entonces*

$$A^{-1} = \frac{1}{\det(A)} adj(A).$$

Como puede observarse en el caso en el cual $A \in M_{2\times 2}$, la matriz $adj(A)$ es sumamente simple de obtener. El caso general es más elaborado y se abordará mas adelante en este capítulo.

A continuación proporcionamos una definición del determinante basada en la dada por E. Artin[1]. Más adelante veremos que ésta concide con la que hemos proporcionado anteriormente para matrices de 2×2.

Definición 8.5 *Decimos que una función*

$$F : M_{n\times n} \longrightarrow \mathbb{R}$$

*es **multilineal en los renglones** si se cumplen las siguientes dos propiedades:*

[1]Galois Theory, Dover Publications, Inc, NY, 1998.

1. *Dada $A \in M_{n\times n}$, si $A(1)$ se obtiene de A multiplicando el $i-$ ésimo renglón por $c \in \mathbb{R}$, entonces*

$$F(A(1)) = cF(A).$$

2. *Dadas las matrices de $n \times n$*

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}, \ A' \ y \ B,$$

en donde A' se obtiene de A sustituyendo el renglón A_i por

$$[a'_{i1} \quad a'_{i2} \cdots \ a'_{in}]$$

y B se obtiene de A sustituyendo el renglón A_i por

$$[(a_{i1} + a'_{i1}) \quad (a_{i2} + a'_{i2}) \cdots (a_{in} + a'_{in})],$$

entonces

$$F(B) = F(A) + F(A').$$

Observación 8.6 *Lo anterior no implica que F sea una función lineal puesto que, dados $A, B \in M_{n\times n}$ y $c \in \mathbb{R}$, no se cumplen*

$$F(cA) = cF(A)$$

ni

$$F(A + B) = F(A) + F(B).$$

La multilinealidad implica que F es lineal en cada renglón. Alternativamente, la multilinealidad puede definirse para las columnas de A de la misma forma que se hizo para los renglones. En ese caso se diría que F es ***multilineal en las columnas.***

Ejemplo 8.7 *Sea* $\det : M_{2\times 2} \longrightarrow \mathbb{R}$ *la función determinante de la Definición 8.1. Observamos que*

$$\begin{aligned} \det\left(\begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{bmatrix}\right) &= (a_1 + a_2)\, d - (b_1 + b_2)\, c \\ &= (a_1 d - b_1 c) + (a_2 d - b_2 c) \\ &= \det\left(\begin{bmatrix} a_1 & b_1 \\ c & d \end{bmatrix}\right) + \left(\begin{bmatrix} a_2 & b_2 \\ c & d \end{bmatrix}\right) \end{aligned}$$

y dada $r \in \mathbb{R}$,

$$\det\left(\begin{bmatrix} ra & rb \\ c & d \end{bmatrix}\right) = rad - rbc$$
$$= r(ad - bc) = r\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).$$

Evidentemente las mismas propiedades se cumplen para el segundo renglón y concluimos que la función det *es multilineal.*

Ejemplo 8.8 *Sean*

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \in M_{2\times 2}$$

y $\det : M_{2\times 2} \longrightarrow \mathbb{R}$ *la función determinante de la Definición 8.1. Entonces,*

$$\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = -2 = -1 + (-1)$$
$$= \det\left(\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\right) + \det\left(\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}\right)$$

y

$$\det\left(\begin{bmatrix} 1 & 2 \\ -6 & -8 \end{bmatrix}\right) = -2\det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\right) = 4.$$

Definición 8.9 *Sea*

$$\det : M_{n\times n} \longrightarrow \mathbb{R}$$

una función que satisface las siguientes propiedades:

1. det *es multilineal en los renglones.*
2. $\det(A) = 0$ *cuando A tiene dos renglones consecutivos iguales.*
3. $\det(I) = 1$.

Dada $A \in M_{n\times n}$, nos referiremos a $\det(A)$ como el **determinante** de A. En algunos textos se utiliza la notación $|A|$ para denotar al determinante de A. Una pregunta natural es si existe una función con las propiedades de $\det(A)$ y en dado caso, ¿es ésta única?

Veamos primero algunas propiedades que se siguen de la definición.

Proposición 8.10 *Si $A \in M_{n\times n}$ tiene un renglón de ceros,*

$$\det(A) = 0.$$

Demostración. Si A' se obtiene de A multiplicando algún renglón por $c = 0$, la multilinealidad nos dice que

$$\det(A') = 0 \times \det(A) = 0.$$

■

Proposición 8.11 *Supongamos que A' se obtiene a partir de A intercambiando dos renglones consecutivos, digamos el i y el $i+1$, es decir,*

$$A \overset{R_i \leftrightarrow R_{i+1}}{\sim} A',$$

entonces

$$\det(A') = -\det(A).$$

Demostración. Utilizando las propiedades (1) y (2) del determinante se tiene que,

$$\begin{aligned}
0 &= \det \begin{bmatrix} \vdots \\ A_i + A_{i+1} \\ A_i + A_{i+1} \\ \vdots \end{bmatrix} \begin{array}{l} \\ \longleftarrow i \\ \longleftarrow i+1 \\ \\ \end{array} \\
&= \det \begin{bmatrix} \vdots \\ A_i \\ A_i + A_{i+1} \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vdots \\ A_{i+1} \\ A_i + A_{i+1} \\ \vdots \end{bmatrix} \\
&= \det \begin{bmatrix} \vdots \\ A_i \\ A_i \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vdots \\ A_i \\ A_{i+1} \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vdots \\ A_{i+1} \\ A_i \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vdots \\ A_{i+1} \\ A_{i+1} \\ \vdots \end{bmatrix} \\
&= 0 + \det(A) + \det(A') + 0,
\end{aligned}$$

por lo tanto

$$\det(A') = -\det(A).$$

■

Corolario 8.12 *La propiedad (2) de la Definición 8.9 puede generalizarse como sigue.*

(2)´ $\det(A) = 0$ *cuando A tiene dos renglones iguales (no necesariamente consecutivos).*

Demostración. Supongamos que en la matriz A los renglones i y k son iguales con $i < k$, es decir $A_i = A_k$ y A_k está por debajo de A_i. Construimos la matriz A' intercambiando A_k en forma sucesiva con los renglones que lo preceden hasta llegar a la posición $i+1$. De acuerdo a la Proposición 8.11 se tendrá que

$$\det(A') = \pm \det(A),$$

en donde el signo depende del número de veces que intercambiamos renglones. Dado que A' tiene dos renglones consecutivos iguales (el i y el $i+1$), se tendrá que por la propiedad (2),

$$\det(A') = 0 = \det(A).$$

■

Utilizando la propiedad $(2)'$, el siguiente corolario es inmediato sustituyendo $i+1$ por k en la demostración de la Proposición 8.11.

Corolario 8.13 *Si A' se obtiene a partir de A intercambiando dos renglones, digamos el $i-$ésimo y el $k-$ésimo; es decir,*

$$A \overset{R_i \leftrightarrow R_k}{\sim} A',$$

entonces

$$\det(A') = -\det(A).$$

En el caso de $n = 2$, podemos encontrar una función determinante a partir de sus propiedades como sigue. Sea

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Calculamos el determinante de la matriz A de la siguiente forma:

$$\begin{aligned}
\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \det \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} + \det \begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \\
&= a \det \begin{bmatrix} 1 & 0 \\ c & d \end{bmatrix} + b \det \begin{bmatrix} 0 & 1 \\ c & d \end{bmatrix} \\
&= ac \det \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + ad \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&+ bc \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + bd \det \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \\
&= 0 + ad(1) + bc(-1) + 0 = ad - bc.
\end{aligned}$$

Es decir, ¡$\det(A)$ es el mismo que el proporcionado en la Definición 8.1! Así, para matrices de 2×2 la función determinante no solo existe sino que es única.

El Corolario 8.13 nos dice qué la función determinante cambia de signo al intercambiar dos renglones en una matriz. A continuación vemos lo que sucede con la función det, en caso de existir, al aplicar las otras dos operaciones elementales entre los renglones de una matriz.

Proposición 8.14 *Si A' se obtiene a partir de A multiplicando el renglón i por un real $c \neq 0$, es decir*

$$A \overset{cR_i}{\sim} A',$$

entonces

$$\det(A') = c\det(A).$$

Demostración. *La demostración es inmediata por la multilinealidad (para renglones) de la función determinante.* ■

Proposición 8.15 *Si $c \in \mathbb{R}$ y A' se obtiene a partir de A sumando c veces el renglón i al renglón j, es decir*

$$A \overset{cR_i+R_j}{\sim} A',$$

entonces

$$\det(A') = \det(A).$$

Demostración. Por la multilinealidad (para renglones) del determinante se tiene que

$$\det(A') = \det \begin{array}{cc} \begin{bmatrix} \vdots \\ A_i \\ \vdots \\ cA_i + A_j \\ \vdots \end{bmatrix} & \begin{array}{c} \\ \longleftarrow i \\ \\ \longleftarrow j \\ \\ \end{array} \end{array} = \det \begin{bmatrix} \vdots \\ A_i \\ \vdots \\ cA_i \\ \vdots \end{bmatrix} + \det \begin{bmatrix} \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \end{bmatrix}$$

$$= c\det \begin{bmatrix} \vdots \\ A_i \\ \vdots \\ A_i \\ \vdots \end{bmatrix} + \det(A) = 0 + \det(A) = \det(A).$$

En el último paso se utilizó que el determinante de una matriz con dos renglones iguales es nulo. Concluimos que

$$\det(A) = \det(A').$$

■

El siguiente corolario enuncia estos resultados en términos de matrices elementales.

Corolario 8.16 *La función*

$$\det : M_{n\times n} \longrightarrow \mathbb{R}$$

cumple las siguientes propiedades para las matrices elementales dadas en la proposición anterior:

Proposición 8.17 1. *Si $E(1) \in M_{n\times n}$ se obtiene a partir de I intercambiando dos renglones, entonces*

$$\det(E(1)) = -1.$$

2. *Si $E(2) \in M_{n\times n}$ se obtiene a partir de I multiplicando un renglón por un número real $c \neq 0$, entonces*

$$\det(E(2)) = c.$$

3. *Sea $c \in \mathbb{R}$, si $E(3) \in M_{n\times n}$ se obtiene a partir de I sumando c veces un renglón a otro renglón, entonces*

$$\det(E(3)) = 1.$$

Demostración. La demostración de los tres incisos es inmediata tomando $A = I$ en las Proposiciones 8.14, 8.11, 8.15. ■

Las consideraciones anteriores nos llevan directamente a los siguientes resultados.

Proposición 8.18 *Sea $A \in M_{n\times n}$ cualquier matriz y $E(1), E(2), E(3) \in M_{n\times n}$ las matrices elementales del Corolario 8.16. Entonces*

$$\begin{aligned}
\det\left(E(1)A\right) &= \det(E(1))\det(A) = -\det(A),\\
\det\left(E(2)A\right) &= \det(E(2))\det(A) = c\det(A),\\
\det\left(E(3)A\right) &= \det(E(3))\det(A) = \det(A),
\end{aligned}$$

es decir, la función determinante preserva productos por matrices elementales.

Demostración. La demostración es inmediata utilizando las Proposiciones 8.14, 8.11, 8.15 y el Corolario 8.16, puesto que la multiplicación por matrices elementales equivale a las operaciones elementales sobre los renglones de A. ■

Corolario 8.19 *Sean $\mathcal{E}(1)$, $\mathcal{E}(2), \ldots,$ $\mathcal{E}(k) \in M_{n\times n}$ cualesquiera matrices elementales y $A \in M_{n\times n}$. Entonces*

$$\det\left(\mathcal{E}(k)\mathcal{E}(k-1)\cdots\mathcal{E}(1)A\right) = \det\left(\mathcal{E}(k)\right)\det\left(\mathcal{E}(k-1)\right)\cdots\det\left(\mathcal{E}(1)\right)\det(A).$$

Demostración. La prueba es inmediata aplicando la Proposición 8.18 en forma recursiva. ∎

Observación 8.20 *Si en el Corolario 8.19 tomamos $A = I$, entonces*

$$\det(\mathcal{E}(k)\mathcal{E}(k-1)\cdots\mathcal{E}(1)) = \det(\mathcal{E}(k))\det(\mathcal{E}(k-1))\cdots\det(\mathcal{E}(1)).$$

Ahora estamos listos para extender el resultado que se tenía para matrices de 2×2, que proporciona la relación entre el determinante y la invertibilidad de una matriz.

Proposición 8.21 *Si $A \in M_{n\times n}$, entonces A es invertible syss* $\det(A) \neq 0$.

Demostración. Sean $\tilde{A}$ la FER de A y $\mathcal{E}(1), \mathcal{E}(2), \ldots \mathcal{E}(k)$ las matrices elementales tales que

$$A = \mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k)\tilde{A}.$$

Ahora bien, si A es invertible $\tilde{A} = I$; por los Corolarios 8.19, 8.16 y la Proposición 8.10, se tiene que

$$\begin{aligned}\det(A) &= \det(\mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k)) \\ &= \det(\mathcal{E}(1))\det(\mathcal{E}(2))\cdots\det(\mathcal{E}(k)) \neq 0.\end{aligned}$$

Si A no es invertible, entonces $\tilde{A}$ tiene un renglón de ceros y la Proposición 8.10 nos dice que $\det(\tilde{A}) = 0$. Por lo tanto,

$$\begin{aligned}\det(A) &= \det\left(\mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k)\tilde{A}\right) \\ &= \det(\mathcal{E}(1))\det(\mathcal{E}(2))\cdots\det(\mathcal{E}(k))\det(\tilde{A}) = 0,\end{aligned}$$

con lo cual se concluye la demostración. ∎

El siguiente resultado es uno de los más importantes en relación a la función determinante.

Proposición 8.22 *Si A y B son matrices de $n \times n$, entonces*

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

Demostración. Si A o B no es invertible, entonces AB tampoco lo es y por la Proposición 8.21 se cumple

$$0 = \det(AB) = \det(A)\det(B) = 0.$$

Si A y B son invertibles, AB también lo es y la prueba es inmediata de la Observación 8.20, recordando que las matrices invertibles son productos de matrices elementales. ∎

Observación 8.23 *Si A es invertible,*

$$\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1,$$

por lo que

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Ejemplo 8.24 *Sea*

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

entonces $\det(A) = -5$. *La matriz inversa está dada por*

$$A^{-1} \begin{bmatrix} -\frac{2}{5} & \frac{1}{5} \\ \frac{3}{5} & \frac{1}{5} \end{bmatrix}$$

y

$$\det(A^{-1}) = -\frac{1}{5} = \frac{1}{\det(A)}.$$

Ejemplo 8.25 *La matriz invertible*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -2 \\ 5 & 1 & 0 \end{bmatrix},$$

puede expresarse como producto de matrices elementales como

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

Entonces, como

$$\det\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\right) = -1, \quad \det\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}\right) = -2,$$
$$\det\left(\begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = 1,$$

se tiene que

$$\det(A) = (-1)(-2)(1) = 2.$$

Si nos remitimos a la discusión final del Capítulo 6 en relación a las transpuestas de matrices elementales, tenemos que si $E(1), E(2), E(3) \in M_{n\times n}$ son las matrices elementales del Corolario 8.16, entonces

$$E(1)^T = E(1),$$
$$E(2)^T = E(2)$$

y $E(3)^T$ es una matriz elemental del mismo tipo que $E(3)$. De esta forma,

$$\det(E(i)^T) = \det(E(i)),\ i = 1, 2, 3.$$

Estas consideraciones nos llevan al siguiente resultado.

Proposición 8.26 *Si $A \in M_{n\times n}$, entonces*

$$\det(A) = \det(A^T).$$

Demostración. *Si A no es invertible A^T tampoco lo es y se tiene*

$$0 = \det(A) = \det(A^T) = 0.$$

Si A es invertible, entonces A^T también lo es. Adicionalmente, A puede expresarse como producto de matrices elementales, digamos

$$A = \mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k).$$

Por lo tanto,

$$A^T = (\mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k))^T = \mathcal{E}(k)^T\mathcal{E}(k-1)^T\cdots\mathcal{E}(1)^T$$

y

$$\begin{aligned}\det(A^T) &= \det\left(\mathcal{E}(k)^T\mathcal{E}(k-1)^T\cdots\mathcal{E}(1)^T\right)\\ &= \det\left(\mathcal{E}(k)^T\right)\det\left(\mathcal{E}(k-1)^T\right)\cdots\det\left(\mathcal{E}(1)^T\right)\\ &= \det\left(\mathcal{E}(k)\right)\det\left(\mathcal{E}(k-1)\right)\cdots\det\left(\mathcal{E}(1)\right)\\ &= \det(A).\end{aligned}$$

■

Observación 8.27 *Como consecuencia inmediata de este resultado se tiene que las condiciones (1) y (2) de la Definición 8.9 pueden expresarse, alternativamente, en términos de las columnas de la matriz como sigue.*

1c. La función det *es multilineal en las columnas.*

2c. $\det(A) = 0$ *cuando A tiene dos columnas consecutivas iguales.*

Antes de mostrar una función que cumpla todas las propiedades de la función determinante para $n > 2$, demostraremos que dicha función es única. En realidad esto es relativamente simple dadas todas las propiedades que conocemos de la función determinante como se ve a continuación.

Teorema 8.28 (Unicidad de la función determinante) *Si existe la función* det *de la Definición 8.9, ésta es única.*

Demostración. Supongamos que existen dos funciones, det y det́, que cumplen con las tres condiciones de la Definición 8.9. Dada $A \in M_{n\times n}$, si A no es invertible, por la Proposición 8.10 se tiene que

$$\det(A) = 0 = \acute{\det}(A).$$

Si A es invertible, existen $\mathcal{E}(1), \mathcal{E}(2), \ldots, \mathcal{E}(k)$ matrices elementales de $n \times n$ tales que

$$A = \mathcal{E}(1)\mathcal{E}(2)\cdots\mathcal{E}(k).$$

Por el Corolario 8.16 y la Observación 8.20 se tiene que las funciones det y det́ coinciden en las matrices elementales y preservan productos, por lo tanto

$$\begin{aligned}
\det(A) &= \det(\mathcal{E}(1))\det(\mathcal{E}(2))\cdots\det(\mathcal{E}(k)) \\
&= \acute{\det}(\mathcal{E}(1))\acute{\det}(\mathcal{E}(2))\cdots\acute{\det}(\mathcal{E}(k)) \\
&= \acute{\det}(A)
\end{aligned}$$

y ambas funciones coinciden. ■

Observación 8.29 *La descomposición de A como producto de matrices elementales no es única ya que es posible llegar a la FER de A de diversas maneras. Esto podría ocasionar problemas para definir la función determinante; sin embargo, el teorema de unicidad que acabamos de demostrar no tiene que ver con la existencia y definición del determinante. Simplemente nos dice que si existen las funciones* det *y* det́*, entonces tienen que coincidir.*

Hasta el momento, todo el desarrollo acerca de las propiedades de la función determinante ha sido un simple ejercicio teórico. Despues de todo, ¡es posible que esta maravillosa función ni siquiera exista para $n > 2$! Afortunadamente, este no es el caso y en la siguiente sección construiremos una función explícita para calcular el deteminante de una matriz de $n \times n$.

Existencia del determinante.

Antes de definir el determinante en el caso general se requiere de la siguiente definición.

Definición 8.30 *Dados* $A \in M_{n\times n}$ *y dos números naturales* i *y* j *en* $\{1, 2, \ldots, n\}$, *definimos* $\widetilde{A}_{ij} \in M_{(n-1)\times(n-1)}$ *como la matriz que se obtiene a partir de* A *eliminando el renglón* i *y la columna* j.

Ejemplo 8.31 *Consideremos a las matrices*

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

Entonces

$$\widetilde{A}_{13} = \begin{bmatrix} -2 & 6 \\ 8 & -1 \end{bmatrix}, \ \tilde{A}_{31} = \begin{bmatrix} -7 & 0 \\ 6 & 9 \end{bmatrix}, \ y\ \widetilde{B}_{43} = \begin{bmatrix} 2 & 5 & 9 \\ 3 & 2 & \pi \\ 1 & 8 & 4 \end{bmatrix}.$$

Recordemos que en el caso de matrices

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2}.$$

la función determinante está dada por

$$\det(A) = ad - bc.$$

Definición 8.32 *Sea* A *una matriz de* $n \times n$ *y sea* j *cualquier columna de* A. *La función*

$$D : M_{n\times n} \longrightarrow \mathbb{R}$$

se define como

$$D(A) = \det(A)$$

para el caso $n = 2$ *y recursivamente para* $n > 2$ *como*

$$D(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} D(\widetilde{A}_{ij}). \tag{8.1}$$

Observemos que la función definida por (8.1) parecería depender de la columna elegida. Lo sorprendente (y el primero en notarlo fue P.S. Laplace) es que esta elección será irrelevante para el desarrollo de $D(A)$. La razón es muy simple: independientemente de la columna j que se elija para calcular $D(A)$, la expresión (8.1) satisface las propiedades de la Definición 8.9 y por lo tanto, se tendrá que por la unicidad del determinante, se cumple

$$D(A) = \det(A).$$

Procedamos a verificar que, en efecto, esto sucede.

Proposición 8.33 *Sean A una matriz de $n \times n$ y A^j cualquier columna de A. Entonces, $D(A)$ definido como en (8.1) satisface las propiedades de la Definición 8.9.*

Demostración. Las tres propiedades de 8.9 son inmediatas para $n = 2$, utilizando la Definición 8.1. Sea $n > 2$, se utiliza inducción sobre n en todos los casos. Vamos a demostrar la primera parte de la multilinealidad y la propiedad (2) de 8.9. Las demás propiedades se demuestran en forma semejante y se dejan como ejercicio al lector. Consideremos primero a la matriz A' que se obtiene multiplicando el $k-\acute{e}simo$ renglón de A por una constante c. Observemos que si $i \neq k$ las matrices $\widetilde{A'}_{ij}$ son de tamaño $(n-1) \times (n-1)$ y se obtienen de $\widetilde{A}_{ij}$ multiplicando el renglón k por la constante c; entonces, por hipótesis de inducción,

$$D(\widetilde{A'}_{ij}) = cD(\widetilde{A}_{ij})$$

para toda $i \neq k$. En el caso para el cual $i = k$, se tiene que $\widetilde{A'}_{ij} = \widetilde{A}_{ij}$ pero $a'_{kj} = ca_{kj}$. De esta forma se cumple

$$\begin{aligned} D(A') &= \sum_{i=1}^{n} (-1)^{i+j} a'_{ij} D(\widetilde{A'}_{ij}) \\ &= (-1)^{k+j} ca_{kj} D(\widetilde{A}_{kj}) + \sum_{\substack{i=1 \\ i \neq k}}^{n} (-1)^{i+j} a_{ij} cD(\widetilde{A}_{ij}) \\ &= cD(A). \end{aligned}$$

Ahora supongamos que la matriz A tiene dos renglones consecutivos iguales, digamos $A_k = A_{k+1}$. Si $i \neq k$, $i \neq k+1$, las matrices $\widetilde{A}_{ij}$ son de tamaño $(n-1) \times (n-1)$ y sus renglones k y $k+1$ son iguales. Por hipótesis de inducción,

$$D(\widetilde{A}_{ij}) = 0$$

para toda $i \neq k$, $i \neq k+1$. Por lo tanto, en el desarrollo de $D(A)$ sólo aparecen los términos correspondientes a $i = k$ e $i = k+1$. Adicionalmente, $a_{kj} = a_{(k+1)j}$ y $\widetilde{A}_{kj} = \widetilde{A}_{(k+1)j}$, de manera que

$$D(A) = (-1)^{k+j} a_{kj} D(\widetilde{A}_{kj}) + (-1)^{(k+1)+j} a_{kj} D(\widetilde{A}_{kj}).$$

Dado que los términos son iguales y tienen signos opuestos se concluye que

$$D(A) = 0.$$

■

Como se mencionó anteriormente, la unicidad de la función determinante implica que, $D(A) = \det(A)$, independientemente de la columna

j que haya sido elegida para calcular $D(A)$. Se tiene como corolario el resultado conocido como *Teorema de Laplace.* Éste nos dice que el determinante puede calcularse no solo por medio de cualquier columna, sino también cualquier renglón.

Teorema 8.34 *Para cualesquiera i, $j = 1, 2, \ldots, n$ se tiene que*

$$\begin{aligned}\det(A) &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\widetilde{A}_{ij}) \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\widetilde{A}_{ij}).\end{aligned}$$

Demostración. La demostración es inmediata de la Proposición 8.26 que nos dice que $\det(A) = \det(A^T)$. ∎

En este contexto, el determinante $\det(\widetilde{A}_{ij})$ se conoce como el **menor** ij de A. Definiendo

$$C_{ij} = (-1)^{i+j} \det(\widetilde{A}_{ij}), \tag{8.2}$$

el número C_{ij} es conocido como **el cofactor** ij de A. Asimismo, la suma

$$\sum_{j=1}^{n} C_{ij} a_{ij}$$

es llamada **la expansión por menores o por cofactores** a lo largo del $i-ésimo$ renglón de A. Similarmente, la suma

$$\sum_{i=1}^{n} C_{ij} a_{ij}$$

es **la expansión por menores o cofactores** a lo largo de la $j-ésima$ columna de A. El Teorema de Laplace nos dice que

$$\det(A) = \sum_{j=1}^{n} C_{ij} a_{ij} = \sum_{i=1}^{n} C_{ij} a_{ij}.$$

Cálculo de determinantes

En esta sección se utiliza el método de Laplace para el cálculo de determinantes.

Ejemplo 8.35 *Consideremos a la matriz*

$$A = \begin{bmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{bmatrix}$$

y supongamos que deseamos calcular el determinante de A *y fijamos* $i = 1$, *es decir, el desarrollo por cofactores del determinante se hace utilizando el primer renglón. Para ello notamos que*

$$\begin{aligned} \det(A) = (-1)^{1+1} \times 1 \times \det \begin{bmatrix} -5 & 2 \\ 4 & -6 \end{bmatrix} \\ + (-1)^{1+2} \times 3 \times \det \begin{bmatrix} -3 & 2 \\ -4 & -6 \end{bmatrix} \\ +(-1)^{1+3} \times (-3) \times \det \begin{bmatrix} -3 & -5 \\ -4 & 4 \end{bmatrix} \\ &= (30 - 8) + (-3)(18 + 8) + (-3)(-12 - 20) \\ &= 22 - 78 + 96 = 40. \end{aligned}$$

Si alternativamente el desarrollo se realiza tomando la segunda columna de A, *entonces* $j = 2$ *y tenemos que*

$$\begin{aligned} \det(A) = (-1)^{1+2} \times 3 \times \det \begin{bmatrix} -3 & 2 \\ -4 & -6 \end{bmatrix} \\ + (-1)^{2+2} \times (-5) \times \det \begin{bmatrix} 1 & -3 \\ -4 & -6 \end{bmatrix} \\ +(-1)^{3+2} \times 4 \times \det \begin{bmatrix} 1 & -3 \\ -3 & 2 \end{bmatrix} \\ &= (-3)(18 + 8) + (-5)(-6 - 12) - 4(2 - 9) \\ &= -78 + 90 + 28 = 40, \end{aligned}$$

obteniéndose el mismo resultado.

Ejemplo 8.36 *Consideremos a la matriz*

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

y supongamos que deseamos calcular el determinante de B. *Como el segundo renglón contiene el mayor número de entradas iguales a cero,* *tomamos $i = 2$ y notamos que,*

$$\begin{aligned}\det(B) &= \sum_{j=1}^{4}(-1)^{2+j}b_{2j}\det(\widetilde{B}_{2j})\\ &=(-1)^{2+1}\times 2\times\det\begin{bmatrix}1 & 3 & -3\\ -3 & -5 & 2\\ -4 & 4 & -6\end{bmatrix}+0+0+0\\ &= (-1)^{2+1}\times 2\times 40 = -80.\end{aligned}$$

La penúltima igualdad se debe a que en el ejemplo anterior habíamos mostrado que $\det(A) = 40$ *(notar que $\widetilde{B}_{21} = A$).*

Ejemplo 8.37 *Consideremos a la matriz,*

$$A = \begin{bmatrix}8 & 4 & 0\\ 8 & 4 & 1\\ 1 & 3 & 0\end{bmatrix}$$

y supongamos que deseamos calcular el determinante de A utilizando la expansión por cofactores a lo largo de la tercera columna de A (debido a que en ella aparecen el mayor número de ceros). Entonces, tomamos $j = 3$ para obtener

$$\begin{aligned}\det(A) = (-1)^{1+3}\times 0\times\det\begin{bmatrix}8 & 4\\ 1 & 3\end{bmatrix}\\ + (-1)^{2+3}\times 1\times\det\begin{bmatrix}8 & 4\\ 1 & 3\end{bmatrix}\\ + (-1)^{3+3}\times 0\times\det\begin{bmatrix}8 & 4\\ 8 & 4\end{bmatrix}\\ = -(24-4) = -20.\end{aligned}$$

Ejemplo 8.38 *Consideremos a la matriz*

$$B = \begin{bmatrix}8 & 0 & 8 & 1\\ 4 & 0 & 4 & 3\\ 0 & 7 & 1 & 0\\ -1 & 2 & 3 & 0\end{bmatrix}$$

y supongamos que deseamos calcular el determinante de B utilizando la expansión por cofactores a lo largo de la segunda columna de B (simplemente porque en esa columna aparecen “muchos ceros”). Para *ello notamos que*

$$\begin{aligned}
\det(B) &= \sum_{k=1}^{4}(-1)^{k+2} b_{k2}\det(\widetilde{B}_{k2}) \\
&= (-1)^{1+2}\times 0\times\det\begin{bmatrix} 4 & 4 & 3 \\ 0 & 1 & 0 \\ -1 & 3 & 0 \end{bmatrix} + (-1)^{2+2}\times 0\times\det\begin{bmatrix} 8 & 8 & 1 \\ 0 & 1 & 0 \\ -1 & 3 & 0 \end{bmatrix} \\
&+ (-1)^{3+2}\times 7\times\det\begin{bmatrix} 8 & 8 & 1 \\ 4 & 4 & 3 \\ -1 & 3 & 0 \end{bmatrix} + (-1)^{4+2}\times 2\begin{bmatrix} 8 & 8 & 1 \\ 4 & 4 & 3 \\ 0 & 1 & 0 \end{bmatrix} \\
&= -7(-80) + 2(-20) = 520.
\end{aligned}$$

La penúltima igualdad se debe a que en el ejemplo anterior habíamos mostrado que $\det(A) = -20$ *(notar que* $\widetilde{B}_{42} = A$*) y a que*

$$\det\begin{bmatrix} 8 & 8 & 1 \\ 4 & 4 & 3 \\ -1 & 3 & 0 \end{bmatrix} =$$
$$(-1)^{3+1}\times(-1)(24-4) + (-1)^{3+2}\times 3(24-4) + 0 = -80.$$

Ejemplo 8.39 *Sean*

$$A = \begin{bmatrix} 1 & 2 \\ -3 & -4 \end{bmatrix},\ A^T = \begin{bmatrix} 1 & -3 \\ 2 & -4 \end{bmatrix},$$

entonces $\det(A) = -4-(-6) = 2 = \det(A^T)$.

Ejemplo 8.40 *Si*

$$A = \begin{bmatrix} 1 & -2 & 4 & 7 \\ \pi & 13 & 666 & -11 \\ 3 & -1 & 0 & 2 \\ 1 & -2 & 4 & 7 \end{bmatrix},$$

entonces $\det(A) = 0$ *puesto que los renglones* 1 *y* 4 *son iguales. En particular,* A *no es invertible y sus columnas son LD.*

Recordemos que una matriz cuadrada D es diagonal syss todas sus componentes fuera de la diagonal principal son cero (i.e., syss $d_{ij} = 0$ para toda $i \neq j$). Entonces, si

$$D = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix}$$

es una matriz diagonal, por la multilinealidad del determinante se tiene

$$\begin{aligned}\det(A) &= d_{11}d_{22}\cdots d_{nn}\det\begin{bmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{bmatrix}\\ &= d_{11}d_{22}\cdots d_{nn}.\end{aligned}$$

De esta forma, el determinante de una matriz diagonal es, simplemente el producto de los elementos de la diagonal. Este resultado puede extenderse a matrices un poco más generales como se ve a continuación.

Definición 8.41 *Sea $A \in M_{n\times n}$. Decimos que*

(a) A es triangular superior syss todas sus componentes por debajo de la diagonal principal son cero (i.e., syss $a_{ij} = 0$ para toda $i > j$).

(b) A es triangular inferior syss todas sus componentes por arriba de la diagonal principal son cero (i.e., syss $a_{ij} = 0$ para toda $i < j$).

Observación 8.42 *Si A es una matriz de $n \times n$, entonces*

1. *A es triangular superior syss A^T es triangular inferior.*
2. *A es diagonal syss A es triangular superior y triangular inferior.*

Ejemplo 8.43 *Las matrices*

$$\begin{bmatrix}8 & 7\\ 0 & \pi\end{bmatrix} \; y \; \begin{bmatrix}0 & 7 & 6\\ 0 & -1 & 4\\ 0 & 0 & 0\end{bmatrix}$$

son triangulares superiores. De igual manera la matriz identidad y la matriz cero de $n \times n$ son triangulares superiores.

Ejemplo 8.44 *Las matrices*

$$\begin{bmatrix}8 & 0\\ 9 & 0\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\ 2 & 666 & 0\\ 7 & 0 & 9\end{bmatrix} \; y \; \begin{bmatrix}-1 & 0 & 0\\ 0 & \pi & 0\\ 8 & 0 & 0\end{bmatrix}$$

son triangulares inferiores. De igual manera la matriz identidad y la matriz cero de $n \times n$ son triangulares inferiores.

Proposición 8.45 *Sea $A \in M_{n\times n}$. Si A es triangular superior o triangular inferior, entonces*

$$\det(A) = a_{11}a_{22}\ldots a_{nn}.$$

Esto es, el determinante de A es igual al producto de todas las componentes de A que están en su diagonal principal.

Demostración. Se utiliza el desarrollo de Laplace con respecto a la primera columna, en el caso de una matriz triangular superior, o con respecto al primer renglón, si la matriz es triangular inferior. ■

Ejemplo 8.46 *Como la matriz*

$$A = \begin{bmatrix} 7 & 8 & 1 \\ 0 & \pi & -666 \\ 0 & 0 & 9 \end{bmatrix}$$

es triangular superior, $\det(A) = 63\pi$.

Ejemplo 8.47 *Sea*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 1 & 4 & 4 \end{bmatrix}.$$

Utilizando operaciones elementales entre renglones podemos transformar a la matriz A en una triangular como sigue.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 7 \\ 1 & 4 & 4 \end{bmatrix} \overset{-2R_1+R_2}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 1 & 4 & 4 \end{bmatrix} \overset{-R_1+R_3}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}$$
$$\overset{\frac{1}{2}R_3}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \end{bmatrix} \overset{R_2 \leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}.$$

Las primeras dos operaciones no alteran el determinante. Utilizando la multilinealidad y el Corolario 8.13 se tiene,

$$\det(A) = 2(-1)\det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}\right) = -2.$$

La siguiente serie de equivalencias (todas han sido demostradas) recapitula gran parte del trabajo hecho hasta ahora.

Proposición 8.48 *Sea $A \in M_{n\times n}$. Las siguientes afirmaciones son equivalentes.*

1. A es invertible.
2. Las columnas de A forman un conjunto LI.
3. La FER de A es la matriz identidad.
4. La ecuación $A\vec{x} = \vec{0}$ tiene una única solución (a saber, $\vec{x} = \vec{0}$).

5. La ecuación $A\vec{x} = \vec{b}$ tiene solución para toda $\vec{b} \in \mathbb{R}^n$.
6. $Null(A) = \{\vec{0}\}$.
7. A^T es invertible.
8. Los renglones de A forman un conjunto LI.
9. $\det(A) \neq 0$

Ejemplo 8.49 *Consideremos a la matriz*

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 \\ \pi & 10 & 12 & 14 \end{bmatrix}$$

y supongamos que deseamos calcular el determinante de B utilizando la expansión por cofactores a lo largo del primer renglón de B. Para ello, notamos que

$$\det(B) = (1) \times \det \begin{bmatrix} 3 & 4 & 5 \\ 7 & 8 & 9 \\ 10 & 12 & 14 \end{bmatrix} = 1 \times 0 = 0,$$

pues

$$\det \begin{bmatrix} 3 & 4 & 5 \\ 7 & 8 & 9 \\ 10 & 12 & 14 \end{bmatrix} = 0$$

ya que el tercer renglón de esta matriz es combinación lineal de los otros dos renglones.

Ejemplo 8.50 *Supongamos que deseamos encontrar todos los valores de a tales que el conjunto*

$$\left\{ \begin{bmatrix} a-2 \\ a \\ \pi \end{bmatrix}, \begin{bmatrix} 0 \\ a-3 \\ a^4+9 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} \right\}$$

es LI. Para ello construimos la matriz cuyas columnas son estos vectores:

$$A = \begin{bmatrix} a-2 & 0 & 0 \\ a & a-3 & 0 \\ \pi & a^4+9 & 6 \end{bmatrix}.$$

Lo que estámos buscando son todos los valores de a tales que $\det(A) \neq 0$. Como A es triangular inferior,

$$\det(A) = (a-2)(a-3)6.$$

Por lo tanto, nuestro conjunto es LI syss $a \neq 2$ y $a \neq 3$.

Regla de Cramer

Finalizamos este capítulo con un método sumamente popular (aunque no muy eficiente) para resolver sistemas de ecuaciones lineales que tienen tantas ecuaciones como variables. La utilidad de este resultado es más bien en el cálculo diferencial de varias variables, dado que se utiliza para la demostración de un teorema importante: el Teorema de la Función Implícita.

Sea $A \in M_{n\times n}$, recordemos la definición de cofactor dada en (8.2) como

$$C_{ij} = (-1)^{i+j} det(\widetilde{A}_{ij}),$$

en donde $\widetilde{A}_{ij}$ es la matriz que se obtiene a partir de A eliminando el renglón i y la columna j. Se define la **matriz adjunta** de A, denotada por $adj(A)$, como la transpuesta de la matriz cuyas entradas son los cofactores de A, concretamente

$$(adj(A))_{ij} = C_{ji}.$$

La siguiente proposición, que enunciamos sin demostrar, nos dice que la matriz adjunta de A está íntimamente relacionada con $\det(A)$.

Proposición 8.51 *Sea $A \in M_{n\times n}$, entonces se cumple*

$$adj(A)A = (A)\, adj(A) = \det(A)I.$$

Para el caso en el cual la matriz A es invertible, se satisface $\det(A) \neq 0$ y se tiene el siguiente corolario.

Corolario 8.52 *Sea $A \in M_{n\times n}$ una matriz invertible, entonces se cumple*

$$A^{-1} = \left(\frac{1}{\det(A)}\right) adj(A).$$

Una aplicación de este corolario es la llamada **Regla de Cramer** para resolver ciertos sistemas de ecuaciones.

Teorema 8.53 (Regla de Cramer) *Sea $A \in M_{n\times n}$ una matriz invertible, sea $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ un vector de $\mathbb{R}^n$ y consideremos al sistema lineal $A\vec{x} = \vec{b}$ en donde*

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Entonces, este sistema tiene solución única y además, para cada $j \in \{1, \ldots, n\}$*, se tiene que*

$$x_j = \frac{\det(A^{j \leftrightarrow \vec{b}})}{\det(A)},$$

en donde $A^{j \leftrightarrow \vec{b}}$ *es la matriz que resulta de reemplazar la* j*-ésima columna de* A *por el vector* $\vec{b}$ *de términos independientes.*

Demostración. Como A es invertible sabemos que la solución del sistema lineal $A\vec{x} = \vec{b}$ es única y puede expresarse como

$$\vec{x} = A^{-1}\vec{b}.$$

Utilizando el Corolario 8.52, reescribimos esta expresión para la solución como

$$\vec{x} = \left(\frac{1}{\det(A)}\right) adj(A)\vec{b},$$

o bien

$$\begin{aligned}
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} &= \left(\frac{1}{\det(A)}\right) \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\
&= \left(\frac{1}{\det(A)}\right) \begin{bmatrix} \sum_{i=1}^{n} b_i C_{i1} \\ \sum_{i=1}^{n} b_i C_{i2} \\ \vdots \\ \sum_{i=1}^{n} b_i C_{in} \end{bmatrix}.
\end{aligned}$$

Observamos que las sumas $\sum_{i=1}^{n} b_i C_{ij}$ coinciden con la expansión por cofactores de Laplace (utilizando la columna j) para $\det(A)$, excepto que la $j-\acute{e}sima$ columna ha sido reemplazada por el vector de términos independientes $\vec{b}$. Tenemos así que

$$\sum_{i=1}^{n} b_i C_{ij} = \det(A^{j \leftrightarrow \vec{b}})$$

y se concluye que

$$x_j = \frac{\det(A^{j \leftrightarrow \vec{b}})}{\det(A)}.$$

■

Ejemplo 8.54 *Supongamos que deseamos encontrar la solución del sistema*

$$\begin{aligned} 3x_1 + x_2 + 0 &= 5, \\ -x_1 + 2x_2 + x_3 &= -2, \\ 0 - x_2 + 2x_3 &= -1. \end{aligned}$$

Para ello, tomamos

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 2 & 1 \\ 0 & -1 & 2 \end{bmatrix} \ y \ \vec{b} = \begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix},$$

por lo que nuestro sistema puede reescribirse como $A\vec{x} = \vec{b}$. *Notamos ahora que como*

$$\det(A) = 17 \neq 0,$$

puede utilizarse la Regla de Cramer para encontrar a x_1, x_2 *y* x_3. *Tenemos así que*

$$x_1 = \frac{\det(A^{1\leftrightarrow\vec{b}})}{\det(A)} = \frac{1}{17} \det \begin{bmatrix} 5 & 1 & 0 \\ -2 & 2 & 1 \\ -1 & -1 & 2 \end{bmatrix} = \frac{28}{17},$$

$$x_2 = \frac{\det(A^{2\leftrightarrow\vec{b}})}{\det(A)} = \frac{1}{17} \det \begin{bmatrix} 3 & 5 & 0 \\ -1 & -2 & 1 \\ 0 & -1 & 2 \end{bmatrix} = \frac{1}{17},$$

y

$$x_3 = \frac{\det(A^{3\leftrightarrow\vec{b}})}{\det(A)} = \frac{1}{17} \det \begin{bmatrix} 3 & 1 & 5 \\ -1 & 2 & -2 \\ 0 & -1 & -1 \end{bmatrix} = \frac{-8}{17}.$$

Capítulo 9

Factorización PLU(opcional)

Introducción

Supongamos que $A \in M_{n\times n}$ es una matriz triangular superior con $a_{ii} \neq 0$, $i = 1, 2, \ldots, n$, de manera que

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}.$$

Si consideramos el sistema lineal $A\vec{x} = \vec{b}$ con

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n,$$

observemos que las ecuaciones correspondientes están dadas como

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

Este sistema es consistente y puede resolverse fácilmente en forma recursiva, comenzando con la última ecuación.

En forma similar, si $A \in M_{n\times n}$ es una matriz triangular inferior, con $a_{ii} \neq 0$, $i = 1, 2, \ldots, n$, dada como

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix},$$

las ecuaciones correspondientes al sistema lineal el sistema lineal $A\vec{x} = \vec{b}$ están dadas por

$$\begin{aligned} a_{11}x_1 &= b_{1,} \\ a_{21}x_1 + a_{22}x_2 &= b_{2,} \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n. \end{aligned}$$

En este caso, el sistema es consistente y puede resolverse recursivamente, comenzando por la primera ecuación.

Ejemplo 9.1 *Sean*

$$A = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix}, \ \vec{b} = \begin{bmatrix} -2 \\ 15 \\ 10 \end{bmatrix}.$$

El sistema de ecuaciones correspondiente está dado por

$$\begin{aligned} -x_1 + 2x_2 + x_3 &= -2, \\ 2x_2 - x_3 &= 15, \\ 2x_3 &= 10. \end{aligned}$$

De la tercera ecuación se tiene que $x_3 = 5$; *sustituyendo este resultado en la segunda ecuación se tiene que*

$$2x_2 - 5 = 15,$$

por lo tanto, $x_2 = 10$. *Finalmente, sustituyendo para* x_2 *y* x_3 *en la primer ecuación tenemos*

$$-x_1 + 2 \times 10 + 5 = -2,$$

por lo que $x_1 = 27$. *En este caso resolvimos el sistema de ecuaciones de "abajo hacia arriba" (lo que se denomina* ***sustitución hacia atrás****) y, si la matriz de coeficientes hubiese sido triangular inferior, se resolvería de "arriba hacia abajo" (o* ***sustitución hacia adelante****).*

Factorización LU

Las consideraciones anteriores resaltan la simplicidad de los sistemas lineales correspondientes a matrices triangulares. Convertir una matriz cuadrada en una triangular (superior o inferior) mediante operaciones elementales de renglones, requiere de menos pasos que la obtención de la FER. Por esta razón es deseable, computacionalmente hablando, la descomposición de una matriz en matrices triangulares.

Sea

$$A(1) = \begin{bmatrix} a_{11}^1 & a_{12}^1 & \cdots & a_{1n}^1 \\ a_{21}^1 & a_{22}^1 & \cdots & a_{2n}^1 \\ \vdots & \vdots & & \vdots \\ a_{n1}^1 & a_{n2}^1 & \cdots & a_{nn}^1 \end{bmatrix}$$

con $a_{11}^1 \neq 0$. Queremos convertir a $A(1)$ en una matriz triangular superior utilizando únicamente matrices elementales que corresponden a sumar un múltiplo de un renglón i a otro renglón $j > i$; es decir, matrices elementales del tipo

$$\begin{bmatrix} 1 & 0 & & & 0 \\ 0 & 1 & & 0 & \\ & & \ddots & & \\ & & c & & 0 \\ 0 & & & & 1 \end{bmatrix}. \qquad (\bigstar)$$

El procedimiento sería como sigue: como $a_{11}^1 \neq 0$, eliminamos a_{21}^1 realizando la operación $-\frac{a_{21}^1}{a_{11}^1}R_1 + R_2$ sobre la matriz A^1. Equivalentemente, se multiplica $A(1)$ por la matriz elemental

$$L(2,1) = \begin{bmatrix} 1 & 0 & & 0 \\ -\frac{a_{21}^1}{a_{11}^1} & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}.$$

De forma similar eliminamos $a_{31}^1, \ldots, a_{n1}^1$, multiplicando por matrices elementales[2] $L(3,1), \ldots, L(n,1)$ hasta obtener

$$L(1)A(1) = A(2),$$

[1]La palabra "eliminar", en este contexto, significa convertir en cero la entrada a_{21}^1 de la matriz mediante la operación elemental correspondiente.

[2]Utilizamos la letra L para denotar a estas matrices elementales para hacer énfasis en que se trata de matrices triangulares inferiores.

en donde la matriz $A(2)$ es de la forma

$$\begin{bmatrix} a_{11}^2 & a_{12}^2 & \cdots & a_{1n}^2 \\ 0 & a_{22}^2 & \cdots & a_{2n}^2 \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^2 & \cdots & a_{nn}^2 \end{bmatrix}$$

y

$$L(1) = L(n,1)\cdots L(2,1) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\frac{a_{21}^1}{a_{11}^1} & 1 & 0 & \cdots & 0 \\ -\frac{a_{31}^1}{a_{11}^1} & 0 & 1 & & \vdots \\ \vdots & \vdots & & \ddots & \\ -\frac{a_{n1}^1}{a_{11}^1} & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Suponiendo que $a_{22}^2 \neq 0$ se procede de la misma forma, multiplicando por matrices elementales del tipo ($\bigstar$) hasta obtener una matriz $A(3)$ de la forma

$$L(2)A(2) = A(3) = \begin{bmatrix} a_{11}^3 & a_{12}^3 & a_{13}^3 & \cdots & a_{1n}^3 \\ 0 & a_{22}^3 & a_{23}^3 & \cdots & a_{2n}^3 \\ 0 & 0 & a_{33}^3 & \cdots & a_{3n}^3 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & a_{n3}^3 & \cdots & a_{nn}^3 \end{bmatrix},$$

en donde

$$L(2) = L(n,2)\cdots L(2,2) = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -\frac{a_{32}^2}{a_{22}^2} & 1 & & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & -\frac{a_{n2}^2}{a_{22}^2} & 0 & \cdots & 1 \end{bmatrix}.$$

Siguiendo con este proceso, y suponiendo que para la matriz $A(k)$ se cumple que $a_{kk}^k \neq 0$ ($k = 1, \ldots, n-1$), podemos eliminar todas las entradas de la forma a_{ij}^k con $i > j$ con lo que $A(n-1)$ será una matriz triangular superior que se denota por

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix}.$$

Si

$$\hat{L} = \prod_{i=1}^{n} L(i),$$

en donde el símbolo "Π" se refiere al producto, es un ejercicio simple verificar que

$$\hat{L} = \begin{bmatrix} 1 & 0 & & & 0 \\ -\frac{a_{21}^1}{a_{11}^1} & 1 & & 0 & \\ -\frac{a_{31}^1}{a_{11}^1} & -\frac{a_{32}^2}{a_{22}^2} & 1 & & \\ \vdots & & & \ddots & 0 \\ -\frac{a_{n1}^1}{a_{11}^1} & -\frac{a_{n2}^2}{a_{22}^2} & \cdots & -\frac{a_{n(n-1)}^{n-1}}{a_{(n-1)(n-1)}^{n-1}} & 1 \end{bmatrix},$$

en donde notamos que se trata de una matriz triangular inferior con unos en su diagonal.

Sea $A = A(1) \in M_{n\times n}$ con $a_{11}^1 \neq 0$ y supongamos que puede construirse la sucesión de matrices $A(2), \ldots A(n-1)$ como arriba (sabemos que esto es posible si $a_{22}^2, \ldots, a_{(n-1)(n-1)}^{n-1} \neq 0$). Entonces, las consideraciones anteriores nos dicen que existen $U \in M_{n\times n}$,triangular superior y $\hat{L} \in M_{n\times n}$, triangular inferior, tales que se cumple

$$\hat{L}A = U. \tag{9.1}$$

Ahora bien, como $\hat{L}$ es invertible ($\det(\hat{L}) = 1$). Es fácil verificar que

$$\hat{L}^{-1} = \begin{bmatrix} 1 & 0 & & & 0 \\ \frac{a_{21}^1}{a_{11}^1} & 1 & & 0 & \\ \frac{a_{31}^1}{a_{11}^1} & \frac{a_{32}^2}{a_{22}^2} & 1 & & \\ \vdots & & & \ddots & 0 \\ \frac{a_{n1}^1}{a_{11}^1} & \frac{a_{n2}^2}{a_{22}^2} & \cdots & \frac{a_{n(n-1)}^{n-1}}{a_{(n-1)(n-1)}^{n-1}} & 1 \end{bmatrix}.$$

En esta situación se dice que la matriz A admite una **factorización** LU y se tiene

$$A = LU,$$

tomando $L = \hat{L}^{-1}$. Observemos que esto último sucede cuando A puede transformarse en una matriz triangular superior U, utilizando únicamente operaciones elementales entre renglones del tipo 3 (equivalentemente, multiplicando por matrices elementales del tipo (★))[3]

Ejemplo 9.2 *Sea*

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

[3]En el caso particular para el cual $L = U^T$, esta factorización se denomina **factorización de Cholesky**.

entonces utilizando únicamente operaciones elementales del tipo 3 tenemos,

$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 3 \\ -3 & 1 & -4 \end{bmatrix} \overset{-2R_1+R_2}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ -3 & 1 & -4 \end{bmatrix} \overset{3R_1+R_3}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & -2 & 2 \end{bmatrix}$$
$$\overset{2R_2+R_3}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Se tiene así que

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -3 & -1 & 1 \end{bmatrix}, \ U = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

y $A = LU$.

Ejemplo 9.3 *Sea* $A \in M_{3\times 3}$ *tal que su descomposición* LU *está dada como*

$$\underbrace{\begin{bmatrix} 1 & -1 & 2 \\ -2 & 5 & -4 \\ 2 & -5 & 5 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U}.$$

Considérese el sistema lineal (que tendrá solución única pues A *es invertible)*

$$\begin{bmatrix} 1 & -1 & 2 \\ -2 & 5 & -4 \\ 2 & -5 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 6 \end{bmatrix},$$

mismo que puede escribirse como

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 6 \end{bmatrix}.$$

De aquí definimos

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{9.2}$$

y resolvemos

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 6 \end{bmatrix},$$

obteniendo

$$\begin{aligned} x' &= 3, \\ -2x' + y' &= 0, \ \textit{por lo que } y' = 6, \\ 2x' - y' + z' &= 6, \ \textit{por lo que } z' = 6. \end{aligned}$$

Después sustituimos estos valores en (9.2) y resolvemos

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 6 \end{bmatrix},$$

obteniendo

$$\begin{aligned} z &= 6, \\ 3y &= 6, \ \textit{por lo que } y = 2, \\ x - y + 2z &= 3, \ \textit{por lo que } x = -7. \end{aligned}$$

Ejemplo 9.4 *Sea*

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

en este caso A es triangular superior por lo que $L = I$ y la factorización LU es simplemente

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

Factorización PLU

Supongamos que tenemos una matriz $A \in M_{n\times n}$ que no admite una factorización LU. Por ejemplo, la matriz

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

no puede llevarse a una matriz triangular superior utilizando únicamente operaciones elementales del tipo 3 (el problema es que $a_{11} = 0$). En este caso se requiere intercambiar los renglones 1 y 2 para llegar a la matriz triangular superior. Ahora bien, mediante el intercambio adecuado de renglones, una matriz invertible siempre puede transformarse en una matriz que admite una factorización LU.

Ejemplo 9.5 *La matriz*

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

no admite una factorización LU (esto puede verificarse fácilmente). Sin embargo, intercambiando los renglones 2 y 3 obtenemos la matriz

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

por lo tanto

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \overset{-R_1+R_3}{\sim} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

y se tiene la siguiente factorización LU :

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

Definición 9.6 *Sean* $E(1), \ldots, E(k) \in M_{n\times n}$ *matrices elementales obtenidas intercambiando renglones en la matriz identidad. La matriz* P *definida como*

$$P = E(k) \ldots E(1),$$

se denomina una ***matriz de permutación.***

Ejemplo 9.7 *La matriz*

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

es una matriz de permutación que intercambia los renglones 1 y 3 y 2 y 4.

La siguiente proposición resume la discusión anterior.

Proposición 9.8 *Sea* $A \in M_{n\times n}$*, entonces existen matrices* $L, U, P' \in M_{n\times n}$*, en donde* L *es triangular inferior con unos en su diagonal,* U *es triangular superior y* P' *es una matriz de permutación, tales que se cumple*

$$P'A = LU,$$

es decir, la matriz $P'A$ *admite una factorización* LU.

Observación 9.9 *Como P' es invertible y su inversa es otra matriz de permutación (concretamente, $(P')^{-1} = (P')^T$, se deja esta demostración como ejercicio al lector), definiendo*

$$P = (P')^{-1} = (P')^T$$

se sigue que

$$A = PLU.$$

A esta factorización de la matriz A se le conoce como ***factorización PLU****.*

Ejemplo 9.10 *La matriz*

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

no tiene una descomposición LU, sin embargo la matriz dada por,

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \begin{bmatrix} 0 & -1 & 2 \\ -2 & 0 & -4 \\ 2 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 & -4 \\ 0 & -1 & 2 \\ 2 & -1 & 1 \end{bmatrix},$$

obtenida intercambiando los renglones 1 y 2 puede descomponerse como

$$\begin{bmatrix} -2 & 0 & -4 \\ 0 & -1 & 2 \\ 2 & -1 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} -2 & 0 & -4 \\ 0 & -1 & 2 \\ 0 & 0 & -5 \end{bmatrix}}_{U}.$$

Dado que, en este caso

$$P^{-1} = P^T = P,$$

se sigue que

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

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} -2 & 0 & -4 \\ 0 & -1 & 2 \\ 0 & 0 & -5 \end{bmatrix}}_{U}.$$

Observación 9.11 *Recordemos que el determinante de las matrices triangulares es simplemente el producto de los elementos de su diagonal y el determinante preserva productos, por lo tanto, las descomposiciones LU y PLU nos facilitan el cálculo del determinante. Por ejemplo, utilizando la descomposición del Ejemplo 9.10 se tiene que*

$$\det\left(\begin{bmatrix} 0 & -1 & 2 \\ -2 & 0 & -4 \\ 2 & -1 & 1 \end{bmatrix}\right) =$$
$$\det\left(\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -2 & 0 & -4 \\ 0 & -1 & 2 \\ 0 & 0 & -5 \end{bmatrix}\right)$$
$$= -1 \times 1 \times (-10) = 10.$$

Capítulo 10

Subespacios

Introducción

Cuando tenemos conjuntos con alguna estructura algebraica es común que éstos posean subconjuntos con el mismo tipo de estructura. Tenemos así que los racionales ($\mathbb{Q}$) son un subconjunto de los reales ($\mathbb{R}$) y ambos conjuntos de números tienen el mismo comportamiento con respecto a las operaciones de suma y producto entre sus elementos[1]. En el caso de los espacios vectoriales $\mathbb{R}^n$, éstos también tienen subconjuntos que son cerrados bajo las operaciones de suma y producto por escalares. Formalmente tenemos la siguiente definición.

Definición 10.1 *Sea S un subconjunto de $\mathbb{R}^n$. Decimos que S es* ***un subespacio de*** $\mathbb{R}^n$ *syss se cumplen las siguientes tres condiciones:*

1. $\vec{0}_n \in S$, *donde* $\vec{0}_n$ *denota al vector cero de* $\mathbb{R}^n$.

2. *S es cerrado bajo sumas. Esto es, si $\vec{x}$ y $\vec{y}$ son elementos de S, entonces $\vec{x}+\vec{y} \in S$.*

3. *S es cerrado bajo multiplicación escalar. Esto es, si $\vec{x}$ es un elemento de S y c es un escalar, entonces $c\vec{x} \in S$.*

Observación 10.2 *El lector debe notar que, con las propiedades de la definición anterior, los elementos del conjunto S satisfacen las propiedades V1 a V8 dadas en 3.6. De esta forma, S es un verdadero*

[1] Concretamente, decimos que tanto $\mathbb{Q}$ como $\mathbb{R}$ son lo que se denominan **campos**.

*espacio vectorial contenido en $\mathbb{R}^n$. En ocasiones, nos referimos a estos subespacios como **subespacios lineales**.*

Un resultado inmediato a partir de las propiedades de la definición de subespacio es el siguiente.

Proposición 10.3 *Si $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ son elementos de un subespacio S, entonces $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\} \subset S$.*

Demostración. Supongamos que $\vec{v} \in span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$. Esto es, existen escalares $c_1, c_2, \ldots c_m$ tales que

$$v = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m.$$

Debemos probar que $\vec{v} \in S$. Sin embargo, como $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m \in S$ y S es cerrado bajo multiplicación escalar se tiene que

$$c_1\vec{v}_1, c_2\vec{v}_2, \ldots, c_m\vec{v}_m \in S.$$

Esto último junto con el hecho de que S es cerrado bajo sumas implican que

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m \in S.$$

■

En lo que sigue veremos bajo que circunstancias algunos subconjuntos de $\mathbb{R}^n$ son o no subespacios. En vista de la proposición anterior, el primer candidato a examinar es el span de un conjunto de vectores como se ve a continuación.

Proposición 10.4 *Si $\vec{v}_1, \vec{v}_2, \ldots \vec{v}_p$ son vectores de $\mathbb{R}^n$, entonces*

$$span\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_p\}$$

es un subespacio de $\mathbb{R}^n$.

Demostración. Como

$$\vec{0} = 0\vec{v}_1 + 0\vec{v}_2 + \cdots + 0\vec{v}_p$$

se tiene que $\vec{0} \in span\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_p\}$. Si $\vec{x}, \vec{y} \in span\{\vec{v}_1, \vec{v}_2, \ldots \vec{v}_p\}$, entonces existen x_i, y_i, $i = 1, \ldots, p$, escalares, tales que

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_p\vec{v}_p,$$
$$\vec{y} = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_p\vec{v}_p.$$

Por lo tanto,

$$\vec{x} + \vec{y} = (x_1 + y_1)\,\vec{v}_1 + (x_2 + y_2)\,\vec{v}_2 + \cdots + (x_p + y_p)\,\vec{v}_p$$

y $\vec{x}+\vec{y} \in span\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_p\}$. Finalmente, si $c \in \mathbb{R}$ y $\vec{x} \in span\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_p\} \in span\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_p\}$ es como arriba, se tiene que

$$c\vec{x} = cx_1\vec{v}_1 + cx_2\vec{v}_2 + \cdots + cx_p\vec{v}_p$$

y $c\vec{x} \in span\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_p\}$. Como $span\{\vec{v}_1, \vec{v}_2, \dots \vec{v}_p\}$ cumple con las propiedades de la Definición 10.1 se concluye que es un subespacio de $\mathbb{R}^n$. ■

Corolario 10.5 *$\{\vec{0}_n\}$ y $\mathbb{R}^n$ son subespacios de $\mathbb{R}^n$.*

Demostración. Basta notar que $\{\vec{0}_n\} = span\{\vec{0}_n\}$ y que $\mathbb{R}^n$ evidentemente cumple las propiedades de la Definición 10.1. ■

Definición 10.6 *$\{\vec{0}_n\}$ y $\mathbb{R}^n$ son conocidos como los **subespacios triviales** de $\mathbb{R}^n$.*

ES claro que todos los espacios $\mathbb{R}^n$ contienen al menos los subespacios triviales. ¿Qué otros subespacios contendrán? En el caso de $n = 1$, es decir, $\mathbb{R}^1 = \mathbb{R}$, no existen subespacios no triviales; en efecto, si $S \subset \mathbb{R}$ y $S \neq \{\vec{0}_n\}$ es un subespacio distinto del cero, entonces existirá un vector $\vec{v} \in S$, $\vec{v} \neq \vec{0}$ y por la Proposición 10.3

$$span\{\vec{v}\} \subset S.$$

Como $\{\vec{v}\}$ es un conjunto LI en $\mathbb{R}$, el Corolario 4.31 nos dice que

$$span\{\vec{v}\} = \mathbb{R},$$

de manera que $S = \mathbb{R}$.

Si $n = 2$ y S es un subespacio no trivial de $\mathbb{R}^2$, existirá un vector $\vec{v} \neq \vec{0}$ tal que $\vec{v} \in S$. Más aún, cualquier otro elemento de S debe ser multiplo de $\vec{v}$ porque en caso contrario, existiría $\vec{w} \in S$ tal que $\vec{w} \notin span\{\vec{v}\}$, con lo cual $\{\vec{v}, \vec{w}\}$ sería un conjunto LI y, por el Corolario 4.31 y la Proposición 10.4, tendríamos que

$$\mathbb{R}^2 = span\{\vec{v}, \vec{w}\} \subset S.$$

Por lo tanto, $S = \mathbb{R}^2$, lo cual es imposible pues $S \subsetneq \mathbb{R}^2$. De aquí que si S es un subespacio no trivial de $\mathbb{R}^2$, S puede visualizarse en el plano como una recta que pasa por el origen.

De manera análoga, puede justificarse que si S es un subespacio no trivial de $\mathbb{R}^3$, entonces S podrá ser de la forma $span\{\vec{v}\}$, en cuyo caso se visualiza en el espacio como una recta que pasa por el origen o bien, $S = span\{\vec{v}, \vec{w}\}$ con $\{\vec{v}, \vec{w}\} \subset S$ un conjunto LI, de forma que S representa un plano que contiene al origen. S no puede poseer más de dos vectores linealmente independientes ya que si $\{\vec{u}, \vec{v}, \vec{w}\}$ fuese tal conjunto de vectores, entonces, una vez más por el Corolario 4.31 y la Proposición 10.4 tendríamos

$$\mathbb{R}^3 = span\{\vec{u}, \vec{v}, \vec{w}\} \subset S.$$

De aquí que $S = \mathbb{R}^3$ lo cual sería imposible ya que $S \subsetneq \mathbb{R}^3$.

Generalizando los ejemplos anteriores, más adelante veremos que los subespacios de $\mathbb{R}^n$ serán, los subespacios triviales $\{\vec{0}_n\}$ y $\mathbb{R}^n$ y todos los subespacios S tales que

$$S = span\{v_1, v_2, \ldots, v_m\},$$

en donde $m < n$ y el conjunto $\{v_1, v_2, \ldots, v_m\}$ es LI.

En $\mathbb{R}^3$, la intersección de dos planos (distintos) que atraviesan el origen es una recta que pasa por el origen, de manera que dicha intersección también es un subespacio de $\mathbb{R}^3$. De forma más general, tenemos el siguiente resultado.

Proposición 10.7 *Si S_1 y S_2 son subespacios de $\mathbb{R}^n$, entonces $S_1 \cap S_2$ es subespacio de $\mathbb{R}^n$.*

Demostración. Verificamos, primero que $\vec{0}_n \in S_1 \cap S_2$, lo cual es obvio pues al ser S_1 y S_2 subespacios, $\vec{0}_n \in S_1$ y $\vec{0}_n \in S_2$. Verifiquemos ahora que $S_1 \cap S_2$ es cerrado bajo sumas. Si $\vec{x}$ y $\vec{y}$ son elementos de $S_1 \cap S_2$, debemos verificar que

$$\vec{x} + \vec{y} \in S_1 \cap S_2.$$

Ahora bien, como

$$\vec{x}, \vec{y} \in S_i, \ i = 1, 2$$

y S_1, S_2 son cerrados bajo sumas se tiene que

$$\vec{x} + \vec{y} \in S_1 \cap S_2.$$

Finalmente, la verificación de que $S_1 \cap S_2$ es cerrado bajo multiplicación escalar se deja como ejercicio para el lector. ■

Concluimos esta sección con algunos ejemplos de subconjuntos que satisfacen dos de las propiedades que definen a un subespacio pero que no son subespacios.

Ejemplo 10.8 *Sean*

$$S_1 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x, y \geq 0 \right\} \text{ y } S_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x, y \leq 0 \right\}.$$

y supongamos que deseamos determinar si los siguientes conjuntos son subespacios de $\mathbb{R}^2$: S_1, S_2 y $S_1 \cup S_2$. En el caso de S_1 es claro que satisface la primera y la segunda propiedad de subespacio, sin embargo,

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \in S_1 \text{ pero } -2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \notin S_1,$$

por lo que S_1 no es subespacio de $\mathbb{R}^2$ (pues S_1 no es cerrado bajo multiplicación escalar). Análogamente puede verificarse que S_2 tampoco es subespacio de $\mathbb{R}^2$ (pues S_2 no es cerrado bajo multiplicación escalar). Finalmente, en el caso de $S_1 \cup S_2$ observamos que se satisfacen la primera y la tercera propiedad, pero

$$\begin{bmatrix}1\\0\end{bmatrix} \in S_1 \cup S_2, \begin{bmatrix}0\\-1\end{bmatrix} \in S_1 \cup S_2, \ y \ \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\-1\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix} \notin S_1 \cup S_2,$$

por lo que $S_1 \cup S_2$ no es cerrado bajo sumas y, en consecuencia, no es subespacio de $\mathbb{R}^2$.

Ejemplo 10.9 *Consideremos a los subespacios (es fácil verificar que lo son)*

$$S_1 = \left\{ \begin{bmatrix}x\\0\end{bmatrix} \mid x \in \mathbb{R} \right\} \ y \ S_2 = \left\{ \begin{bmatrix}0\\y\end{bmatrix} \mid y \in \mathbb{R} \right\}.$$

y supongamos que deseamos determinar si $S_1 \cup S_2$ también es subespacio de $\mathbb{R}^2$. Para ello, notamos que

$$\begin{bmatrix}1\\0\end{bmatrix} \in S_1 \cup S_2, \begin{bmatrix}0\\\pi\end{bmatrix} \in S_1 \cup S_2, \ pero \ \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\\pi\end{bmatrix} = \begin{bmatrix}1\\\pi\end{bmatrix} \notin S_1 \cup S_2,$$

por lo que $S_1 \cup S_2$ no es subespacio de $\mathbb{R}^2$ (pues $S_1 \cup S_2$ no es cerrado bajo sumas).

Espacio nulo y rango

Dada una transformación lineal,

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$

se tienen dos subespacios, uno de $\mathbb{R}^n$ y otro de $\mathbb{R}^m$, naturalmente asociados a ella. Recordemos que

$$Rango(T) = \{T(\vec{x}) \in \mathbb{R}^m\} \subset \mathbb{R}^m.$$

Si $A \in M_{m\times n}$ es la matriz estándar que representa a T, sabemos que

$$Rango(T) = span\{A^1, \ldots, A^n\}$$

y la Proposición 10.4 nos dice que $Rango(T)$ es un subespacio de $\mathbb{R}^m$ asociado a T.

Ejemplo 10.10 *Sea $T:\mathbb{R}^4 \longrightarrow \mathbb{R}^2$ la transformación lineal dada por*

$$T\left(\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}\right) = \begin{bmatrix} x-w \\ w \end{bmatrix}$$

y supongamos que queremos encontrar un conjunto LI de $\mathbb{R}^2$ cuyo span sea igual al rango de T. La matriz A que representa a T está dada por

$$\begin{aligned} A &= [T(\vec{e}_1)\ T(\vec{e}_2)\ T(\vec{e}_3)\ T(\vec{e}_4)] \\ &= \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

y como $Rango(T) = span\{A^1, A^2, A^3, A^4\}$, es claro que el conjunto

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

es LI y genera a $Rango(T)$.

El subespacio de $\mathbb{R}^n$ asociado a la transformación lineal T tiene que ver con los elementos que, bajo la transformación, son enviados al vector $\vec{0}_m$. Formalmente se define el siguiente subespacio.

Definición 10.11 *Sea $T:\mathbb{R}^n \longrightarrow R^m$ una transformación lineal. Definimos al* ***espacio nulo*** *de T, denotado por $Null(T)$, como el conjunto*

$$Null(T) = \{\vec{x} \in \mathbb{R}^n \mid T(\vec{x}) = \vec{0}_m\}.$$

El espacio nulo de T será, en efecto, un subespacio de $\mathbb{R}^n$ como se ve a continuación.

Proposición 10.12 *Si $T:\mathbb{R}^n \longrightarrow R^m$ una transformación lineal, entonces $Null(T)$ es un subespacio de $\mathbb{R}^n$.*

Demostración. Como $T(\vec{0}_n) = \vec{0}_m$ se tiene que $\vec{0}_n \in Null(T)$. Ahora bien, si $\vec{x}, \vec{y} \in Null(T)$, entonces

$$T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y}) = \vec{0}_m + \vec{0}_m = \vec{0}_m,$$

por lo que

$$\vec{x}+\vec{y} \in Null(T).$$

Finalmente, dados $\vec{x} \in Null(T)$ y c un escalar, se tiene que

$$T(c\vec{x}) = cT(\vec{x}) = c\vec{0}_m = \vec{0}_m,$$

de forma que $c\vec{x} \in Null(T)$. Concluimos que $Null(T)$ cumple las tres propiedades requeridas de un subespacio. ∎

Observación 10.13 *En algunos textos, al espacio nulo de T se le conoce como el* ***kernel*** [2] *de T o* ***nucleo*** *de T.*

Si $A \in M_{m\times n}$ es la matriz estándar que representa a T, es evidente que

$$\begin{aligned} Null(T) &= \{\vec{x} \in \mathbb{R}^n \mid T(\vec{x}) = \vec{0}_m\} \\ &= \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}_m\}, \end{aligned}$$

es decir, $Null(T)$ corresponde al conjunto de soluciones del sistema homogéneo $A\vec{x} = \vec{0}_m$. La siguiente definición y la proposición que la sigue son una consecuencia natural de que la multiplicación por cualquier matriz $A \in M_{m\times n}$ determina una transformación lineal.

Definición 10.14 *Sea $A \in M_{m\times n}$. Definimos el espacio* ***nulo*** *de A, denotado por $Null(A)$, como el conjunto*

$$Null(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}_m\},$$

es decir, se trata del conjunto de soluciones del sistema homogéneo.

Observación 10.15 *Es claro que si $T : \mathbb{R}^n \longrightarrow R^m$ es una transformación lineal y $A \in M_{m\times n}$ es la matriz estándar asociada a T, por definición,*

$$Null(T) = Null(A).$$

Observación 10.16 *Observemos que si la transformación T es suprayectiva, entonces $Rango(T) = \mathbb{R}^m$ y si T es inyectiva $Null(T) = \{\vec{0}_n\}$*

Proposición 10.17 *Si $A \in M_{m\times n}$, entonces $Null(A)$ es un subespacio de $\mathbb{R}^n$.*

Ejemplo 10.18 *Sea*

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 5 \end{bmatrix}$$

y supongamos que deseamos expresar a $Null(A)$ como un cierto "span". Resolvemos el sistema homogéneo

$$\begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

notando que la FER de la matriz aumentada del sistema está dada por

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right].$$

[2] La palabra "kernel" en el idioma inglés,se refiere al centro o corazón de una fruta o vegetal.

Por lo tanto, si $\vec{x} \in Null(A)$, *sus componentes satisfacen*

$$x_1 + 2x_3 = 0,\ x_2 + x_3 = 0,\ x_3 \in \mathbb{R}$$

y concluimos que, por la Proposición 10.4,

$$Null(A) = \left\{ x_3 \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\} = span \left\{ \begin{bmatrix} -2 \\ -1 \\ 1 \end{bmatrix} \right\}$$

es un subespacio de $\mathbb{R}^3$.

Ejemplo 10.19 *Sea* $T : \mathbb{R}^4 \longrightarrow \mathbb{R}^2$ *la transformación lineal del Ejemplo 10.10 dada por*

$$T\left(\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \right) = \begin{bmatrix} x - w \\ w \end{bmatrix},$$

representada por la matriz estándar

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

Supongamos que queremos encontrar un conjunto LI de $\mathbb{R}^4$ *cuyo span sea igual al espacio nulo de* T.

Como $Null(T) = Null(A)$, *resolvemos el sistema homogéneo*

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

La FER de la matriz aumentada $[A \mid \vec{0}]$ *está dada por*

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array} \right],$$

de manera que $\vec{x} \in Null(A) = Null(T)$ *syss sus componentes satisfacen*

$$x_1 = 0,\ x_4 = 0,\ x_2 \in \mathbb{R},\ x_3 \in \mathbb{R}.$$

Concluimos que

$$\begin{aligned} Null(T) &= \left\{ x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \mid x_2, x_3 \in \mathbb{R} \right\} \\ &= span \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}. \end{aligned}$$

En los ejemplos anteriores se encontraron conjuntos LI que generaban a ciertos subespacios de $\mathbb{R}^2$ y de $\mathbb{R}^4$. Este tipo de conjuntos generadores es de suma importancia en la teoría de espacios vectoriales y será el motivo de estudio del siguiente capítulo.

Capítulo 11

Bases y dimensión

Introducción

Dado cualquier subespacio $S \subset \mathbb{R}^n$, decimos que el conjunto $\{\vec{v}_1, \dots \vec{v}_p\}$ genera a S [1] si

$$S = span\{\vec{v}_1, \dots, \vec{v}_p\}.$$

Ahora bien, si $\{\vec{v}_1, \dots \vec{v}_p\}$ fuese LD, la intuición nos dice que deberían poder extraerse algunos vectores de manera que el conjunto resultante, $\{\vec{v}_{i_1}, \dots, \vec{v}_{i_m}\}$, fuese LI y

$$S = span\{\vec{v}_{i_1}, \dots, \vec{v}_{i_m}\}, \quad m < p.$$

En otras palabras, se eliminan los vectores "redundantes" para generar a S más eficientemente. Estos conjuntos linealmente independientes de generadores son de suma importancia en el estudio de los espacios vectoriales. Anteriormente ya hemos introducido conjuntos como éstos. Concretamente, el conjunto

$$\{\vec{e}_1, \dots, \vec{e}_n\},$$

al que denominamos la base estándar de $\mathbb{R}^n$, es un conjunto LI que genera a $\mathbb{R}^n$, es decir,

$$span\{\vec{e}_1, \dots, \vec{e}_n\} = \mathbb{R}^n.$$

Este tipo de conjuntos serán el objeto de estudio de este capítulo.

[1] Posteriormente, veremos que todos los subespacios son de esta forma.

Bases

La siguiente definición generaliza las características de las base estándar.

Definición 11.1 *Sea S un subespacio de $\mathbb{R}^n$ y sean $\vec{v}_1, \ldots, \vec{v}_p$ vectores de S. Decimos que el conjunto $\{v_1, \ldots, v_p\}$ es una **base** para S syss se cumplen las siguientes dos condiciones.*

1. *$span\{\vec{v}_1, \ldots, \vec{v}_p\} = S$.*
2. *$\{\vec{v}_1, \ldots, \vec{v}_p\}$ es LI.*

Sabemos (por el Corolario 4.31) que si un conjunto de vectores de $\mathbb{R}^n$ es LI y tiene exactamente n elementos, entonces ese conjunto genera a todo $\mathbb{R}^n$. Por lo tanto, el siguiente resultado se sigue en forma inmediata.

Proposición 11.2 *Sean $\vec{v}_1, \ldots \vec{v}_n$ vectores de $\mathbb{R}^n$. Si $\{\vec{v}_1, \ldots, \vec{v}_n\}$ es LI, entonces $\{\vec{v}_1, \ldots, \vec{v}_n\}$ es una base para $\mathbb{R}^n$. En particular, el conjunto $\{\vec{e}_1, \ldots, \vec{e}_n\} \subset \mathbb{R}^n$, conocido como la base estándar de $\mathbb{R}^n$, es efectivamente una base para $\mathbb{R}^n$ en el sentido de la Definición 11.1.*

Ejemplo 11.3 *Los conjuntos*

$$\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \text{ y } \mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 9 \end{bmatrix}, \begin{bmatrix} \pi \\ -2 \end{bmatrix} \right\}$$

son bases para $\mathbb{R}^2$. Los conjuntos

$$\mathcal{C}_1 = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\} \text{ y } \mathcal{C}_2 = \left\{ \begin{bmatrix} 5 \\ 0 \end{bmatrix} \right\}$$

no son bases para $\mathbb{R}^2$ ya que $\mathcal{C}_1$ es LD y

$$span\ \mathcal{C}_2 = \left\{ \begin{bmatrix} a \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\} \subsetneq \mathbb{R}^2.$$

Sin embargo, $\mathcal{C}_2$ es una base para el subespacio

$$\left\{ \begin{bmatrix} a \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\}.$$

Ejemplo 11.4 *Los conjuntos*

$$\mathcal{B}_3 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \quad y \quad \mathcal{B}_4 = \left\{ \begin{bmatrix} 1 \\ 9 \\ 0 \end{bmatrix}, \begin{bmatrix} \pi \\ -2 \\ 0 \end{bmatrix} \right\}$$

son bases para el subespacio

$$\left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \mid a, b \in \mathbb{R} \right\}.$$

Por supuesto que si S es un subespacio, $\{\vec{v}_1, \ldots, \vec{v}_p\}$ es una base para S y $\vec{v} \in S$, entonces $\vec{v} \in S = span\{\vec{v}_1, \ldots, \vec{v}_p\}$ y por lo tanto, $\vec{v}$ puede escribirse como combinación lineal de $\vec{v}_1, \ldots, \vec{v}_p$. El siguiente resultado dice que, salvo por el orden de los términos, existe una **única** manera de expresar a este vector $\vec{v}$ como combinación lineal de $\vec{v}_1, \ldots, \vec{v}_p$.

Proposición 11.5 *Sean S un subespacio de $\mathbb{R}^n$, $\{\vec{v}_1, \ldots, \vec{v}_p\}$ una base para S y $\vec{v} \in S$. Si $c_1, \ldots, c_p$ y $d_1, \ldots, d_p$ son escalares tales que*

$$\vec{v} = c_1\vec{v}_1 + \cdots + c_p\vec{v}_p \quad y \quad \vec{v} = d_1\vec{v}_1 + \cdots + d_p\vec{v}_p,$$

entonces $c_j = d_j$, para toda j.

Demostración. Por hipótesis,

$$\begin{aligned} \vec{0} &= \vec{v} - \vec{v} = c_1\vec{v}_1 + \cdots + c_p\vec{v}_p - (d_1\vec{v}_1 + \cdots + d_p\vec{v}_p) \\ &= (c_1 - d_1)\vec{v}_1 + \cdots + (c_p - d_p)\vec{v}_p \end{aligned}$$

Ahora bien, como $\{\vec{v}_1, \ldots, \vec{v}_p\}$ es LI,

$$c_1 - d_1 = \cdots = c_p - d_p = 0.$$

Por lo tanto, $c_1 = d_1, \ldots, c_p = d_p$. ■

Observación 11.6 *Sean $S \subset \mathbb{R}^n$ un subespacio y $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_p\}$ una base (ordenada) de S. Si $\vec{v} \in S$, entonces existen escalares $c_1, \ldots, c_p$, determinados en forma única por la base $\mathcal{B}$, tales que*

$$\vec{v} = c_1\vec{v}_1 + \cdots + c_p\vec{v}_p.$$

Los escalares $c_1, \ldots, c_p$ se conocen como las coordenadas de $\vec{v}$ con respecto a la base $\mathcal{B}$.

Ejemplo 11.7 *Sea*

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

y supongamos que deseamos encontrar una base para

$$Null(A) = \{\vec{x} \mid A\vec{x} = \vec{0}\}.$$

Para ello, basta notar que la FER de la matriz aumentada $\left[A \mid \vec{0}\right]$ *es igual a la matriz*

$$\left[\begin{array}{cccr|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Por lo tanto, si $\vec{x} \in Null(A)$*, entonces sus componentes satisfacen*

$$x_1 = -2x_2 - x_4,\ x_3 = x_4,\ x_2 \in \mathbb{R},\ x_4 \in \mathbb{R}$$

y concluimos que

$$Null(A) = \left\{ \begin{bmatrix} -2x_2 - x_4 \\ x_2 \\ x_4 \\ x_4 \end{bmatrix} \mid x_2, x_4 \in \mathbb{R} \right\}$$

$$= \left\{ x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \mid x_3, x_4 \in \mathbb{R} \right\} = span \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Además, el conjunto

$$\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\} |$$

es LI (pues es un conjunto de dos vectores de los cuales ninguno de ellos es múltiplo del otro). Por lo tanto, $\mathcal{B}$ *es una base para* $Null(A)$*.*

Dimensión

A continuación, probaremos que cualesquiera dos bases de un subespacio tienen exactamente el mismo número de elementos. Este resultado es crucial para definir el concepto de dimensión de un espacio vectorial.

Teorema 11.8 *Si* S *es un subespacio de* $\mathbb{R}^n$ *tal que* $span\{\vec{v}_1, \ldots, \vec{v}_k\} = S$ *y* $\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}$ *es un subconjunto de* S *con* $k+1$ *elementos, entonces* $\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}$ *es LD.*

Demostración. Sea $\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}$ cualquier subconjunto de S con $k+1$ elementos. Como $S = span\{\vec{v}_1, \ldots, \vec{v}_k\}$ se tiene que para toda $i = 1, 2, \ldots, k+1$,

$$\vec{w}_i \in span\{\vec{v}_1, \ldots, \vec{v}_k\}.$$

En particular, si definimos a la matriz $A \in M_{n\times k}$ como la matriz cuyas columnas son los vectores $\vec{v}_j$, es decir

$$A = [\vec{v}_1\ \vec{v}_2 \cdots \vec{v}_k]\,,$$

se tiene que existen $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_{k+1} \in \mathbb{R}^k$ tales que[2]

$$A\vec{x}_i = \vec{w}_i,\ i = 1, 2, \ldots, k+1. \qquad (\bigstar)$$

Ahora bien, definiendo las matrices B y C como

$$\begin{aligned} B &= [\vec{x}_1\ \vec{x}_2 \cdots \vec{x}_{k+1}] \in M_{k\times(k+1)}, \\ C &= [\vec{w}_1\ \vec{w}_2 \cdots \vec{w}_{k+1}] \in M_{n\times(k+1)}, \end{aligned}$$

se tiene que, por la Proposición 6.18 y las igualdades dadas en $(\bigstar)$, se cumple, para toda $j = 1, 2, \ldots, k+1$

$$(AB)^j = AB^j = A\vec{x}_j = \vec{w}_j = C^j,$$

por lo tanto

$$AB = C. \qquad (\spadesuit)$$

Observemos que las columnas de B son LD ya que $\{\vec{x}_1,\ \vec{x}_2, \cdots, \vec{x}_{k+1}\}$ es un conjunto de $k+1$ vectores en $\mathbb{R}^k$. Entonces, existe $\vec{x} \in \mathbb{R}^{k+1}$, $\vec{x} \neq \vec{0}_{k+1}$, tal que se cumple

$$B\vec{x} = \vec{0}_k.$$

En virtud de $(\spadesuit)$ se tiene que

$$C\vec{x} = AB\vec{x} = \vec{0}_n$$

y como $\vec{x} \neq \vec{0}_{k+1}$ concluimos que las columnas de la matriz C, es decir el conjunto $\{\vec{w}_1, \ldots, \vec{w}_k, \vec{w}_{k+1}\}$, son LD. $\blacksquare$

Corolario 11.9 *Si S es un subespacio de $\mathbb{R}^n$ y $\{\vec{v}_1, \ldots, \vec{v}_k\}$ y $\{\vec{w}_1, \ldots, \vec{w}_l\}$ son bases para S, entonces $k = l$. Este número, denotado por $dim(S)$, será llamado* ***la dimensión de*** *S.*

[2] Puesto que todos los vectores $\vec{w}_i$ son combinación lineal de las columnas de A, es decir, los vectores $\vec{v}_j$.

Demostración. Como $span\{\vec{v}_1, \ldots, \vec{v}_k\} = S$ y $\{\vec{w}_1, \ldots, \vec{w}_l\}$ es LI tenemos, por el Teorema 11.8, que $l \leq k$ (de otra forma $\{\vec{w}_1, \ldots, \vec{w}_l\}$ sería LD), . De igual manera, como $span\{\vec{w}_1, \ldots, \vec{w}_l\} = S$ y $\{\vec{v}_1, \ldots, \vec{v}_k\}$ es LI tenemos, por el Teorema 11.8, que $k \leq l$. Concluimos así que $k = l$. ■

Observación 11.10 *La dimensión nos dice cuántos elementos son necesarios y suficientes para generar un espacio vectorial. Esto puede interpretarse como una medida de qué tan grande es este espacio.*

Ejemplo 11.11 *Como la base estándar de $\mathbb{R}^n$ (i.e., el conjunto $\{\vec{e}_1, \ldots \vec{e}_n\}$) es una base para $\mathbb{R}^n$, es claro que $dim(\mathbb{R}^n) = n$.*

Ejemplo 11.12 *Como el conjunto*

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

es una base para el subespacio

$$S = \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \mid a, b \in \mathbb{R} \right\},$$

se tiene que $dim(S) = 2$.

Ejemplo 11.13 *Como el conjunto*

$$\left\{ \begin{bmatrix} 5 \\ 0 \end{bmatrix} \right\}$$

es una base para el subespacio

$$S = \left\{ \begin{bmatrix} a \\ 0 \end{bmatrix} \mid a \in \mathbb{R} \right\},$$

entonces $dim(S) = 1$.

A la luz de los tres ejemplos anteriores y por motivos de conveniencia, resulta natural asignarle una dimensión nula al espacio vectorial cuyo único elemento es el vector cero.

Definición 11.14 *Si $\vec{0}_n$ denota al vector cero de $\mathbb{R}^n$, entonces*

$$\dim\left(\left\{\vec{0}_n\right\}\right) = 0.$$

Observación 11.15 *Al asignarle dimensión nula al subespacio* $S = \left\{\vec{0}_n\right\}$ *estamos diciendo, literalmente, que cualquier base de* S *tiene cero elementos, por lo tanto, la única base posible es el conjunto vacío:* $\varnothing$.

Observación 11.16 *Consideremos ahora el conjunto*

$$\mathcal{D} = \left\{ \begin{bmatrix} 7 \\ 8 \\ 9 \\ 6 \end{bmatrix}, \begin{bmatrix} -9 \\ 7 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 23 \\ 22 \\ 13 \end{bmatrix} \right\}$$

y supongamos que deseamos probar que $span\mathcal{D} \subsetneq \mathbb{R}^4$. *Para ello, basta notar que* $dim(span\mathcal{D}) \leq 3$ *(de hecho,* $dim(span\mathcal{D}) = 2$ *porque el tercer elemento de* $\mathcal{D}$ *es combinación lineal de los otros dos) mientras que* $dim(\mathbb{R}^4) = 4$. *Por lo tanto,* $span\mathcal{D} \subsetneq \mathbb{R}^4$.

La siguiente proposición es una generalización de la observación anterior.

Proposición 11.17 (El famoso pecado por defecto) *Si* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ *son vectores de* $\mathbb{R}^n$ *con* $p < n$, *entonces* $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subsetneq \mathbb{R}^n$.

Demostración. Como $dim(span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}) \leq p < n = dim(\mathbb{R}^n)$, es claro que $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subsetneq \mathbb{R}^n$. ■

Si $S \subset \mathbb{R}^n$ es un subespacio, la idea de tener subconjuntos minimales de generadores (subconjuntos generadores con el menor número de elementos posible) y subconjuntos maximales LI (subcojuntos LI con el mayor número de elementos posible) son los argumentos clave de los siguientes resultados.

Proposición 11.18 *Si* $S \subset \mathbb{R}^n$ *es un subespacio,* $S \neq \{\vec{0}\}$ *y* p *es el* ***máximo*** *número natural tal que existen* $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in S$ *de forma que el conjunto* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *es LI (decimos que se trata de un conjunto* ***maximal*** *LI en* S.*); entonces* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *es una base para* S.

Demostración. La demostración se llevará a cabo por contradicción, es decir, probamos que la invalidez de la conclusión nos conduce a la invalidez de alguna de las hipótesis. Primero observamos que $1 \leq p$ (pues S tiene al menos un vector distinto del vector cero) y que $p \leq n$ (pues cualquier subconjunto LI de $\mathbb{R}^n$ tiene a lo más n elementos y porque $S \subset \mathbb{R}^n$). Ahora bien, si $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subset S$ no fuese una base para S, entonces $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subsetneq S$ y existiría un vector $\vec{v} \in S$ tal que

$$\vec{v} \notin span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}.$$

Por lo tanto, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p, \vec{v}\}$ sería un subconjunto linealmente independiende de S con $p + 1$ elementos, lo cual contradice la supuesta maximalidad de p. ■

Proposición 11.19 *Si $S \subset \mathbb{R}^n$ es un subespacio con $S \neq \{\vec{0}\}$ y p es el* ***mínimo*** *número natural tal que existen $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in S$ de forma que $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = S$ (decimos que el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto* ***minimal*** *de generadores de S); entonces $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base para S.*

Demostración. La demostración también se llevará a cabo por contradicción. Primero observemos que $1 \leq p$ (pues S tiene al menos un vector distinto del vector cero). Ahora bien, si $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subset S$ no fuese una base para S, tendría que ser LD de manera que alguno de los vectores $\vec{v}_i$ sería combinación lineal de los demás. Sin perdida de general (reordenando los vectores si es necesario) supongamos que

$$\vec{v}_p \in span\{\vec{v}_1, \ldots, \vec{v}_{p-1}\},$$

lo cual implicaría que

$$span\{\vec{v}_1, \ldots, \vec{v}_{p-1}\} = S.$$

Por lo tanto, $\{\vec{v}_1, \ldots, \vec{v}_{p-1}\}$ sería un conjunto de generadores con $p-1$ elementos, lo cual contradice la hipótesis de que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto minimal de generadores. ■

Corolario 11.20 *Si $S \subset \mathbb{R}^n$ es un subespacio con $dim(S) = p > 0$ y $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subset S$ es un conjunto linealmente independiente con p elementos, entonces $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base para S.*

Demostración. Gracias al Teorema 11.8 sabemos cualquier subconjunto de S de tamaño $p+1$ es LD (pues S es generado por cualquiera de sus bases y todas ellas son de tamaño p). De esta forma, p y $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ son como en las hipótesis de la Proposición 11.18 y por lo tanto, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base para S. ■

Observación 11.21 *En particular, los resultados anteriores nos dicen que dado $S \subset \mathbb{R}^n, S \neq \{\vec{0}\}$ un subespacio, Cualquier subconjunto minimal de generadores o cualquier subconjunto maximal LI, forman una base para S.*

Ejemplo 11.22 *Sea S el subespacio de $\mathbb{R}^4$ que consiste de todos aquellos vectores cuya cuarta componente es igual a cero y supongamos que deseamos mostrar que el conjunto $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ es una base para S, donde*

$$\vec{w}_1 = \begin{bmatrix} 1 \\ \pi \\ 7 \\ 0 \end{bmatrix}, \vec{w}_2 = \begin{bmatrix} -1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \vec{w}_3 \begin{bmatrix} 0 \\ 3 \\ 0 \\ 0 \end{bmatrix}.$$

Para ello, y en vista del corolario anterior, basta notar que $dim(S) = 3$ *(pues* $\{\vec{e}_1, \vec{e}_2, \vec{e}_3)$ *es una base para* S*), que* w_1*,* w_2 *y* w_3 *son elementos de* S *y que el conjunto* $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ *es un conjunto LI.*

Ahora presentamos la forma dual del corolario anterior.

Proposición 11.23 *Si* $S \subset \mathbb{R}^n$ *es un subespacio con* $dim(S) = p > 0$ *y* $span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = S$*, entonces* $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ *es una base para* S*.*

Demostración. Una vez más procedemos por contradicción. Si $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ no es una base para S, entonces $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto LD. Sin perdida de generalidad, podemos suponer que $\vec{v}_1 \in span\{\vec{v}_2, \ldots, \vec{v}_p\}$ con lo cual se tiene que,

$$S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = span\{\vec{v}_2, \ldots, \vec{v}_p\}.$$

Por lo tanto,

$$p = dim(S) = dim(span\{\vec{v}_2, \ldots, \vec{v}_p\}) \leq p - 1,$$

lo cual es una contradicción. ■

Minería (extracción) de bases

En el caso de la Observación 11.16, el lector minucioso seguramente habrá observado que, a pesar de que $\mathcal{D}$ tiene 3 elementos y $dim(span\mathcal{D}) = 2$, es posible encontrar una base para $span\mathcal{D}$ cuyos vectores son elementos de $\mathcal{D}$. De hecho, gracias a la advertencia de que "el tercer elemento de $\mathcal{D}$ es combinación lineal de los otros dos", es claro que una base para dicho subespacio es el conjunto

$$\left\{ \begin{bmatrix} 7 \\ 8 \\ 9 \\ 6 \end{bmatrix}, \begin{bmatrix} -9 \\ 7 \\ 4 \\ 1 \end{bmatrix} \right\}.$$

De forma más general, supongamos que $S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$. Si $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es LI, por definición se tiene que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base para S. En caso contrario, alguno de los elementos de este conjunto es combinación lineal de los otros $p-1$ elementos. Sin perdida de generalidad, podemos suponer que $v_1 \in span\{\vec{v}_2, \ldots, \vec{v}_p\}$ (de no ser así, simplemente se reordenan los elementos). Entonces, $S = span\{\vec{v}_2, \ldots, \vec{v}_p\}$ y tenemos dos posibilidades:

1. $\{\vec{v}_2, \ldots, \vec{v}_p\}$ es LI y por lo tanto, este conjunto con $p-1$ elementos constituye una base para S, o bien
2. $\{\vec{v}_2, \ldots, \vec{v}_p\}$ es LD y por lo tanto, es posible seguir extrayendo un número finito elementos de este conjunto hasta obtener un subconjunto LI que genere a todo S (i.e., una base para S).

El siguiente teorema es una consecuencia inmediata del razonamiento anterior.

Teorema 11.24 *Si $S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$, entonces existe un subconjunto de $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ que es base para S y por lo tanto, $dim(S) \leq p$.*

El teorema anterior nos dice que si $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto de generadores de un subespacio S, es decir, $S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$, entonces podemos extraer un subconjunto LI que siga generando S. Sin embargo, es necesario tener algún método para encontrar estas bases. Para este propósito se requiere de los siguientes resultados preliminares.

Proposición 11.25 *Sea $A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \in M_{m\times n}$ y sea $\tilde{A} = \begin{bmatrix} \tilde{A}_1 \\ \tilde{A}_2 \\ \vdots \\ \tilde{A}_m \end{bmatrix}$ su FER. Entonces $span\{A_1, A_2, \ldots, A_m\} = span\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m\}$, es decir, los renglones de ambas matrices generan el mismo subespacio en $\mathbb{R}^n$.*

Demostración. Todos los renglones de $\tilde{A}$ se obtienen de los renglones de A mediante operaciones elementales sobre los mismos. Concretamente, dado cualquier renglón $\tilde{A}_i$ de $\tilde{A}$, necesariamente este renglón será una combinación lineal de renglones de A y por lo tanto,

$$span\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m\} \subset span\{A_1, A_2, \ldots, A_m\}.$$

Ahora bien, cada paso utilizado para obtener la FER de A es reversible, simplemente utilizando la operación elemental inversa (que claramente es el mismo tipo de operación elemental). De aquí que cualquier renglón A_i de A será una combinación lineal de renglones de $\tilde{A}$, de manera que

$$span\{A_1, A_2, \ldots, A_m\} \subset span\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m\}$$

y se tiene la igualdad

$$span\{A_1, A_2, \ldots, A_m\} = span\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_m\}.$$

■

Observación 11.26 *El conjunto de los renglones de la matriz $\tilde{A}$ puede dividirse en dos subconjuntos: el subconjunto de aquellos renglones cuyas entradas son todas cero y el subconjunto de aquellos renglones que tienen un uno principal. Sin embargo, es claro que el segundo subconjunto es LI y genera a todos los elementos del primero. Por ello, es más preciso decir que el conjunto generado por los renglones de A es igual al conjunto generado por los renglones de $\tilde{A}$ que tienen un uno principal y que tal conjunto es LI. En resumen, el conjunto generado por los renglones de A tiene como base al conjunto formado por los renglones de $\tilde{A}$ que tienen un uno principal. El siguiente corolario es inmediato a partir de esta observación.*

Corolario 11.27 *Sea* $A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \in M_{m\times n}$ *y sea* $\tilde{A} = \begin{bmatrix} \tilde{A}_1 \\ \tilde{A}_2 \\ \vdots \\ \tilde{A}_m \end{bmatrix}$ *su FER. Entonces, si $\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_l$ con $l \leq m$, son los renglones de $\tilde{A}$ que tienen unos principales, se tiene que $\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_l\}$ es LI y*

$$span\{A_1, A_2, \ldots, A_m\} = span\{\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_l\}.$$

Ejemplo 11.28 *Sea*

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 5 \\ -4 & 4 & -9 \end{bmatrix},$$

de manera que su FER se obtiene (hacemos dos operaciones por paso en aras de ser eficientes) como

$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 5 \\ -4 & 4 & -9 \end{bmatrix} \underset{\sim}{\substack{-2R_1+R_2 \\ 4R_1+R_3}} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \underset{\sim}{\substack{R_2+R_3 \\ -2R_2+R_1}} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \tilde{A},$$

Dado que

$$\begin{aligned} \tilde{A}_1 &= A_1 - 2\,(-2A_1 + A_2) = 5A_1 - 2A_2, \\ \tilde{A}_2 &= -2A_1 + A_2, \\ \tilde{A}_3 &= (-2A_1 + A_2) + (4A_1 + A_3) = 2A_1 + A_2 + A_3, \end{aligned}$$

entonces es claro que $span\{\tilde{A}_1, \tilde{A}_2, \tilde{A}_3\} \subset span\{A_1, A_2, A_3\}$. Por otro lado, utilizando las operaciones elementales inversas se tiene,

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \underset{\sim}{\substack{-R_2+R_3 \\ 2R_2+R_1}} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \underset{\sim}{\substack{2R_1+R_2 \\ -4R_1+R_3}} \begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 5 \\ -4 & 4 & -9 \end{bmatrix} = A,$$

por lo que, en forma análoga, se obtiene que

$$span\{A_1, A_2, A_3\} \subset span\{\tilde{A}_1, \tilde{A}_2, \tilde{A}_3\}.$$

Por lo tanto,

$$span\{A_1, A_2, A_3\} = span\{\tilde{A}_1, \tilde{A}_2, \tilde{A}_3\}.$$

Más aún, los renglones de $\tilde{A}$ que tienen unos principales son $\tilde{A}_1$ y $\tilde{A}_2$ de manera que $\{\tilde{A}_1, \tilde{A}_2\}$ es LI y

$$span\{A_1, A_2, A_3\} = span\{\tilde{A}_1, \tilde{A}_2\}.$$

A continuación presentaremos dos algoritmos muy efectivos para encontrar bases de subespacios S de la forma $S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$. El primero de ellos se limita a encontrar una base (aunque la base no tenga ninguna relación con el conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$) mientras que el segundo, efectivamente extrae una base del conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ en el sentido del Teorema 11.24.

ALGORITMO 1. El primer algoritmo está basado en el Corolario 11.27. Así, si S es un subespacio de $\mathbb{R}^n$ de la forma

$$S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$$

y queremos encontrar una base para S, consideramos primero a la matriz $A \in M_{m\times n}$ tal que $A_1^T = \vec{v}_1, A_2^T = \vec{v}_2, \ldots, A_m^T = \vec{v}_m$. Después, encontramos la FER de A; supongamos que $\tilde{A}$ es dicha FER. Finalmente, notamos que

$$S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\} = span\{\tilde{A}_1^T, \tilde{A}_2^T, \ldots \tilde{A}_l^T\}$$

en donde $\{\tilde{A}_1, \tilde{A}_2, \ldots \tilde{A}_l\}$ es el conjunto formado por los renglones de $\tilde{A}$ que tienen un uno principal. Dado que este conjunto es LI, concluimos que $\{\tilde{A}_1^T, \tilde{A}_2^T, \ldots \tilde{A}_l^T\}$ es una base para S.

Ejemplo 11.29 *Sea*

$$S = Span\left\{\begin{bmatrix}-1\\-1\\0\end{bmatrix}, \begin{bmatrix}-1\\-1\\5\end{bmatrix}, \begin{bmatrix}-1\\-1\\-10\end{bmatrix}\right\}$$

y supongamos que deseamos encontrar una base para S. Para utilizar el Algoritmo 1, basta notar que la FER de la matriz

$$\begin{bmatrix}-1 & -1 & 0\\-1 & -1 & 5\\-1 & -1 & -10\end{bmatrix},$$

cuyos renglones son estos vectores, es la matriz

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

por lo que el conjunto (de renglones con unos principales)

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

es una base para S.

El algoritmo que acabamos de utilizar no es útil cuando queremos obtener una base a partir de un conjunto de generadores de un subespacio, misma que esté constituida por elementos de este conjunto. Para este efecto es necesario analizar la estructura de dependencia lineal entre los vectores columna de una matriz y su FER.

Recordemos que si $A = \begin{bmatrix} A^1 & A^2 & \cdots & A^n \end{bmatrix} \in M_{m\times n}$ y $\tilde{A} = \begin{bmatrix} \tilde{A}^1 & \tilde{A}^2 & \cdots & \tilde{A}^n \end{bmatrix}$ es su FER, entonces $Null(A) = Null(\tilde{A})$, es decir

$$A\vec{x} = \vec{0} \;\text{ syss }\; \tilde{A}\vec{x} = \vec{0}.$$

Equivalentemente, tomando $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$,

$$x_1A^1+x_2A^2+\cdots+x_nA^n = \vec{0} \quad\text{syss}\quad x_1\tilde{A}^1+x_2\tilde{A}^2+\cdots+x_n\tilde{A}^n = \vec{0}. \quad (\bigstar)$$

La equivalencia dada por ($\bigstar$) nos dice que la relación de dependencia lineal entre los vectores columna de A y de $\tilde{A}$ es exactamente la misma. En virtud de esto, un subconjunto de columnas de A es LI syss el subconjunto de columnas correspondientes de $\tilde{A}$ es LI.

Ejemplo 11.30 *Sea*

$$A = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 6 & 2 \\ 1 & 2 & 4 \end{bmatrix}$$

cuya FER está dada por

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

Por lo tanto,

$$Null(A) = Null(\tilde{A}) = \left\{ \begin{bmatrix} -2x_3 \\ -x_3 \\ x_3 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\} \subset \mathbb{R}^3.$$

Si tomamos

$$\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} \in Null(A) = Null(\tilde{A})$$

se cumple,

$$2\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ 6 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

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

de manera que

$$2\underbrace{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}}_{A^1} + \underbrace{\begin{bmatrix} -2 \\ 6 \\ 2 \end{bmatrix}}_{A^2} = \underbrace{\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}}_{A^3},$$

$$2\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\tilde{A}^1} + \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\tilde{A}^2} = \underbrace{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}}_{\tilde{A}^3}.$$

Asimismo, se observa que los conjuntos $\{A^1, A^2\}$ *y* $\left\{\tilde{A}^1, \tilde{A}^2\right\}$ *son LI.*

Supongamos que $A \in M_{m\times n}$ y sea $\tilde{A} \in M_{m\times n}$ su FER conteniendo k unos principales, $k \leq n$. Dado el sistema homogéneo $\tilde{A}\vec{x} = \vec{0}$, existen $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$, variables restringidas que corresponden a las columnas $\tilde{A}^{i_1}, \ldots, \tilde{A}^{i_k}$ que contienen los unos principales. Por construcción, cada una de estas columnas es de la forma

$$\tilde{A}^{i_j} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; j = 1, 2, \ldots k, \qquad (\spadesuit)$$

en donde el "1" se encuentra en la posición i_j. Asimismo, existen

$$x_{i_{(k+1)}}, x_{i_{(k+2)}}, \ldots, x_{i_n},$$

variables libres, que pueden tomar cualquier valor, mismas que corresponden a las columnas $\tilde{A}^{i_{(k+1)}}, \ldots, \tilde{A}^{i_n}$. Sea

$$\mathcal{L} = \{\tilde{A}^{i_{(k+1)}}, \ldots, \tilde{A}^{i_n}\};$$

dada $\tilde{A}^{i_j} \in \mathcal{L}$, se asigna el valor $x_{i_j} = -1$ a la variable libre correspondiente y cero a todas las variables libres restantes. Entonces, dado que

$$x_1\tilde{A}^1 + x_2\tilde{A}^2 + \cdots + x_n\tilde{A}^n = \vec{0},$$

$\tilde{A}^{i_j}$ puede despejarse como

$$\tilde{A}^{i_j} = x_{i_1}\tilde{A}^{i_1} + x_{i_2}\tilde{A}^{i_2} \ldots + x_{i_k}\tilde{A}^{i_k}$$

y por lo tanto,

$$\tilde{A}^{i_j} \in span\left\{\tilde{A}^{i_1}, \ldots, \tilde{A}^{i_k}\right\}.$$

Adicionalmente, como el conjunto $\left\{\tilde{A}^{i_1}, \ldots, \tilde{A}^{i_k}\right\}$ de las columnas correspondientes a las variables restringidas es LI (esto es claro pues son de la forma (♠)), el conjunto de columnas $\{A^{i_1}, A^{i_2} \ldots, A^{i_k}\}$ también será LI y constituirá una base para el conjunto de columnas de A. La siguiente proposición resume la discusión anterior.

Proposición 11.31 *Sean* $A = \begin{bmatrix} A^1 & A^2 & \cdots & A^n \end{bmatrix} \in M_{m\times n}$ *y* $\tilde{A} = \begin{bmatrix} \tilde{A}^1 & \tilde{A}^2 & \cdots & \tilde{A}^n \end{bmatrix}$ *su FER. Entonces si* $\tilde{A}^{i_1}, \tilde{A}^{i_2}, \ldots, \tilde{A}^{i_k}$ *son las columnas de* $\tilde{A}$ *que poseen unos principales, éstas forman un conjunto LI. Asimismo, el conjunto correspondiente* $\{A^{i_1}, A^{i_2}, \ldots, A^{i_k}\}$ *de columnas de* A *también es LI. Adicionalmente,*

$$span\left\{\tilde{A}^{i_1}, \tilde{A}^{i_2}, \ldots, \tilde{A}^{i_k}\right\} = span\left\{\tilde{A}^1, \tilde{A}^2, \ldots, \tilde{A}^n\right\}$$

y

$$span\left\{A^{i_1}, A^{i_2}, \ldots, A^{i_k}\right\} = span\left\{A^1, A^2, \ldots, A^n\right\}.$$

Este resultado nos lleva al siguiente algoritmo.

ALGORITMO 2. Sea S un subespacio de $\mathbb{R}^n$ de la forma $S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. Supongamos que queremos encontrar una base para S con la restricción adicional de que ésta debe ser un subconjunto del conjunto original $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$. La filosofía del algoritmo que resuelve este problema se basa en la Proposición 11.31. Sea A la matriz cuyas columnas son $A^1 = \vec{v}_1, A^2 = \vec{v}_2, \ldots, A^n = \vec{v}_n$ y sea $\tilde{A}$ la FER de A. Notamos que la columna A^j (i.e., el vector $\vec{v}_j$) estará en nuestra base $\mathcal{B}$

syss la columna $\tilde{A}^j$ tiene uno uno principal. Así, si $\left\{\tilde{A}^{i_1}, \tilde{A}^{i_2}, \ldots, \tilde{A}^{i_k}\right\}$ es el conjunto de columnas de $\tilde{A}$ con unos principales, por la Proposición 11.31 tenemos que el conjunto correspondiente de columnas de A dado por$\{A^{i_1}, A^{i_2}, \ldots, A^{i_k}\}$, es una base de S.

Ejemplo 11.32 *Sean*

$$\vec{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -1 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \vec{v}_4 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \vec{v}_5 = \begin{bmatrix} 0 \\ 4 \\ 0 \\ 4 \end{bmatrix}$$

vectores en $\mathbb{R}^4$ *y supongamos que deseamos encontrar una base para*

$$span\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\},$$

con la restricción de que dicha base debe de ser un subconjunto de

$$\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}.$$

Para hacer uso del Algoritmo 2 calculamos la FER de la matriz cuyas columnas son los vectores $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5$. *Sea*

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

dicha matriz, cuya FER está dada por

$$\tilde{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 4 \\ 0 & 1 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Esta matriz tiene unos principales en la primera, segunda y cuarta columnas, por lo que el Algoritmo 2 asegura que el conjunto

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix} \right\}$$

es una base para S. *Si el lector quisiera una verificación de que* $\mathcal{B}$ *es LI, sería suficiente hacerle notar que la FER de la matriz aumentada*

$$[A^1 \ A^2 \ A^4 \mid \ \vec{0}]$$

es la matriz

$$[\tilde{A}^1\ \tilde{A}^2\ \tilde{A}^4 \mid\ \vec{0}],$$

De igual manera, si el lector quisiera verificar, por ejemplo, que $\vec{v}_5 \in span\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$*, sería suficiente hacerle notar que la FER de la matriz aumentada* $[A^1\ A^2\ A^4 \mid\ A^5]$ *es la matriz*

$$[\tilde{A}^1\ \tilde{A}^2\ \tilde{A}^4 \mid\ \tilde{A}^5],$$

la cual representa un sistema consistente. Las verificaciones de que $\vec{v}_2 \in span\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ *y de que* $\vec{v}_4 \in span\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ *son análogas.*

Ejemplo 11.33 *Considérese el subconjunto de* $\mathbb{R}^3$ *dado por*

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Es claro que $\mathcal{A}$ *es LD y span*$\mathcal{A} = \mathbb{R}^3$*. Para extraer una base para* $\mathbb{R}^3$ *consideramos a la matriz cuyas columnas son los elementos de* $\mathcal{A}$*, es decir*

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

cuya FER está dada por

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

Los unos principales están en las columnas uno dos y tres y, por lo tanto, por el Algoritmo 2

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathbb{R}^3$ *formada por elementos de* $\mathcal{A}$*.*

Observemos que si reordenamos los elementos de $\mathcal{A}$ *como*

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\},$$

la matriz asociada será ahora

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

cuya FER se obtiene fácilmente como

$$\begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

En consecuencia

$$\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

es otro subconjunto LI de $\mathcal{A}$ tal que span$\mathcal{C}$ $=\mathbb{R}^3$. En general, el conjunto inicial $\mathcal{A}$ puede reordenarse de cualquier manera para incluir a los primeros vectores (LI) en una base de $\mathbb{R}^3$.

Observación 11.34 *En el ejemplo anterior vemos que pueden extraerse múltiples bases de un conjunto de vectores dado. Asimismo, si queremos garantizar que uno o varios vectores LI pertenezcan a una base, entonces reordenamos el conjunto de manera que estos vectores sean los primeros vectores del mismo.*

Existencia de bases

Los Algoritmos 1 y 2 nos permiten encontrar una base a partir de un conjunto de generadores de un subespacio $S \subset \mathbb{R}^n$. Esencialmente, dado un conjunto de generadores, se eliminan vectores hasta llegar a un conjunto minimal de generadores que constituye una base para S. Ahora bien, que pasaría si en lugar de comenzar con un conjunto de vectores que genera a S iniciamos simplemente con un subespacio S y queremos encontrarle una base. En este caso, se construye una base, agregando uno a uno sus elementos hasta obtener un conjunto maximal LI ¿Cómo procederíamos?

Sea $S \subset \mathbb{R}^n$ cualquier subespacio. Si $S = \{\vec{0}\}$, entonces el conjunto vacío, $\emptyset$, es la base buscada. Si $S \neq \{\vec{0}\}$, existe $\vec{v}_1 \in S$, $\vec{v}_1 \neq \vec{0}$, entonces hay dos posibilidades:

$$span\{\vec{v}_1\} = S \text{ o } span\{\vec{v}_1\} \subsetneq S.$$

En el primer caso $\{\vec{v}_1\}$ es una base para S y el proceso termina. En el segundo, existirá $\vec{v}_2 \in S, \vec{v}_2 \notin span\{\vec{v}_1\}$ de manera que $\{\vec{v}_1, \vec{v}_2\}$ es LI. Una vez más hay dos posibilidades

$$span\{\vec{v}_1, \vec{v}_2\} = S \text{ o } span\{\vec{v}_1, \vec{v}_2\} \subsetneq S.$$

Al igual que antes, en el primer caso $\{\vec{v}_1, \vec{v}_2\}$ es una base para S y en el segundo, existirá $\vec{v}_3 \in S, \vec{v}_2 \notin span\{\vec{v}_1, \vec{v}_2\}$ de manera que $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ es LI. Continuamos agregando vectores hasta que

$$span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = S.$$

El proceso termina, necesariamente, ya que $S \subset \mathbb{R}^n$ y, por lo tanto, el número de vectores que se agregan debe ser tal que, $p \leq n$. El conjunto $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base para S y es maximal LI. El siguiente teorema y su corolario son inmediatos.

Teorema 11.35 *Sea $S \subset \mathbb{R}^n$ cualquier subespacio diferente del $\{\vec{0}\}$, entonces existen $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in S$, tales que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es una base de S.*

Corolario 11.36 *Sea $S \subset \mathbb{R}^n$ cualquier subespacio diferente del $\{\vec{0}\}$ y sea $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} \subset S$ cualquier subconjunto LI. Entonces existen vectores $\vec{v}_{p+1}, \vec{v}_{p+2}, \ldots, \vec{v}_r \in S$ tales que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ es una base para S. En otras palabras, cualquier subconjunto LI de S puede completarse a una base.*

Observación 11.37 *Si $S \subset \mathbb{R}^n$ es cualquier subespacio, entonces $S = \{\vec{0}\}$ o existen $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in S$ tales que*

$$S = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}.$$

Ejemplo 11.38 *Supongamos que se desea encontrar una base para $\mathbb{R}^3$ a partir del conjunto LI dado por*

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

Tenemos que

$$\begin{aligned} & span\left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \\ &= \left\{ r \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \middle| r, s \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} r \\ -r+s \\ 2r \end{bmatrix} \middle| r, s \in \mathbb{R} \right\}. \end{aligned}$$

Observando que la tercera coordenada de cualquier elemento de este span es dos veces su primera coordenada se tiene que (como $0 \neq 2 \times 1$)

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \notin span \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

y el conjunto

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

es LI independiente. En este caso, el proceso termina aquí pues este conjunto LI tiene tres elementos y forma una base para $\mathbb{R}^3$.

Dados $S_1, S_2 \subset \mathbb{R}^n$ subespacios, el siguiente resultado es intuitivamente claro.

Proposición 11.39 *Si S_1 y S_2 son subespacios de $\mathbb{R}^n$ tales que $S_1 \subset S_2$, entonces $dim(S_1) \leq dim(S_2)$. En caso de que los subespacios S_1 y S_2 sean tales que $S_1 \subset S_2$ y $dim(S_1) = dim(S_2)$, entonces $S_1 = S_2$.*

Demostración. Sea $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ una base para S_1 y sea $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_q\}$ una base para S_2, mismas que sabemos que existen por el Teorema 11.35. Como $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto LI cuyos elementos están en S_1 (y por tanto, en S_2) y como $S_2 = span\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_q\}$, el Teorema 11.8 asegura que $p \leq q$, es decir, $dim(S_1) \leq dim(S_2)$.

Supongamos ahora que $S_1 \subset S_2$ y que $dim(S_1) = dim(S_2)$. Una vez más tenemos que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ es un conjunto LI cuyos elementos están en S_1 (y por tanto, en S_2). Como el tamaño de este conjunto es igual a la dimensión de S_2, el Corolario 11.20 asegura que $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ también es base de S_2. Por lo tanto,

$$S_1 = span\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = S_2.$$

■

Unicidad de la FER(opcional)

Recordemos que en el Capítulo 2 se demostró que, mediante operaciones elementales sobre sus renglones, toda matriz puede transformarse en una matriz que está en forma escalonada reducida. Una pregunta natural es si esta FER es única y la respuesta es afirmativa. A continuación se demuestra la unicidad de la FER.

Proposición 11.40 *Sea* $A \in M_{m\times n}$*, entonces si* U *y* V *son formas escalonadas reducidas de* A*, se tiene que* $U = V$*.*

Demostración. Si $A = \Theta$, el resultado es inmediato pues A es su propia FER. Si $A \neq \Theta$, dado que U y V son FER's de A, se tiene que existen matrices elementales $E(1), E(2), \ldots, E(k)$ y $F1), F(2), \ldots, F(l)$ tales que se cumplen

$$E(k)E(k-1)\cdots E(1)A = U,$$
$$F(k)F(k-1)\cdots F(l)A = V.$$

Por lo tanto, recordando que todas las matrices elementales son invertibles, existe una matriz invertible

$$C = E(k)\cdots E(1)F(1)^{-1}\cdots F(l)^{-1} \in M_{m\times m}$$

tal que

$$U = CV,$$
$$V = C^{-1}U.$$

De aquí se obtiene que para toda $j = 1, 2, \ldots, n$,

$$U^j = (CV)^j = CV^j;$$

de manera que $U^j = \vec{0}$ syss $V^j = \vec{0}$ y las matrices U y V tienen las mismas columnas de ceros. Por lo tanto, sin pérdida de generalidad consideramos las matrices $\hat{U}$ y $\hat{V}$ que se obtienen a partir de U y V eliminando todas las columnas de ceros. Asimismo, definimos a la matriz $\hat{A}$ como aquella que se obtiene a partir de A eliminando las mismas columnas que se eliminaron en U y V para obtener $\hat{U}$ y $\hat{V}$.

Por construccíon, $\hat{U}$ y $\hat{V}$ son FER's de $\hat{A}$ y observamos que, como no hay columnas de ceros, $\hat{U}$ y $\hat{V}$ tienen un uno principal en la posición $1-1$, es decir, $\hat{U}_{11} = \hat{V}_{11} = 1$ y $\hat{U}_{i1} = \hat{V}_{i1} = 0$ para todo renglón $i \neq 1$. Realizamos inducción sobre el número de columnas de $\hat{A}$ para probar que $\hat{U} = \hat{V}$. Si $n = 1$ el resultado es inmediato puesto que

$$\hat{U} = \hat{V} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Supongamos ahora que $n > 1$. Por el argumento que acabamos de presentar $U^1 = V^1$. Considérense las matrices $\tilde{A}, \tilde{U}$ y $\tilde{V}$ que se obtienen a partir de $\hat{A}, \hat{U}$ y $\hat{V}$, respectivamente, eliminando la primera columna. Entonces, $\tilde{A}, \tilde{U}$ y $\tilde{V}$ tienen $n-1$ columnas y, por hipótesis de inducción, $\tilde{U} = \tilde{V}$, por lo tanto, $\hat{U} = \hat{V}$ y $U = V$. ■

Capítulo 12

Subespacios asociados a una matriz

En este breve capítulo mostraremos la íntima relación que existe entre **tres espacios fundamentales asociados a una matriz**. Recordemos que si A es una matriz de $m \times n$, entonces **el espacio nulo de** A es el conjunto de soluciones a la ecuación homogénea, es decir,

$$Null(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}_m\}.$$

Definición 12.1 *Si $A \in M_{m\times n}$, entonces la* ***nulidad*** *de A, denotada por $nulidad(A)$, se define como la dimensión del espacio nulo de A.*

La siguiente proposición justifica el razonamiento utilizado en el Ejemplo 11.7 en el cual una base para $Null(A)$ tenía dos elementos, con lo cual $nulidad(A) = 2$.

Proposición 12.2 *Si $A \in M_{m\times n}$, la nulidad de A coincide con el número de variables libres de la FER de A.*

Demostración. Si $\tilde{A}$ es la FER de A entonces $A\vec{x} = \vec{0}$ syss $\tilde{A}\vec{x} = \vec{0}$, es decir, $Null(A) = Null(\tilde{A})$. Supongamos que la matriz aumentada $\left[\tilde{A} \mid \vec{0}\right]$ tiene k unos principales correspondientes a las variables restringidas $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$. Las $n-k$ variables restantes son libres y las denotamos por $x_{i_{k+1}}, x_{i_{k+2}}, \ldots, x_{i_n}$. Utilizando $\tilde{A}$, cada variable restringida puede despejarse en términos de las $n-k$ variables libres. Se tiene así que dada $\vec{x} \in Null(\tilde{A}) \subset \mathbb{R}^n$, sus coordenadas satisfacen

$$\begin{aligned} x_{i_j} &= \sum_{l=1}^{n-k} a_{jl} x_{i_{k+l}} \text{ si } j = 1, \ldots, k, \\ x_{i_j} &\in \mathbb{R} \ \text{ si } j = k+1, \ldots, n, \end{aligned}$$

para escalares a_{jl} ($j = 1, \ldots, k$ y $l = 1, \ldots, n-k$) determinados por la matriz $\tilde{A}$.

Sean $\vec{v}_{i_{k+1}}, \vec{v}_{i_{k+2}}, \ldots, \vec{v}_{i_n} \in Null(\tilde{A})$ los vectores inducidos por las variables libres en el siguiente sentido: $\vec{v}_{i_{k+l}}$ se obtiene tomando los valores de las variables libres dados por

$$x_{i_{k+l}} = 1 \text{ y } x_{i_{k+h}} = 0 \text{ si } h \neq l.$$

Así, las coordenadas i_j, $j = 1, \ldots, n$, del vector $\vec{v}_{i_{k+l}}$ quedan dadas como

$$\vec{v}_{i_{k+l}} \cdot \vec{e}_{i_j} = \begin{cases} a_{jl} & \text{si } j = 1, \ldots, k, \\ 1 & \text{si } j = k+l, \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

En esta expresión, $\{\vec{e}_{i_1}, \vec{e}_{i_2}, \ldots, \vec{e}_{i_n}\}$ es el conjunto de vectores de la base canónica de $\mathbb{R}^n$ y apelamos a la Observación 4.12 para simplificar la notación de la "coordenada i_j del vector $\vec{v}_{i_{k+l}}$". Observemos que si $\vec{x} \in Null(\tilde{A})$, entonces puede expresarse como

$$\vec{x} = \sum_{l=1}^{n-k} x_{i_{k+l}} \vec{v}_{i_{k+l}},$$

es decir, $\vec{x} \in span\left\{\vec{v}_{i_{k+1}}, \vec{v}_{i_{k+2}}, \ldots, \vec{v}_{i_n}\right\}$ y se cumple

$$Null(\tilde{A}) = span\left\{\vec{v}_{i_{k+1}}, \vec{v}_{i_{k+2}}, \ldots, \vec{v}_{i_n}\right\} = Null(A).$$

Adicionalmente, el conjunto $\vec{v}_{i_{k+1}}, \vec{v}_{i_{k+2}}, \ldots, \vec{v}_{i_n}$ es LI puesto que $\vec{v}_{i_{k+l}}$ tiene un uno en la posición i_{k+l} y los otros vectores tienen un cero en dicha posición. Concluimos que la dimensión del espacio nulo está dada por

$$nulidad(A) = \# \text{ de variables libres} = n - k.$$

■

Ahora veremos que el número de variables restringidas de A es igual a la dimensión del **espacio generado por las columnas de** A y, sorprendéntemente, también coincide con la dimensión del **espacio generado por los renglones de** A. Observemos que estos dos subespacios no son necesariamente iguales incluso si $n = m$ (evidentemente si $n \neq m$, ambos subespacios son distintos pues "habitan" distintos espacios).

Definición 12.3 *Sea $A \in M_{m\times n}$. Definimos el espacio columna de A, denotado por $col(A)$, como el conjunto*

$$\begin{aligned} col(A) &= span\{A^1, \ldots, A^n\} \\ &= \{\vec{b} \in \mathbb{R}^m \mid \textit{el sistema } A\vec{x} = \vec{b} \textit{ es consistente}\} \end{aligned}$$

y a su dimensión como

$$rank(A) = \dim(\mathrm{col}(A)).$$

Observación 12.4 *De acuerdo al Algoritmo 2 del capítulo anterior, si queremos hallar una base para $col(A)$, con la restricción de que dicha base sea un subconjunto de $\{A^1, \ldots, A^n\}$, primero encontramos la FER de A; supongamos que $\tilde{A}$ es dicha FER. Segundo, notamos que la columna A^j estará en nuestra base $\mathcal{B}$ syss la columna $\tilde{A}^j$ contiene un uno principal. Esto nos lleva al siguiente resultado.*

Proposición 12.5 *Sea $A \in M_{m\times n}$, entonces $rank(A)$ es igual al número de unos principales de la FER de A.*

Ejemplo 12.6 *Consideremos a la matriz*

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

y supongamos que deseamos encontrar una base para $null(A)$, la nulidad de A, una base para $col(A)$ y $rank(A)$. Para ello, notamos que la FER de A es la matriz

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

por lo que el Algoritmo 2 asegura que el conjunto

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

es una base para $col(A)$ y $rank(A) = 2$. La matriz $\tilde{A}$ también nos dice que la nulidad de A es igual a 3 (pues tiene tres variables libres). Más aún, como la FER de la matriz aumentada del sistema homogéneo $A\vec{x} =$

$\vec{0}$ *es igual a* $(\tilde{A} \mid \vec{0})$ *tenemos que* $\vec{x} \in Null(A)$ *syss sus componentes satisfacen*

$$\begin{aligned} x_1 - x_2 &= 0, \\ x_3 - x_4 + 2x_5 &= 0x_2, \\ x_4, x_5 &\in \mathbb{R}. \end{aligned}$$

Por lo tanto,

$$null(A) = \left\{ \begin{bmatrix} x_2 \\ x_2 \\ x_4 - 2x_5 \\ x_4 \\ x_5 \end{bmatrix} \mid x_2, x_4, x_5 \in \mathbb{R} \right\}$$

$$= \left\{ x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \mid x_2, x_4, x_5 \in \mathbb{R} \right\}$$

$$= span \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Concluimos pues, que el conjunto

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

es una base para $Null(A)$ *por lo que la nulidad de* A *es igual a 3.*

El ejemplo anterior ilustra como el número de variables libres (nulidad de A) más el número de variables restringidas ($rank(A)$) coincide, evidentemente, con el número total de columnas (o variables). Esto puede enunciarse en forma general como sigue.

Teorema 12.7 (Teorema de la dimensión.) *Si* A *es una matriz de* $m \times n$*, entonces*

$$n = rank(A) + nulidad(A).$$

Demostración. Supogamos que $\tilde{A}$ es la FER de A y que k es el número de unos principales que allí aparecen. Entonces, de acuerdo a la observación anterior, $rank(A) = k$. Por otro lado, la matriz aumentada $(\tilde{A} \mid \vec{0})$ correspondiente al sistema $\tilde{A}\vec{x} = \vec{0}$, tiene exactamente $n - k$ variables libres. Por lo tanto,

$$nulidad(A) = n - k = n - rank(A)$$

o lo que es lo mismo,

$$n = rank(A) + nulidad(A).$$

$\blacksquare$

Definición 12.8 *Sea $A \in M_{m\times n}$. Definimos el espacio renglón de A, denotado por $row(A)$, como el conjunto*

$$row(A) = span\{A_1, \ldots, A_m\}.$$

Proposición 12.9 *Si $A \in M_{m\times n}$, entonces*

$$\dim(\mathrm{row}(A)) = \dim(\mathrm{col}(A)) = rank(A),$$

es decir, los espacios renglón y columna de una matriz tienen la misma dimensión: $rank(A)$.

Demostración. Sea $\tilde{A}$ la FER de A. De acuerdo al Algoritmo 1, los renglones de $\tilde{A}$ que contienen un uno principal forman una base para $\mathrm{row}(A)$. Asimismo, el Algoritmo 2 nos dice que las columnas de A que corresponden a las columnas de $\tilde{A}$ que tienen unos principales forman una base para $\mathrm{col}(A)$. Por lo tanto, las dimensiones de ambos subespacios coinciden con el número de unos principales de $\tilde{A}$. El resultado se sigue observando que $rank(A)$ también es igual al número de unos principales de $\tilde{A}$. $\blacksquare$

El siguiente corolario se sigue de la Proposición 12.9, dado que las columnas de la matriz A corresponden a los renglones de la matriz A^T.

Corolario 12.10 *Dada $A \in M_{m\times n}$, se tiene que $rank(A) = rank(A^T)$.*

Observación 12.11 *En este punto es convenienete señalar que aunque la dimensión del espacio columna de A es siempre igual a la dimensión del espacio renglón de A (dicho número es igual a $rank(A)$), esto no implica que $col(A) = row(A)$ ya que, en general, estos subespacios habitan distintos espacios. En general, si $A \in M_{m\times n}$, tenemos que $col(A)$ es un subespacio de $\mathbb{R}^m$ mientras que $row(A)$ es un subespacio de $\mathbb{R}^n$.*

Ejemplo 12.12 *Consideremos a la matriz*

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

cuya FER es la matriz

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

Dado que hay dos unos principales se tiene que $rank(A) = 2$. *Más aún, el Algoritmo 2 asegura que el conjunto*

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix} \right\} \subset \mathbb{R}^3$$

es una base para $col(A)$. *Por otro lado, el Algoritmo 1 asegura que el conjunto*

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} \subset R^4$$

es una base para $row(A)$. *Como* $\tilde{A}$ *tiene dos columnas que corresponden a las variables libres del sistema* $\tilde{A}\vec{x} = \vec{0}$ *se tiene que* $nulidad(A) = 2$. *Asimismo, el espacio nulo está dado por*

$$Null(A) = \left\{ x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \end{bmatrix} \mid x_3, x_4 \in \mathbb{R} \right\}.$$

Claramente se satisface el teorema de la dimensión pues

$$rank(A) + nulidad(A) = 2 + 2 = 4$$

Tomemos ahora

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 5 & 2 \\ 3 & 6 & 3 \\ 4 & 7 & 4 \end{bmatrix},$$

cuya FER está dada como

$$\tilde{A}^T = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Entonces $rank(A^T) = 2$ *y una base para* $\mathrm{col}(A^T)$ *está dada por*

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 6 \\ 7 \end{bmatrix} \right\}.$$

Observemos que esta base es distinta a la que se obtuvo arriba para los renglones de A *ya que en ese caso se utilizó el Algoritmo 1 y aquí se utilizó el algoritmo 2. El espacio* $\mathrm{row}(A^T)$ *tiene como base a*

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

que también difiere de la base obtenida para las columnas de A *por la misma razón que arriba (se utiliza otro algoritmo para obtenerla). Finalmente, es claro que* $nulidad(A^T) = 1$, *que*

$$Null(A^T) = \left\{ x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\},$$

y el Teorema de la dimensión se satisface puesto que

$$rank(A) + nulidad(A) = 2 + 1 = 3.$$

Ahora presentamos un resultado importante cuya demostración está autocontenida en el enunciado del mismo.

Corolario 12.13 *Sea* $A \in M_{n\times n}$. *Las siguientes afirmaciones son equivalentes.*

1. *A es invertible.*
2. *Las columnas de A forman un conjunto LI.*
3. *La ecuación $A\vec{x} = \vec{0}$ tiene una única solución (a saber, $\vec{x} = \vec{0}$).*
4. *$Null(A) = \{\vec{0}\}$.*
5. *$nulidad(A) = 0$.*
6. *$rank(A) = n$.*
7. *Los renglones de A forman un conjunto LI.*

Observación 12.14 *Si* $A \in M_{m\times n}$ *y* $T : \mathbb{R}^n \rightharpoonup \mathbb{R}^m$ *es la transformación lineal dada por* $T(\vec{x}) = A\vec{x}$, *claramente*

$$nulidad(A) = \dim(Null(T))$$

y

$$rank(A) = \dim(Rango(T)).$$

De esta forma, el Teorema de la Dimensión puede expresarse en términos de la transformación lineal T *como*

$$\dim(Rango(T)) + \dim(Null(T)) = n.$$

Finalizamos este capítulo con algunos ejemplos que muestran cómo los conceptos de rank y nulidad delimitan las posibilidades de que existan cierto tipo de matrices y transformaciones.

Ejemplo 12.15 *Supongamos que* $A \in M_{7\times 16}$. *¿ Es posible que la nulidad de A sea igual a 8?. Para responder a esta preguntar notamos que si existiese tal matriz A, entonces*

$$rank(A) = 16 - nulidad(A) = 16 - 8 = 8,$$

por lo que la FER de A tendría al menos 8 renglones (distintos de cero). A partir de esta contradicción concluimos que si A es una matriz de 7×16, *no es posible que su nulidad sea igual a* 8.

Ejemplo 12.16 *Supongamos que* $A \in M_{5\times 13}$. *¿Cuál es el posible valor máximo para* $rank(A)$ *y cuál es el posible valor mínimo para* $nulidad(A)$*?. Para responder la primera pregunta recordamos que* $rank(A)$ *es igual a la dimensión del espacio renglón de A, por lo que dicha dimensión puede ser como mucho 5. Por otro lado, como*

$$13 = rank(A) + nulidad(A),$$

tenemos que si $rank(A)$ *toma su valor máximo, entonces* $nulidad(A)$ *toma su valor mínimo y viceversa. Por lo tanto, el valor mínimo para la nulidad de A es 8.*

Ejemplo 12.17 *Sea* $T : \mathbb{R}^5 \rightharpoonup \mathbb{R}^4$ *y supongamos que la FER de la matriz estándar asociada a* T *tiene 3 renglones diferentes de cero. ¿Cuáles son las dimensiones de* $Null(T)$ *y de* $Rango(T)$*? Si* $\tilde{A}$ *es la FER mencionada, entonces* $rank(\tilde{A}) = 3 = \dim(Rango(T))$. *Por el Teorema de la Dimensión se tiene que*

$$\dim(Null(T)) = 5 - 3 = 2.$$

Observemos que T *no es suprayectiva (pues* $\dim(Rango(T)) < 4$*) ni tampoco inyectiva (pues* $\dim(Null(T)) > 0$*).*

Capítulo 13

Eigenvalores y eigenvectores

Introducción

Comenzamos este capítulo con una observación bastante natural. Supongamos que

$$T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

es una transformación lineal y que $\vec{x} \in \mathbb{R}^n$. ¿En qué sentido podemos decir que la transformación T no afecta "mucho" al vector $\vec{x} \in \mathbb{R}^n$?. Queda claro que si

$$T(\vec{x}) = \vec{x} = 1\vec{x},$$

entonces T no cambia en nada al vector $\vec{x}$. Análogamente, podríamos decir que T "perturba poco" a $\vec{x}$ si **existe** un escalar λ tal que[1]

$$T(\vec{x}) = \lambda\vec{x}.$$

Si A es la matriz estándar asociada a T, lo anterior equivale a

$$A\vec{x} = \lambda\vec{x}.$$

Esta situación se ilustra en la figura 13.1.

En el caso del vector cero siempre se cumple

$$A\vec{0} = \vec{0} = \lambda\vec{0}$$

[1]En este caso T simplemente reescala y/o cambia la dirección del vector $\vec{x}$.

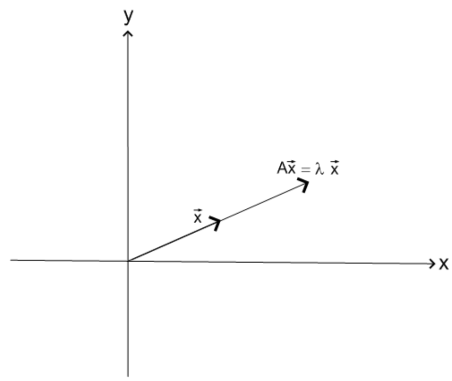

Figura 13-1 $A\vec{x} = \lambda\vec{x}$

para cualquier escalar λ, por lo que no es un caso de interés para este análisis y únicamente se consideran vectores $\vec{x} \neq \vec{0}$. Nuestro objetivo es determinar si, dada una transformación lineal, existen o no vectores (no nulos) que se perturban poco, en el sentido que acabamos de explicar. En las definiciones que siguen, la matriz A puede ser sustituida por su transformación lineal asociada T como se hará explícito más adelante.

Obtención de eigenvalores y eigenvectores

Definición 13.1 *Sean A es una matriz de $n \times n$ y λ un escalar. Decimos que λ es un* ***eigenvalor (****también llamado* ***valor propio o valor característico)*** *de A, si existe un vector $\vec{x}$ de $\mathbb{R}^n$, con $\vec{x} \neq \vec{0}$, tal que $A\vec{x} = \lambda\vec{x}$. El vector $\vec{x}$ es llamado un* ***eigenvector (o vector propio o característico)*** *de A con valor propio λ.*

Observación 13.2 *Los eigenvectores nunca son nulos, sin embargo $\lambda = 0$ si puede ser un eigenvalor. Notemos que $\vec{x}$ es un eigenvector con eigenvalor $\lambda = 0$ syss $A\vec{x} = 0\vec{x} = \vec{0}$. Equivalentemente, $\vec{x} \in Null(A)$.*

Definición 13.3 *Sea A una matriz de $n \times n$ y sea λ un eigenvalor de A. El* ***eigenespacio*** *de A asociado a λ, denotado por* $\mathrm{Ei}\,g(\lambda)$*, se define como el conjunto*

$$\mathrm{Ei}\,g(\lambda) = \{\vec{x} \mid A\vec{x} = \lambda\vec{x}\},$$

es decir, todos los eigenvectores con eigenvalor λ junto con el vector $\vec{0}$.

Una vez definidos estos conceptos, evidentemente nos gustaría saber, dada una matriz $A \in M_{n\times n}$, cómo encontrar sus eigenvalores y eigenvectores, si es que éstos existen. Para entender que es lo que estamos buscando, a continuación vemos que $\mathrm{Ei}\,g(\lambda)$ será, precisamente, el espacio nulo de una matriz, y, en particular, un subespacio de $\mathbb{R}^n$.

Proposición 13.4 *Si A es una matriz de $n \times n$ y λ es un eigenvalor*[2]*, entonces*

$$\mathrm{Ei}\,g(\lambda) = Null(A - \lambda I),$$

y por lo tanto, $\mathrm{Ei}\,g(\lambda)$ *es un subespacio de* $\mathbb{R}^n$.

Demostración. Basta notar las siguientes equivalencias: $\vec{x} \in \mathrm{Ei}\,g(\lambda)$ syss

$$A\vec{x} = \lambda\vec{x} = \lambda I\vec{x}$$

syss

$$A\vec{x} - \lambda I\vec{x} = \vec{0},$$

syss

$$(A - \lambda I)\vec{x} = \vec{0},$$

syss

$$\vec{x} \in Null(A - \lambda I).$$

■

Corolario 13.5 *Sea A una matriz de $n \times n$. Entonces las siguientes afirmaciones son equivalentes.*

(a) *λ es un eigenvalor de A.*

(b) *Existe un vector distinto cero que pertence a*

$$\mathrm{Ei}\,g(\lambda) = Null(A - \lambda I).$$

(c) $dim\,(\mathrm{Ei}\,g(\lambda)) = nulidad(A - \lambda I) \geq 1.$

(d) *Las columnas de $A - \lambda I$ forman un conjunto LD.*

(e) *La matriz $A - \lambda I$ no es invertible.*

(f) $det(A - \lambda I) = 0.$

(g) *El sistema homogéneo $(A - \lambda I)\vec{x} = \vec{0}$ tiene una infinidad de soluciones.*

Observación 13.6 *Notemos que, en particular, si $\vec{x}$ es eigenvector con eigenvalor λ, para cualquier escalar $r \neq 0$ el vector $r\vec{x}$ también lo será.*

[2] En realidad aquí λ podría ser cualquier escalar.

Ejemplo 13.7 *Sea*

$$A = \begin{bmatrix} 1 & 9 \\ 1 & 1 \end{bmatrix},$$

$\lambda \in \mathbb{R}$ *será un eigenvalor de* A *syss dada*

$$\lambda I = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

se tiene que

$$\begin{aligned} \mathrm{Ei}\, g(\lambda) &= Null(A - \lambda I) \\ &= Null\left(\begin{bmatrix} 1-\lambda & 9 \\ 1 & 1-\lambda \end{bmatrix} \right) \neq \{\vec{0}\}, \end{aligned}$$

equivalentemente se cumple

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 1-\lambda & 9 \\ 1 & 1-\lambda \end{bmatrix} \right) =$$
$$(1-\lambda)(1-\lambda) - 9 = (1-\lambda)^2 - 9 = 0.$$

Resolviendo para λ *obtenemos* $\lambda_1 = 4, \lambda_2 = -2$*, por lo que*

$$\begin{aligned} \mathrm{Ei}\, g(\lambda_1) &= Null\left(\begin{bmatrix} -3 & 9 \\ 1 & -3 \end{bmatrix} \right), \\ \mathrm{Ei}\, g(\lambda_2) &= Null\left(\begin{bmatrix} 3 & 9 \\ 1 & 3 \end{bmatrix} \right). \end{aligned}$$

Es fácil ver (se deja como ejercicio al lector) que

$$\begin{aligned} \mathrm{Ei}\, g(\lambda_1) &= span\left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}, \\ \mathrm{Ei}\, g(\lambda_1) &= span\left\{ \begin{bmatrix} -3 \\ 1 \end{bmatrix} \right\}. \end{aligned}$$

En particular puede verificarse que si $\vec{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ *y* $\vec{v}_2 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ *se tiene*

$$\begin{aligned} A\vec{v}_1 &= \begin{bmatrix} 1 & 9 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 12 \\ 4 \end{bmatrix} \\ &= 4 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \lambda_1 \vec{v}_1, \\ A\vec{v}_2 &= \begin{bmatrix} 1 & 9 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \end{bmatrix} \\ &= -2 \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \lambda_2 \vec{v}_2. \end{aligned}$$

En el ejemplo anterior observamos que los eigenvalores de A resultan ser las raices de un polinomio, se tiene así la siguiente definición.

Definición 13.8 *Sea $A \in M_{n\times n}$ y λ una variable. Definimos al* ***polinomio característico*** *de A como*

$$p_A(\lambda) = det(A - \lambda I).$$

Utilizando un argumento inductivo es posible demostrar que, efectivamente, $p_A(\lambda)$ es un polinomio y que su grado es igual a n. Recordemos que si

$$p(x) = a_n x^n + \ldots a_1 x + a_0 = a_n x^n + \ldots a_1 x + a_0 x^0$$

es un polinomio con coeficientes reales distinto del polinomio constante cero, entonces el **grado** de $p(x)$, denotado por $\delta(p(x))$, se define como el máximo $m \leq n$ tal que $a_m \neq 0$ (el grado del polinomio constante cero se define como $-\infty$). Por ejemplo,

$$\delta(0x^8 + 6x^2 + 7) = 2 \text{ y } \delta(6) = \delta(6x^0) = 0.$$

En lo anterior se hace énfasis en que

$$a_0 = a_0 \times 1 = a_0 x^0.$$

También recordamos que si

$$p(x) = a_n x^n + \ldots a_1 x + a_0$$

es un polinomio y r es un número real, decimos que r es una **raíz (**o un **"cero")** de $p(x)$ syss $p(r) = 0$. Utilizando esta terminología, el resultado anterior (o más bien parte de él) puede reescribirse como sigue.

Corolario 13.9 *Si A es una matriz de $n \times n$ y λ^* es un número real, entonces λ^* es un eigenvalor de A syss λ^* es una raíz del polinomio característico $p_A(\lambda)$.*

Ejemplo 13.10 *Consideremos a la matriz*

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

y supongamos que deseamos determinar los eigenvalores de A así como una base para cada uno de sus eigenespacios. Para ello, notamos que el polinomio característico de A es igual a

$$\begin{aligned} p_A(\lambda) = det(A - \lambda I) &= det\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= det\begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 - 1 \end{aligned}$$

Por lo tanto, $p_A(\lambda) = 0$ *syss* $\lambda = 1$ *o* $\lambda = -1$ *y los dos eigenvalores de* A *son* $\lambda_1 = 1$ *y* $\lambda_2 = -1$.

Recordamos ahora que

$$\mathrm{Ei}\, g(\lambda_1) = Null(A - \lambda_1 I) = Null \begin{bmatrix} -\lambda_1 & 1 \\ 1 & -\lambda_1 \end{bmatrix} = Null \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Las soluciones al sistema

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

se obtienen fácilmente[3] *y están dadas por*

$$\begin{aligned} &\left\{ \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} \in \mathbb{R}^2 \mid x_2 \in \mathbb{R} \right\} \\ &= \left\{ x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2 \mid x_2 \in \mathbb{R} \right\}. \end{aligned}$$

Por lo tanto,

$$\mathrm{Ei}\, g(\lambda_1) = span \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\, g(\lambda_1)$. *Por otro lado,*

$$\mathrm{Ei}\, g(\lambda_2) = Null(A - \lambda_2 I) = Null \begin{bmatrix} -\lambda_2 & 1 \\ 1 & -\lambda_2 \end{bmatrix} = Null \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Obteniendo las soluciones para el sistema

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

y procediendo como antes, llegamos a que

$$\mathrm{Ei}\, g(\lambda_2) = span \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\, g(\lambda_2)$.

---

[3] Observando que la FER de la matriz del sistema es simplemente

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

Debido a la correspondencia entre matrices y transformaciones lineales las siguientes dos definiciones resultan naturales.

Definición 13.11 *Sea $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una transformación lineal y sea $A \in M_{n\times n}$ la matriz estándar que representa a T (esto es, $T(\vec{x}) = A\vec{x}$). Dado un escalar λ, decimos que λ es un* ***eigenvalor (valor propio)*** *de T syss λ es un eigenvalor de A. Dado $\vec{x} \in \mathbb{R}^n$ también decimos que $\vec{x}$ es un eigenvector (vector propio) de T con respecto a λ syss $\vec{x}$ es un* ***eigenvector (vector propio)*** *de A con respecto a λ.*

Definición 13.12 *Sean $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una transformación lineal, $A \in M_{n\times n}$ la matriz estándar que representa a T y λ un eigenvalor de T. El* ***eigenespacio*** *de T asociado a λ se define como el eigenespacio de A asociado a λ, es decir,* $\mathrm{Ei}\,g(\lambda)$.

Ejemplo 13.13 *Supongamos que $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ es la transformación lineal que satisface*

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ -2x+4y \end{bmatrix}$$

y supongamos que deseamos determinar los eigenvalores de T así como una base para cada uno de sus eigenespacios. Para ello, notamos primero que la matriz estándar que representa a T es igual a

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

y su polinomio característico es igual a

$$\begin{aligned} p_A(\lambda) &= det(A - \lambda I) \\ &= det\left(\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= det\begin{bmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{bmatrix} = (1-\lambda)(4-\lambda)+2 \\ &= 4 - \lambda - 4\lambda + \lambda^2 + 2 = \lambda^2 - 5\lambda + 6. \end{aligned}$$

Por lo tanto, $p_A(\lambda) = 0$ syss $\lambda = 3$ o $\lambda = 2$ y los dos eigenvalores de T (que son los mismos que los de A) son $\lambda_1 = 3$ y $\lambda_2 = 2$. Como

$$\begin{aligned} \mathrm{Ei}\,g(\lambda_1) &= Null(A - \lambda_1 I) \\ &= Null\begin{bmatrix} 1-\lambda_1 & 1 \\ -2 & 4-\lambda_1 \end{bmatrix} = Null\begin{bmatrix} -2 & 1 \\ -2 & 1 \end{bmatrix}. \end{aligned}$$

y el conjunto solución del sistema homogéneo

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

está dado como

$$\left\{ y \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \in \mathbb{R}^2 \mid y \in \mathbb{R} \right\} = \left\{ s \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2 \mid s \in \mathbb{R} \right\}$$

(en donde $s = \frac{1}{2}y$) por lo que

$$\mathrm{Ei}\,g(\lambda_1) = span \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\,g(\lambda_1)$*.*

Análogamente,

$$\mathrm{Ei}\,g(\lambda_2) = Null(A - \lambda_2 I) = Null \begin{bmatrix} 1-\lambda_2 & 1 \\ -2 & 4-\lambda_2 \end{bmatrix} = Null \begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix}.$$

El conjunto solución del sistema homogéneo

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

está dado por

$$\left\{ y \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2 \mid y \in \mathbb{R} \right\},$$

con lo cual

$$\mathrm{Ei}\,g(\lambda_2) = span \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\,g(\lambda_2)$*.*

Hasta ahora sólo hemos considerado ejemplos donde los eigenespacios en cuestión tienen dimensión uno (la menor posible), pero el siguiente ejemplo involucra a un eigenespacio de dimensión mayor.

Ejemplo 13.14 *Consideremos a la matriz*

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

y supongamos que deseamos determinar los eigenvalores de B así como una base para cada uno de sus eigenespacios. Para ello, notamos que el polinomio característico de B es igual a

$$\begin{aligned} p_B(\lambda) &= det(B - \lambda I) \\ &= det\left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \\ &= det \begin{bmatrix} -\lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 1 & 0 & 1-\lambda \end{bmatrix} = -\lambda(1-\lambda)^2. \end{aligned}$$

Por lo tanto, $p_B(\lambda) = 0$ syss $\lambda = 0$ o $\lambda = 1$ y los dos eigenvalores de B son $\lambda_1 = 0$ y $\lambda_2 = 1$. Recordamos ahora que

$$\begin{aligned} \mathrm{Ei}\, g(\lambda_1) &= Null(B - \lambda_1 I) \\ &= Null \begin{bmatrix} 0-\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_1 & 0 \\ 1 & 0 & 1-\lambda_1 \end{bmatrix} = Null \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Dado que la FER de

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

está dada por

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

el conjunto solución del sistema homogéneo

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

es

$$\left\{ z \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \in \mathbb{R}^3 \mid z \in \mathbb{R} \right\}$$

por lo que

$$\mathrm{Ei}\, g(\lambda_1) = span \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\,g(\lambda_1)$.

Análogamente,

$$\begin{aligned}\mathrm{Ei}\,g(\lambda_2) &= Null(B-\lambda_2 I)\\ &= Null\begin{bmatrix}0-\lambda_2 & 0 & 0\\ 0 & 1-\lambda_2 & 0\\ 1 & 0 & 1-\lambda_2\end{bmatrix} = Null\begin{bmatrix}-1 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\end{bmatrix}.\end{aligned}$$

Pero como

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

es la FER de

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

se tiene

$$\mathrm{Ei}\,g(\lambda_2) = span\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

y

$$\left\{\begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

es una base para $\mathrm{Ei}\,g(\lambda_2)$.

Como consecuencia directa del siguiente resultado podemos afirmar que la matriz B del ejemplo anterior no es invertible.

Proposición 13.15 *Si A es una matriz de $n \times n$, entonces A* ***no*** *es invertible syss 0 es un eigenvalor de A.*

Demostración. Basta notar que A no es invertible syss

$$0 = det(A) = det(A - 0I) = p_A(0).$$

Por lo tanto, A no es invertible syss 0 es un eigenvalaor de A. ■

Ejemplo 13.16 *Consideremos a la matriz*

$$C = \begin{bmatrix}4 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}$$

y supongamos que deseamos determinar los eigenvalores de C así como una base para cada uno de sus eigenespacios. Para ello, notamos que el polinomio característico de C es igual a

$$p_C(\lambda) = det(C - \lambda I)$$
$$= det \left(\begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$
$$= det \begin{bmatrix} 4-\lambda & 0 & 0 \\ 0 & 1-\lambda & 1 \\ 0 & 0 & 1-\lambda \end{bmatrix} = (4-\lambda)(1-\lambda)^2.$$

Por lo tanto, $p_C(\lambda) = 0$ syss $\lambda = 4$ o $\lambda = 1$ y los dos eigenvalores de B son $\lambda_1 = 4$ y $\lambda_2 = 1$. Recordamos ahora que

$$\mathrm{Ei}\, g(\lambda_1) = Null(C - \lambda_1 I)$$
$$= Null \begin{bmatrix} 4-\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_1 & 1 \\ 0 & 0 & 1-\lambda_1 \end{bmatrix} = Null \begin{bmatrix} 0 & 0 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & -3 \end{bmatrix}$$

y como

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

es la FER de

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

se tiene que

$$\mathrm{Ei}\, g(\lambda_1) = \left\{ x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid x \in \mathbb{R} \right\} = span \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\, g(\lambda_1)$*. Por otro lado,*

$$\mathrm{Ei}\, g(\lambda_2) = Null(C - \lambda_2 I)$$
$$= Null \begin{bmatrix} 4-\lambda_2 & 0 & 0 \\ 0 & 1-\lambda_2 & 1 \\ 0 & 0 & 1-\lambda_2 \end{bmatrix} = Null \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Como

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

es la FER de

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene

$$\mathrm{Ei}\, g(\lambda_2) = \left\{ y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \in \mathbb{R}^3 \mid y \in \mathbb{R} \right\} = span \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

y

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

es una base para $\mathrm{Ei}\, g(\lambda_2)$*.*

Notamos que el polinomio característico de C puede reescribirse como

$$p_C(\lambda) = -(\lambda - 4)(\lambda - 1)^2,$$

por lo que podemos decir que $\lambda_1 = 4$ tiene multiplicidad 1 mientras que $\lambda_2 = 1$ tiene multiplicidad 2 en este polinomio. En general, si $p(x)$ es un polinomio y r es una raíz de $p(x)$, definimos a la **multiplicidad** de r en el polinomio $p(x)$ como el máximo número natural m tal que $(x - r)^m$ divide al polinomio $p(x)$.

Los Ejemplos 13.14 y 13.16 sugieren el siguiente resultado, cuya demostración omitimos.

Proposición 13.17 *Sea* $A \in M_{n\times n}$ *y supongamos que* λ^* *es un eigenvalor de* A *(i.e.,* λ^* *es una raíz de* $p_A(\lambda)$*). Si* λ^* *tiene multiplicidad* m *en* $p_A(\lambda)$*, entonces* $dim(\mathrm{Ei}\, g(\lambda^*)) \leq m$*.*

Ejemplo 13.18 *En el ejemplo 13.14 encontramos que el polinomio característico de* B *es igual a*

$$p_B(\lambda) = -\lambda(1 - \lambda)^2 = -(\lambda - 0)(\lambda - 1)^2$$

con raices $\lambda_1 = 0$ *y* $\lambda_2 = 1$*. Además,*

$$dim(\mathrm{Ei}\, g(\lambda_1)) = 1 = multiplicidad(\lambda_1),$$
$$dim(\mathrm{Ei}\, g(\lambda_2)) = 2 = multiplicidad(\lambda_2).$$

Ejemplo 13.19 *En el ejemplo 13.16 encontramos que el polinomio característico de C es igual a*

$$p_C(\lambda) = (4-\lambda)(1-\lambda)^2 = -(\lambda-4)(\lambda-1)^2$$

con raices $\lambda_1 = 4$ y $\lambda_2 = 1$. Además,

$$\begin{aligned} dim(\mathrm{Ei}\,g(\lambda_1)) &= 1 = multiplicidad(\lambda_1), \\ dim(\mathrm{Ei}\,g(\lambda_2)) &= 1 \leq 2 = multiplicidad(\lambda_2). \end{aligned}$$

Corolario 13.20 *Sea $A \in M_{n\times n}$ y supongamos que λ^* es un eigenvalor de A. Si λ^* tiene multiplicidad 1 en $p_A(\lambda)$, entonces $dim(\mathrm{Ei}\,g(\lambda^*) = 1$.*

Demostración. Como λ^* es un eigenvalor de A, es claro que

$$dim(\mathrm{Ei}\,g(\lambda^*) \geq 1.$$

Por otro lado, gracias a la Proposición 13.17, se tiene que

$$dim(\mathrm{Ei}\,g(\lambda^*) \leq 1$$

por lo que, necesariamente, $dim(\mathrm{Ei}\,g(\lambda^*) = 1$. ■

Observación 13.21 *Supongamos que se desean obtener los eigenvalores y eigenvectores de la matriz*

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

Procediendo como lo hemos hecho, buscamos las raíces del polinomio característico dado por

$$p_A(\lambda) = \det(A-\lambda I) = \det \begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix} = \lambda^2 + 1,$$

*pero notamos que la ecuación $\lambda^2+1=0$ no tiene soluciones reales. Las raíces de este polinomio son $\pm i$, en donde i es tal que $i^2 = -1$ (podríamos denotar a i como $\sqrt{-1}$). Estas raíces habitan en lo que se conoce como el conjunto de **números complejos** (conjunto que contiene a los números reales, que posee las mismas propiedades algebraicas de campo y que denotamos por $\mathbb{C}$).*

El Teorema Fundamental del Álgebra nos garantiza que cualquier polinomio de grado n con coeficientes reales tendrá n raíces (no necesariamente todas distintas) en los números complejos. Dado que en este texto únicamente consideramos escalares reales y espacios vectoriales $\mathbb{R}^n$, no abordaremos el caso de eigenvalores y eigenvectores complejos. En consecuencia, existirán matrices, como la dada al inicio de esta observación, que no poseen eigenvalores o eigenvectores reales.

Cadenas de Markov (opcional)

En ocasiones es necesario utilizar **potencias de matrices** en el sentido usual: dada una matriz A de $n \times n$, denotamos por $A^{(p)}$ al producto de A consigo misma p veces, es decir,

$$A^{(p)} = \underbrace{A \cdots A}_{p\ veces}.$$

Así, por ejemplo, $A^{(2)}$ es igual al producto AA. Por supuesto, la notación $A^{(p)}$ puede parecer peculiar, pero es simplemente para distinguir la potencia p, de la p-ésima columna de A, a la cual denotamos en este texto por A^p. La siguiente proposición nos da una relación simple entre los eigenvalores de A y de $A^{(p)}$.

Proposición 13.22 *Sea A una matriz de $n \times n$. Si λ es un eigenvalor de A, entonces λ^p es un eigenvalor de $A^{(p)}$, con el mismo eigenvector.*

Demostración. Utilizamos inducción sobre p. Si $p = 1$ el resultado es válido pues nuestra hipótesis sobre λ afirma que existe un vector $\vec{x}$ de $\mathbb{R}^n$ tal que $\vec{x} \neq \vec{0}$ y $A\vec{x} = \lambda\vec{x}$. Sea $p > 1$ y asumimos el resultado válido para $p - 1$. Entonces,

$$\begin{aligned} A^{(p)}\vec{x} &= (\underbrace{A \cdots A}_{p\ veces})\vec{x} = A(\underbrace{A \cdots A}_{(p-1)\ veces})\vec{x} \\ \underset{\text{hipótesis de inducción}}{=} & A(\lambda^{p-1}x) = \lambda^{p-1}(A\vec{x}) = \lambda^{p-1}(\lambda\vec{x}) = \lambda^p\vec{x}. \end{aligned}$$

y, por lo tanto, λ^p es eigenvalor de $A^{(p)}$ con el mismo eigenvector $\vec{x}$. ■

Un buen número de problemas pueden modelarse como sistemas cuyos estados evolucionan de forma “uniforme”. Veamos un ejemplo sencillo.

Ejemplo 13.23 *Los estudiantes de una universidad pueden elegir entre comer en la cafetería universitaria o bien en algún restaurante externo. Supongamos que, el 30 % de los estudiantes que comen en la cafetería van a un restaurante en la siguiente ocasión (el 70 % vuelve a comer en la cafetería) y el 80 % de quienes comen en algún restaurante regresan a la cafetería universitaria para su siguiente comida (el 20 % vuelve a comer en un restaurante). Podemos representar esta situación con la tabla 13.1 en la cual C y R se refieren a la cafetería a a un restaurante, respectivamente.*

Tabla 13.1

	C	***R***
C	*0.7*	*0.8*
R	*0.3*	*0.2*

Aún mejor, podemos ser más eficientes y sustituir la tabla por la matriz

$$P = \begin{bmatrix} 0.7 & 0.8 \\ 0.3 & 0.2 \end{bmatrix}.$$

Al inicio del semestre, el estado inicial consiste en que el 40 % de los estudiantes comen en la cafetería y el 60 % en algún restaurante. Al día siguiente se tendrá que como

$$0.7 \times 0.4 + 0.8 \times 0.6 = 0.76,$$

el 76 % comerán en la cafetería y el 24 %, es decir,

$$0.3 \times 0.4 + 0.2 \times 0.6 = 0.24,$$

comerán en un restaurante. Observemos que si representamos a la distribución o estado inicial de estudiantes que comen en cada lugar por el vector

$$\vec{x}_0 = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix},$$

entonces la distribución (estado) en el periodo siguiente, digamos $\vec{x}_1$*, puede obtenerse como*

$$\vec{x}_1 = \begin{bmatrix} 0.76 \\ 0.24 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.8 \\ 0.3 & 0.2 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = P\vec{x}_0.$$

Siguiendo este mismo procedimiento, en los periodos subsecuentes tendremos que

$$\begin{aligned} \vec{x}_2 &= P\vec{x}_1 = P^{(2)}\vec{x}_0, \\ \vec{x}_3 &= P\vec{x}_2 = P^{(3)}\vec{x}_0, \\ &\vdots \\ \vec{x}_{i+1} &= P\vec{x}_i = P^{(i+1)}\vec{x}_0. \end{aligned}$$

Calculando algunas potencias de P *tenemos,*

$$\begin{aligned} P^{(2)} &= \begin{bmatrix} 0.73 & 0.72 \\ 0.27 & 0.28 \end{bmatrix}, \\ P^{(3)} &= \begin{bmatrix} 0.727 & 0.728 \\ 0.273 & 0.272 \end{bmatrix}, \\ P^{(4)} &= \begin{bmatrix} 0.727\,3 & 0.727\,2 \\ 0.272\,7 & 0.272\,8 \end{bmatrix}, \end{aligned}$$

con lo cual

$$\vec{x}_2 = P^{(2)}\vec{x}_0 = \begin{bmatrix} 0.724 \\ 0.276 \end{bmatrix},$$
$$\vec{x}_3 = P^{(3)}\vec{x}_0 = \begin{bmatrix} 0.727\,6 \\ 0.272\,4 \end{bmatrix},$$
$$\vec{x}_4 = P^{(4)}\vec{x}_0 = \begin{bmatrix} 0.727\,24 \\ 0.272\,76 \end{bmatrix}.$$

A continuación vamos a formalizar la situación descrita en el Ejemplo 13.23.

Definición 13.24 *Un vector $\vec{x}_0 \in \mathbb{R}^n$ se denomina un **vector de probabilidad** si todas sus entradas son no negativas y suman 1. Asimismo, decimos que una matriz $P \in M_{n\times n}$ es una **matriz de transición** (también llamada **matriz estocástica**), si sus columnas son vectores de probabilidad.*

Ejemplo 13.25 *los vectores*

$$\begin{bmatrix} \frac{1}{7} \\ \frac{6}{7} \end{bmatrix}, \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}$$

son vectores de probabilidad. La matriz

$$\begin{bmatrix} 0 & 0.7 & 0.4 \\ 0.5 & 0.1 & 0.3 \\ 0.5 & 0.2 & 0.3 \end{bmatrix}$$

es una matriz de transición, pero la matriz

$$\begin{bmatrix} 0.6 & 0.75 \\ 0.5 & 0.25 \end{bmatrix}$$

no lo es pues su primera columna no es un vector de probabilidad.

La siguiente proposición nos proporciona algunos resultados en relación a los vectores de probabilidad y las matrices de transición.

Proposición 13.26 *Dadas P y $Q \in M_{n\times n}$ matrices de transición y $\vec{x}_0 \in \mathbb{R}^n$ un vector de probabilidad, se cumplen:*

1. *PQ es una matriz de transición.*
2. *$P^{(k)}$ una matriz de transición para todo $k \in \mathbb{N}$.*

3. *Dado $\vec{x}_0$, si definimos $\vec{x}_{i+1} = P\vec{x}_i$, entonces los vectores $\vec{x}_1, \vec{x}_2, \ldots$ son vectores de probabilidad.*

Demostración. Sea $\vec{1} \in \mathbb{R}^n$ el vector cuyas entradas son todas igual a la unidad, es decir,

$$\vec{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Para demostrar la primera parte, observemos que como P es estocástica, la suma de sus vectores renglón es igual al vector renglón en el cual todas sus entradas son iguales a la unidad, es decir,

$$P_1 + P_2 + \cdots + P_n = \vec{1}^T.$$

Como Q es de transición, dada Q^j (su $j - \acute{e}sima$ columna) se tiene que

$$\vec{1} \cdot Q^j = \vec{1}^T Q^j = 1,$$

ya que este producto es simplemente la suma de las entradas de Q^j que es igual a 1, puesto que es un vector de probabilidad. Como $(PQ)_{ij} = P_i Q^j$, se tiene que la suma de los elementos de la $j - \acute{e}sima$ columna de la matriz PQ está dada como

$$\begin{aligned} P_1 Q^j + P_2 Q^j + \cdots P_n Q^j &= (P_1 + P_2 + \cdots P_n)\, Q^j \\ &= \vec{1}^T Q^j = 1. \end{aligned}$$

La segunda parte es inmediata tomando $Q = P$ y utilizando inducción sobre k.

Para la tercera parte, como $\vec{x}_0$ es un vector de probabilidad se cumple que la suma de todas sus entradas es la unidad, lo cual puede expresarse en forma compacta (igual que arriba) como

$$\vec{1} \cdot \vec{x}_0 = \vec{1}^T \vec{x}_0 = 1.$$

Dado $\vec{x}_0$, el vector $\vec{x}_{i+1}$, $i = 0, 1, 2, \ldots$, está definido por

$$\vec{x}_{i+1} = P\vec{x}_i = P^{(i+1)} \vec{x}_0,$$

en donde, por la segunda parte de esta proposición, $P^{(i+1)}$ es una matriz de probabilidad que satisface (igual que arriba)

$$P_1^{(i+1)} + P_2^{(i+1)} + \cdots + P_n^{(i+1)} = \vec{1}^T.$$

Entonces, como

$$\vec{x}_{i+1} = P^{(i+1)}\vec{x}_0 = \begin{bmatrix} P_1^{(i+1)}\vec{x}_0 \\ P_2^{(i+1)}\vec{x}_0 \\ \vdots \\ P_n^{(i+1)}\vec{x}_0 \end{bmatrix},$$

las entradas de este vector son no negativas y suman 1 ya que se cumple,

$$\begin{aligned} P_1^{(i+1)}\vec{x}_0 + P_2^{(i+1)}\vec{x}_0 + \cdots + P_n^{(i+1)}\vec{x}_0 &= \left(P_1^{(i+1)} + P_2^{(i+1)} + \cdots + P_n^{(i+1)}\right)\vec{x}_0 \\ &= \vec{1}^T\vec{x}_0 = 1. \end{aligned}$$

Concluimos así que si $\vec{x}_0$ es un vector de probabilidad, los vectores $\vec{x}_{i+1}$, $i = 0, 1, 2, \ldots,$ también lo serán. ■

Estamos listos para definir lo que se entiende por una cadena de Markov.

Definición 13.27 *Sea $P \in M_{n\times n}$ una matriz de transición y sea $\vec{x}_0 \in \mathbb{R}^n$ un vector de probabilidad. Si $\vec{x}_{i+1} = P\vec{x}_i, i = 0, 1, 2, \ldots,$ entonces la sucesión de vectores de probabilidad, $\{\vec{x}_i\} = \vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ se conoce como una **cadena de Markov**.*

El futuro es estable (a veces)

Una pregunta natural a partir del Ejemplo 13.23 es si los vectores de una cadena de Markov se estabilizan. Esto es, nos preguntamos si existe un vector de probabilidad $\vec{x}$ tal que

$$\lim_{n\to\infty} \vec{x}_n = \vec{x} \text{ y } P\vec{x} = \vec{x} = 1\vec{x}.$$

.

Llamamos a este vector $\vec{x}$ un **vector estable**. En caso de que $\vec{x}$ exista, por definición será un eigenvector (de probabilidad) de la matriz de transición P con eigenvalor igual a 1. Observemos que si

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

es cualquier eigenvector con eigenvalor igual a λ, es decir $\vec{y} \in \mathrm{Ei}\,g(\lambda)$, entonces el vector

$$\vec{x} = \left(\frac{1}{y_1 + y_2 + \cdots y_n}\right)\vec{y}$$

es un eigenvector de probabilidad con el mismo eigenvalor. En otras palabras, cualquier eigenvector puede transformarse en un eigenvector de probabilidad con el mismo eigenvalor con un simple reescalamiento.

Definición 13.28 *Sea $P \in M_{n\times n}$ una matriz de transición. Decimos que P es* regular *syss para algún entero $k \geq 1$ la matriz potencia $P^{(k)}$ es tal que todas sus entradas son estrictamente positivas.*

Nótese que la matriz $P \in M_{2\times 2}$ del Ejemplo 13.23 es regular pues podemos tomar $k = 1$.

Ejemplo 13.29 *La matriz*

$$\begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}$$

es regular si tomamos $k = 2$, en efecto

$$\begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix}\begin{bmatrix} 0 & 0.5 \\ 1 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.25 \\ 0.5 & 0.75 \end{bmatrix}.$$

Ejemplo 13.30 *La matriz*

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

no es regular puesto que

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{(k)} = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \textit{si } k \textit{ es par,} \\ \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \textit{si } k \textit{ es impar.} \end{cases}$$

El siguiente teorema, cuya demostración omitimos, nos proporciona una condición suficiente para la existencia de un vector estable en una cadena de Markov.

Teorema 13.31 *Si $P \in M_{n\times n}$ es una matriz de transición regular, entonces*

(a) Existe un único vector de probabilidad $\vec{x}$ tal que $P\vec{x} = \vec{x}$.

(b) Si $\vec{x}_0$ es cualquier vector de probabilidad, la sucesión $\vec{x}_0, \vec{x}_1, \vec{x}_2, \ldots$ converge a $\vec{x}$, independientemente de la elección de $\vec{x}_0$.

Observación 13.32 *Es claro que el vector $\vec{x}$ del teorema anterior es un vector estable.*

En el caso particular de la matriz regular $P \in M_{2\times 2}$ del Ejemplo 13.23 , para encontrar este vector estable debemos buscar al único vector de probabilidad en el eigenespacio

$$\begin{aligned}\mathrm{Ei}\, g(\lambda = 1) &= \{\vec{x} \mid P\vec{x} = \vec{x}\} = Null(P - I)\\ &= Null \begin{bmatrix} 0.7-1 & 0.8 \\ 0.3 & 0.2-1 \end{bmatrix} = Null \begin{bmatrix} -0.3 & 0.8 \\ 0.3 & -0.8 \end{bmatrix}.\end{aligned}$$

Como la FER la matriz $\begin{bmatrix} -0.3 & 0.8 \\ 0.3 & -0.8 \end{bmatrix}$ está dada por

$$\begin{bmatrix} 1 & -\frac{8}{3} \\ 0 & 0 \end{bmatrix},$$

se tiene que[4]

$$\begin{aligned}\mathrm{Ei}\, g(\lambda = 1) &= \left\{ b \begin{bmatrix} \frac{8}{3} \\ 1 \end{bmatrix} \in \mathbb{R}^2 \mid b \in \mathbb{R} \right\}\\ &= \left\{ s \begin{bmatrix} 8 \\ 3 \end{bmatrix} \in \mathbb{R}^2 \mid s \in \mathbb{R} \right\}.\end{aligned}$$

Entonces

$$\vec{x} = \frac{1}{8+3} \begin{bmatrix} 8 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{8}{11} \\ \frac{3}{11} \end{bmatrix}$$

es el (eigen) vector estable. Esto quiere decir que, conforme avanza el semestre, el porcentaje de estudiantes que come en la cafetería se acerca a $\frac{8}{11} = 0.\overline{72}$ y el que come en algún restaurante converge a $\frac{3}{11} = 0.\overline{27}$.

Ejemplo 13.33 *Si tomamos una vez más la matriz*

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

del Ejemplo 13.30, es fácil verificar que el vector $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$ *es un eigenvector de probabilidad con eigenvalor* $\lambda = 1$*. No obstante, éste no es un vector estable ya que dado cualquier vector de probabilidad inicial,*

$$\vec{x}_0 = \begin{bmatrix} a \\ 1-a \end{bmatrix} \in \mathbb{R}^2,\ 0 \le a \le 1,$$

se tiene que la cadena de Markov es una sucesión alternante que no converge dada por,

$$\begin{bmatrix} a \\ 1-a \end{bmatrix}, \begin{bmatrix} 1-a \\ a \end{bmatrix}, \begin{bmatrix} a \\ 1-a \end{bmatrix}, \ldots$$

Esto último sucede puesto que la matriz de transición no es regular.

[4] Si $b = 3s$.

Diagonalización

En esta última sección, dada una matriz $A \in M_{n\times n}$, daremos una condición necesaria y suficiente para garantizar que hay una base de $\mathbb{R}^n$ que consiste de eigenvectores de A. El siguiente resultado es el primer paso en esta dirección.

Proposición 13.34 *Sea A una matriz de $n \times n$ y sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ eigenvalores distintos de A. Supongamos que $\vec{v}_i$ es un eigenvector para λ_i, $i = 1, 2, \ldots, k$. Entonces el conjunto*

$$\{\vec{v}_1, \vec{v}_2 \ldots, \vec{v}_k\}$$

es LI.

Demostración. La demostración se hará por inducción sobre k. Si $k = 1$, el conjunto $\{\vec{v}_1\}$ es LI, puesto que $\vec{v}_1 \neq \vec{0}$. Para probar que $\{\vec{v}_1, \vec{v}_2 \ldots, \vec{v}_k\}$ es LI, consideramos la expresión

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \ldots + a_k\vec{v}_k = \vec{0} \tag{13.1}$$

para $a_1, a_2, \ldots, a_n$ escalares reales. Multiplicando por A de ambos lados de esta expresión y utilizando el hecho de que $A\vec{v}_i = \lambda_i\vec{v}_i$ para $i = 1, 2, \ldots, k$ se tiene que

$$a_1\lambda_1\vec{v}_1 + a_2\lambda_2\vec{v}_2 + \ldots + a_k\lambda_k\vec{v}_k = A\vec{0} = \vec{0}. \tag{13.2}$$

Multiplicando (13.1) por λ_k y restando la expresión resultante a (13.2) se obtiene

$$a_1(\lambda_1 - \lambda_k)\vec{v}_1 + a_2(\lambda_2 - \lambda_k)\vec{v}_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)\vec{v}_{k-1} = \vec{0}.$$

Por hipótesis de inducción estos $k - 1$ vectores son LI y, por ende, $a_i(\lambda_i - \lambda_k) = 0$ con $\lambda_i \neq \lambda_k$ para $i = 1, 2, \ldots, k - 1$. Por lo tanto, $a_i = 0$ para $i = 1, 2, \ldots, k - 1$. Por (13.1) $a_k\vec{v}_k = \vec{0}$ y como $\vec{v}_k \neq \vec{0}$, necesariamente $a_k = 0$. Por lo tanto, el conjunto $\{\vec{v}_1, \vec{v}_2 \ldots, \vec{v}_k\}$ es LI. ■

Corolario 13.35 *Sea A una matriz de $n \times n$. Si A tiene exactamente n eigenvalores distintos, entonces existe una base para $\mathbb{R}^n$ que consiste de eigenvectores de A.*

Demostración. Supongamos que $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los n eigenvalores distintos de A y supongamos que para, $i \in \{1, \ldots, n\}$, $\vec{v}_i$ es un eigenvector de A con respecto a λ_i. Por la proposición anterior tenemos ahora que el conjunto

$$\{\vec{v}_1, \vec{v}_2 \ldots, \vec{v}_n\}$$

es LI y, por lo tanto, una base para $\mathbb{R}^n$ que consiste de eigenvectores de A. ■

Ejemplo 13.36 *En el Ejemplo 13.10 notamos que la matriz*

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

tiene dos eigenvalores: $\lambda_1 = 1$ *y* $\lambda_2 = -1$*. También notamos que los conjuntos*

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \; y \; \left\{ \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

son, respectivamente, una base para $\mathrm{Ei}\, g(\lambda_1)$ *y* $\mathrm{Ei}\, g(\lambda_2)$*. Por lo tanto, el Corolario 13.35 afirma que el conjunto*

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathbb{R}^2$ *que consiste de eigenvectores de* A*.*

Corolario 13.37 *Sea* A *una matriz de* $n \times n$ *y sean* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *eigenvalores distintos de* A*. Supongamos que* $\vec{v}_i \in \mathrm{Ei}\, g(\lambda_i)$, $i = 1, 2, \ldots, k$*. Entonces*

$$\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_k = \vec{0} \quad syss \quad \vec{v}_i = \vec{0} \text{ para } i = 1, 2, \ldots, k.$$

Demostración. Si $\vec{v}_i = \vec{0}$ para $i = 1, 2, \ldots, k$. evidentemente se cumple

$$\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_k = \vec{0}\,. \tag{13.3}$$

Si algunos de estos vectores fuesen no nulos, sin pérdida de generalidad podemos suponer que $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_j$ con $j \leq k$ son diferentes del $\vec{0}$ (es decir, asumimos que los primeros j vectores son no nulos). Entonces

$$\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_j = \vec{0}$$

sería una combinación lineal de eigenvectores con eigenvalores distintos que es igual a $\vec{0}$.Esto último es imposible por la Proposición 13.34 ya que estos vectores son LI, por lo tanto debe cumplirse que $\vec{v}_i = \vec{0}$ para $i = 1, 2, \ldots, k$. ■

Dada una matriz $A \in M_{n\times n}$ queremos determinar cuando se tiene una base de eigenvectores de A para $\mathbb{R}^n$. Antes de enunciar este resultado se requiere del siguiente lema.

Lema 13.38 *Sea* A *una matriz de* $n\times n$ *y supongamos que* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *son todos los* k *eigenvalores distintos de* A *(*$k \leq n$*). Entonces si* $i \neq j$

$$\mathrm{Ei}\, g(\lambda_i) \cap \mathrm{Ei}\, g(\lambda_j) = \{\vec{0}\},$$

es decir, los eigenespacios no tienen elementos comunes distintos del vector nulo.

Demostración. Sea $\vec{v} \in \mathrm{Ei}\, g(\lambda_i) \cap \mathrm{Ei}\, g(\lambda_j)$, entonces se cumple

$$A\vec{v} = \lambda_i \vec{v} = \lambda_j \vec{v},$$

por lo tanto

$$(\lambda_i - \lambda_j)\vec{v} = \vec{0}$$

y como $\lambda_i \neq \lambda_j$ necesariamente $\vec{v} = \vec{0}$. ■

Proposición 13.39 *Sea A una matriz de $n \times n$ y supongamos que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son todos los k eigenvalores distintos de A ($k \leq n$). Entonces $\mathbb{R}^n$ tiene una base que consiste de eigenvectores de A syss*

$$dim(\mathrm{Ei}\, g(\lambda_1)) + dim(\mathrm{Ei}\, g(\lambda_2)) + \cdots + dim(\mathrm{Ei}\, g(\lambda_k)) = n.$$

En tal caso, si $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ son, respectivamente, bases para los eigenespacios $\mathrm{Ei}\, g(\lambda_1)$, $\mathrm{Ei}\, g(\lambda_2), \ldots,$ $\mathrm{Ei}\, g(\lambda_k)$, entonces

$$\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_k$$

es una base para $\mathbb{R}^n$ que consiste de eigenvectores de A.

Demostración. Sean $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$, respectivamente, bases para $\mathrm{Ei}\, g(\lambda_1), \mathrm{Ei}\, g(\lambda_2), \ldots$ y sea n_i el número de elementos de la base $\mathcal{B}_i$. Por el Lema 13.38 estas bases no tienen elementos en común y, por lo tanto, el conjunto

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_k$$

posee exactamente $n_1 + n_2 + \cdots + n_k$ elementos. Probaremos ahora que $\mathcal{B}$ es LI. En efecto, si $\mathcal{B}_i = \{v_{i1}, v_{i2}, \ldots, v_{in_i}\}$, sea

$$\sum_{j=1}^{n_1} a_{1j} v_{1j} + \sum_{j=1}^{n_2} a_{2j} v_{2j} + \cdots + \sum_{j=1}^{n_k} a_{kj} v_{kj} = \vec{0}$$

una combinación lineal de elementos de $\mathcal{B}$ igual al vector nulo (siempre puede agruparse de esta forma). Como

$$\sum_{j=1}^{n_i} a_{ij} v_{ij} \in \mathcal{B}_i \subset \mathrm{Ei}\, g(\lambda_i),$$

el Corolario 13.37 nos dice que

$$\sum_{j=1}^{n_i} a_{ij} v_{ij} = \vec{0},\ i = 1, 2, \ldots, k.$$

Como $\mathcal{B}_i$ es LI, debe tenerse que $a_{ij} = 0$ para todo $i = 1, 2, \ldots, k$, y $j = 1, 2, \ldots, n_i$, por lo que $\mathcal{B}$ es LI.

Dado que $n_i = \dim \mathrm{Ei}\, g(\lambda_i)$ se sigue que $\mathbb{R}^n$ tiene una base que consiste de eigenvectores de A syss

$$dim(\mathrm{Ei}\, g(\lambda_1)) + dim(\mathrm{Ei}\, g(\lambda_2)) + \cdots + dim(\mathrm{Ei}\, g(\lambda_k)) = n.$$

■

Ejemplo 13.40 *En el Ejemplo 13.14 vimos que la matriz*

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

tiene dos eigenvalores: $\lambda_1 = 0$ *y* $\lambda_2 = 1$. *También notamos que los conjuntos*

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} \; y \; \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

son, respectivamente, bases para $\mathrm{Ei}\,g(\lambda_1)$ *y* $\mathrm{Ei}\,g(\lambda_2)$. *Por lo tanto, la Proposición 13.39 asegura que el conjunto*

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathbb{R}^3$ *que consiste de eigenvectores de* B.

En este punto es importante recordar que una matriz cuadrada es **diagonal** syss todas sus componentes fuera de la diagonal principal son cero. Dada la simplicidad inherente de las matrices diagonales es importante caracterizar cuando una matriz puede "convertirse" en una matriz diagonal en el siguiente sentido.

Definición 13.41 *Sea* $A \in M_{n\times n}$. *Decimos que* A *es* ***diagonalizable*** *syss existe una matriz diagonal* $D \in M_{n\times n}$ *y una matriz invertible* P *tal que*

$$D = P^{-1}AP.$$

Ejemplo 13.42 *Si* B *es una matriz diagonal, entonces* B *es diagonalizable pues* $B = I^{-1}BI$.

Antes de enunciar el resultado principal de esta sección es conveniente recordar el siguiente hecho, inmediato a partir de la definición de producto de una matriz por un vector.

Lema 13.43 *Sean* $B \in M_{m\times n}$ *y* $\{\vec{e}_1, \ldots, \vec{e}_n\}$ *la base estándar de* $\mathbb{R}^n$, *entonces*

$$B\vec{e}_1 = 1B^1 + 0B^2 + \ldots + 0B^n = B^1$$

$$\vdots$$

$$B\vec{e}_n = 0B^1 + 0B^2 + \ldots + 1B^n = B^n.$$

Teorema 13.44 *Sea $A \in M_{n\times n}$. Entonces A es diagonalizable syss existe una base para $\mathbb{R}^n$ consistente de eigenvectores de A.*

Demostración. Por definición, A es diagonalizable syss existen una matriz invertible $P \in M_{n\times n}$ y una matriz diagonal $D \in M_{n\times n}$ tales que $D = P^{-1}AP$, equivalentemente si $AP = PD$. Sea

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix},$$

en donde las λ_i no son necesariamente distintas, y notemos que las columnas de D pueden expresarse como

$$D^j = \lambda_j \vec{e}_j, \quad j = 1, 2, \dots, n.$$

Se tiene entonces, que A es diagonalizable syss se cumple

$$\begin{aligned} AP &= \left[AP^1 \ AP^2 \cdots AP^n\right] = PD \\ &= \left[PD^1 \ PD^2 \cdots PD^n\right] = [P\lambda_1\vec{e}_1 \ \ P\lambda_2\vec{e}_2 \ \cdots \ P\lambda_n\vec{e}_n] \\ &= [\lambda_1 P\vec{e}_1 \ \ \lambda_2 P\vec{e}_2 \ \cdots \ \lambda_n P\vec{e}_n] = \left[\lambda_1 P^1 \ \ \lambda_2 P^2 \ \cdots \ \lambda_n P^n\right], \end{aligned}$$

en donde la última igualdad se sigue del Lema 13.43. De lo anterior se tiene que, para toda $j = 1, 2, \dots, n$ se cumple

$$AP^j = \lambda_j P^j,$$

o lo que es lo mismo, P^j es un eigenvector con eigenvalor λ_j. Como la matriz P es invertible, sus columnas son LI y el conjunto $\{P^1, \dots, P^n\}$ es una base para $\mathbb{R}^n$. Concluimos que A es diagonalizable syss

$$\{P^1, \dots, P^n\}$$

es una base para $\mathbb{R}^n$ que consiste de eigenvectores de A. Más aún, si P es la matriz cuya j-ésima columna es P^j y D es la matriz diagonal tal que la $j-$*ésima* entrada de su diagonal principal es el eigenvalor correspondiente a P^j entonces $D = P^{-1}AP$. ■

Observación 13.45 *Dada una base de eigenvectores, Las matrices P y D son únicas, salvo por el orden de las columnas en P y de las entradas diagonales en D.*

Ejemplo 13.46 *En el Ejemplo 13.36 notamos que la matriz*

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

es tal que el conjunto

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathbb{R}^2$ *consistente de eigenvectores de* A*. El Teorema 13.44 afirma que* A *es diagonalizable. Más aún, gracias al texto final de la demostración de dicho teorema, para encontrar* P *y* D *tales que* D *es diagonal y* $D = P^{-1}AP$ *se procede como sigue: definimos a* P *como la matriz cuyas columnas son los elementos de la base de eigenvectores de* A*, concretamente,*

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Posteriormente, como los eigenvalores de $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *y* $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *son* 1 *y* -1*, respectivamente, definimos a la matriz diagonal como*

$$D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Es fácil ver que el inverso P^{-1} *está dado por*

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

y se cumple

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

Otra posibilidad es invertir el orden de los eigenvectores definiendo

$$\hat{P} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

y, consecuentemente (cuidando de cambiar también el orden de los eigenvalores),

$$\hat{D} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

En este caso, el lector puede verificar que $\hat{D} = \hat{P}^{-1}A\hat{P}$*.*

Observación 13.47 *En el ejemplo anterior podríamos haber tomado otra base para los eigenespacios, por ejemplo*

$$\left\{ \begin{bmatrix} 7 \\ 7 \end{bmatrix} \right\}$$

para $\mathrm{Ei}\, g(\lambda = 1)$ *y*

$$\left\{ \begin{bmatrix} -666 \\ 666 \end{bmatrix} \right\}$$

para $\mathrm{Ei}\, g(\lambda = -1)$. *La base de* $\mathbb{R}^2$ *estaría dada por*

$$\left\{ \begin{bmatrix} 7 \\ 7 \end{bmatrix}, \begin{bmatrix} -666 \\ 666 \end{bmatrix} \right\}$$

de manera que la matriz invertible estaría dada por

$$\tilde{P} = \begin{bmatrix} 7 & -666 \\ 7 & 666 \end{bmatrix}$$

y la matriz diagonal sería, una vez más

$$D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Ejemplo 13.48 *En el Ejemplo 13.40 notamos que la matriz*

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

es tal que el conjunto

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

es una base para $\mathbb{R}^3$ *que consiste de eigenvectores de* B. *El Teorema 13.44 afirma que* B *es diagonalizable. Así,*

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

y D *es la matriz diagonal de los eigenvalores correspondientes a* P^1, P^2 *y* P^3 *dada como*

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Dado que $P^2 = I$, *se tiene que* $P^{-1} = P$ *y se cumple*

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

Al igual que en el ejemplo anterior, existen varias posibles elecciones para encontrar las matrices D, diagonal, y P, invertible, ya que puede cambiarse el orden de los eigenvectores (por lo tanto cambia el orden de los eigenvalores). Por ejemplo si

$$\hat{P} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \text{ junto con } \hat{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

se tiene que $\hat{D} = \hat{P}^{-1} A \hat{P}$.

Ejemplo 13.49 *Observamos que la matriz*

$$C = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

cuyo polinomio característico es

$$p(\lambda) = (\lambda - 1)^2(\lambda - 4),$$

tiene dos eigenvalores: $\lambda_0 = 4$ y $\lambda_1 = 1$. También notamos que los conjuntos

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} \; y \; \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

son respectivamente bases para $\mathrm{Ei}\, g(\lambda_0)$ y $\mathrm{Ei}\, g(\lambda_1)$. Gracias al Corolario 13.39 y dado que

$$dim(\mathrm{Ei}\, g(\lambda_0)) + dim(\mathrm{Ei}\, g(\lambda_1)) = 1 + 1 = 2 < 3,$$

tenemos que no existe una base para $\mathbb{R}^3$ consistente de eigenvectores de C. Este hecho, junto con el Teorema 13.44 implican que C no es diagonalizable.

Observación 13.50 *El Ejemplo 13.49 nos muestra una matriz que no puede diagonalizarse. La razón es que el eigenvalor $\lambda_1 = 1$ es una raiz doble (de multiplicidad 2) del polinomio caractrístico, pero el eigenespacio $\mathrm{Ei}\, g(\lambda_1)$ sólo tiene dimensión igual a 1. En consecuencia, no se tienen suficientes eigenvectores de C para formar una base de $\mathbb{R}^3$.*

Si consideramos a la matriz

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

dada en la Observación 13.21, su polinomio característico dado por $p(\lambda) = \lambda^2 + 1$ no tiene raices reales. Evidentemente A tampoco es diagonalizable puesto que no tiene ningún eigenvector en $\mathbb{R}^2$. Sin embargo, como se mencionó anteriormente, este polinomio si tiene raices en el campo de los números complejos ($\mathbb{C}$). Concretamente, los eigenvalores son i y $-i$ y es posible encontrar dos eigenvectores con coordenadas en $\mathbb{C}$ (vectores en el espacio vectorial $\mathbb{C}^2$). En consecuencia, la matriz A si podrá diagonalizarse sobre los números complejos.

Finalizamos esta sección con la siguiente proposición.

Proposición 13.51 *Si A, P y D son matrices de $n \times n$ tales que $D = P^{-1}AP$, entonces $det(D) = det(A)$.*

Demostración. Como el determinante preserva productos y $\det(P^{-1}) = \frac{1}{\det(P(}$ se tiene,

$$\begin{aligned} \det(D) &= \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) \\ &= \det(P^{-1})\det(P)\det(A) = \det(A). \end{aligned}$$

■

Ejemplo 13.52 *Sea*

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

cuyo polinomio característico está dado por

$$p_A(\lambda) = -(\lambda - 1)(\lambda - 3)(\lambda + 2)$$

de manera que los eigenvalores son $\lambda_1 = 1$, $\lambda_2 = 3$ y $\lambda_3 = -2$. Sin necesidad de encontrar los eigenespacios correspondientes, sabemos que, dado que A es equivalente a la matriz diagonal

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

entonces $\det(A) = \det(D) = -6$.

Capítulo 14

Longitud y Dirección

Introducción

Dado un vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$, podemos visualizarlo en el plano como en la Figura 14.1:

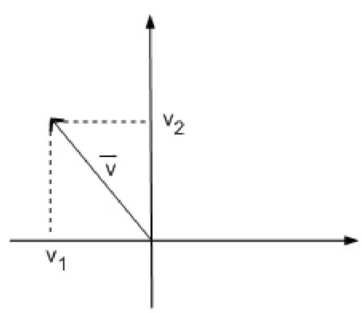

Figura 14.1: Vector $\vec{v}$ en $\mathbb{R}^2$.

Utilizando el Teorema de Pitágoras, resulta natural definir a la **longitud** (o **norma**) de $\vec{v}$, denotada por $||\vec{v}||$, como

$$||\vec{v}|| = \sqrt{v_1^2 + v_2^2}.$$

Por lo tanto,

$$||\vec{v}||^2 = v_1^2 + v_2^2.$$

De esta forma, $||\vec{v}||$ coincide con el "tamaño" del vector $\vec{v}$.

Observación 14.1 *Si $\vec{v} \in \mathbb{R}$ (en cuyo caso el vector también es un escalar), la definición de norma coincide con el valor absoluto de $\vec{v}$, en efecto,*

$$||\vec{v}|| = \sqrt{\vec{v}^2} = |\vec{v}|\,.$$

Dados dos puntos A y B en el plano, correpondientes a los puntos finales de los vectores $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ y $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ en $\mathbb{R}^2$, puede definirse fácilmente la distancia entre A y B como la longitud del segmento que conecta a los puntos, que no es más que el vector $\vec{x} - \vec{y}$ (o alternativamente $\vec{y} - \vec{x}$). Por lo tanto, la distancia entre A y B es igual a

$$\begin{aligned} ||\vec{x} - \vec{y}|| &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\ &= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} = ||\vec{y} - \vec{x}||. \end{aligned}$$

En particular, la distancia entre A y B es igual a la distancia entre B y A. Las consideraciones anteriores se ilustran en la Figura 14.2.

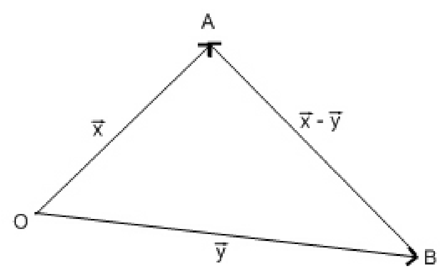

Figura 14.2: Distancia entre dos puntos

Ahora generalizaremos estas nociones geométricas en el plano al contexto de $\mathbb{R}^n$. Recordemos que el producto punto entre los vectores

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},\ \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$$

está dado por

$$\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i,$$

o si pensamos a los vectores como matrices, $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$. Es inmediato comprobar que si $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, entonces

$$\vec{x} \cdot \vec{x} = x_1 x_1 + x_2 x_2 = x_1^2 + x_2^2 = ||\vec{x}||^2.$$

Esto nos sugiere una forma de generalizar la definición de norma y distancia en $\mathbb{R}^n$ como sigue.

Definición 14.2 *Sean $\vec{u}$ y $\vec{v}$ vectores de $\mathbb{R}^n$.*

- *La **longitud** o **norma** de $\vec{u}$ se define como*
$$||\vec{u}|| = \sqrt{\vec{u}\cdot\vec{u}}.$$
- *Si A y B son los puntos determinados por los vectores $\vec{u}$ y $\vec{v}$, la **distancia** entre A y B se define como*
$$d(A,B) = ||\vec{v}-\vec{u}||.$$

Ejemplo 14.3 *Sean*
$$\vec{u} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} \text{ y } \vec{v} = \begin{bmatrix} -1 \\ -5 \\ -3 \\ 0 \end{bmatrix},$$
entonces
$$\vec{u}\cdot\vec{u} = 1^2 + 2^2 + (-3)^2 + 4^2 = 30$$
y se tiene
$$||\vec{u}|| = \sqrt{\vec{u}\cdot\vec{u}} = \sqrt{30}.$$
Asimismo, la distancia entre los puntos determinados por $\vec{u}$ y $\vec{v}$ es igual a
$$\begin{aligned} ||\vec{v}-\vec{u}|| &= \left\| \begin{bmatrix} -1 \\ -5 \\ -3 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ -3 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -2 \\ -7 \\ 0 \\ -4 \end{bmatrix} \right\| \\ &= \sqrt{2^2+7^2+4^2} = \sqrt{69}. \end{aligned}$$

Recordemos ahora las siguientes propiedades del producto punto vistas en el Capítulo 3.

Proposición 14.4 *Si $\vec{u}$, $\vec{v}$ y $\vec{w}$ son vectores de $\mathbb{R}^n$ y $\beta\in\mathbb{R}$, entonces*

1. $\vec{u}\cdot\vec{u} = u_1^2 + \ldots u_n^2$ *y por lo tanto, $\vec{u}\cdot\vec{u} = 0$ syss $\vec{u} = \vec{0}$.*
2. $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$.
3. $\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v} + \vec{u}\cdot\vec{w}$.
4. $(\beta\vec{u})\cdot\vec{v} = \beta(\vec{u}\cdot\vec{v}) =.\vec{u}\cdot(\beta\vec{v})$.

Observación 14.5 *Una consecuencia inmediata de los incisos (2) y (4) es que si $\beta,\gamma\in\mathbb{R}$, se tiene que*
$$(\beta\vec{u})\cdot(\gamma\vec{v}) = \beta\gamma(\vec{u}\cdot\vec{v}).$$

El siguiente resultado es inmediato a partir de la proposición anterior.

Corolario 14.6 *Si $\vec{u}$ es un vector de $\mathbb{R}^n$ y $\beta \in \mathbb{R}$, entonces*

- $||\vec{u}|| = \sqrt{\vec{u} \cdot \vec{u}} \geq 0$ *y* $||\vec{u}|| = 0$ *syss* $\vec{u} = \vec{0}$.
- $||\beta\vec{u}|| = \sqrt{(\beta\vec{u}) \cdot (\beta\vec{u})} = \sqrt{\beta^2(\vec{u} \cdot \vec{u})} = \sqrt{\beta^2}\sqrt{\vec{u} \cdot \vec{u}} = (|\beta|)\, ||\vec{u}||$. *En particular,*

  $$||\vec{v} - \vec{u}|| = (|-1|)\, ||\vec{v} - \vec{u}|| = ||(-1)(\vec{v} - \vec{u})|| = ||\vec{u} - \vec{v}||.$$

  Como antes, si A y B son los puntos de $\mathbb{R}^n$ determinados por los vectores $\vec{u}$ y $\vec{v}$, puede definirse la distancia entre ellos como

  $$d(A, B) = ||\vec{v} - \vec{u}|| = ||\vec{u} - \vec{v}|| = d(B, A).$$

De aquí en adelante omitiremos la mención a los puntos A y B.

Ejemplo 14.7

$$\left\| \begin{bmatrix} -1000 \\ -2000 \\ 0 \end{bmatrix} \right\| = \left\| (-1000) \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\| = (|-1000|) \left\| \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\| = 1000\sqrt{5}.$$

Los conceptos de trigonometría en el plano se generalizan fácilmente a cualquier subespacio de dimensión dos en $\mathbb{R}^n$. Por lo pronto, el concepto de norma o tamaño de un vector nos permite medir distancias entre cualquier pareja de puntos. En lo que sigue, se extiende el concepto de la medición de ángulos en $\mathbb{R}^n$. Para este efecto, primero generalizamos el concepto de ortogonalidad o ángulos rectos entre vectores.

Ortogonalidad

En la trigonometría clásica, el teorema de Pitágoras es, quizás, el resultado más importante en relación a los ángulos rectos. En la figura 14.3 vemos el triángulo rectángulo formado por los vectores $\vec{x}, \vec{y}$ y $\vec{x} - \vec{y}$.

La siguiente definición de ortogonalidad forza el cumplimiento del teorema de Pitágoras cuando se tiene esta configuración geométrica entre los vectores que forman el triángulo.

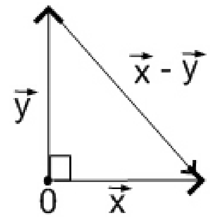

Figura 14-1 Figura 14.3: Triángulo rectángulo

Definición 14.8 *Sean $\vec{x}, \vec{y} \in \mathbb{R}^n$, decimos que estos vectores son **ortogonales** o **perpendiculares** syss*

$$\vec{x} \cdot \vec{y} = 0.$$

El siguiente resultado se sigue de la definición de la norma de un vector y de las propiedades del producto punto.

Proposición 14.9 *$\vec{x}, \vec{y} \in \mathbb{R}^n$ son ortogonales syss*

$$\|\vec{x} - \vec{y}\|^2 = \|x\|^2 + \|y\|^2 .$$

Demostración.

$$\begin{aligned} \|\vec{x} - \vec{y}\|^2 &= (\vec{x} - \vec{y}) \cdot (\vec{x} - \vec{y}) \\ &= \vec{x} \cdot \vec{x} - \vec{x} \cdot \vec{y} - \vec{y} \cdot \vec{x} + \vec{y} \cdot \vec{y} \\ &= \|x\|^2 - 2\vec{x} \cdot \vec{y} + \|y\|^2 , \end{aligned}$$

por lo tanto, los vectores cumplen la Definición 14.8 syss

$$\|\vec{x} - \vec{y}\|^2 = \|x\|^2 + \|y\|^2 .$$

■

Notación 14.1 *Dados $\vec{x}, \vec{y} \in \mathbb{R}^n$ se utiliza la notación*

$$\vec{x} \perp \vec{y}$$

para indicar que $\vec{x}$ y $\vec{y}$ son vectores ortogonales.

Ejemplo 14.10 *Los vectores*

$$\vec{x} = \begin{bmatrix} -1 \\ 2 \\ -3 \\ \pi \end{bmatrix} \quad y \quad \vec{y} = \begin{bmatrix} \pi \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

en $\mathbb{R}^4$ *son ortogonales. En efecto, se tiene que*

$$\begin{bmatrix} -1 \\ 2 \\ -3 \\ \pi \end{bmatrix} \cdot \begin{bmatrix} \pi \\ 3 \\ 2 \\ 1 \end{bmatrix} = -\pi + 6 - 6 + \pi = 0$$

y, por lo tanto $\vec{x} \perp \vec{y}$.

Tomemos, una vez más, $\vec{x}, \vec{y} \in \mathbb{R}^n$. Considérese $S = span(\vec{x})$, que no es más que la recta generada por $\vec{x}$. Todos los elementos de esta recta son de la forma $r\vec{x}$ para algún escalar r. Nos gustaría encontrar un vector $r\vec{x}$ en S de tal suerte que los puntos determinados por $\vec{y}$ y por $r\vec{x}$ estuviesen lo más cerca posible. En otras palabras, se quiere encontrar una $r \in \mathbb{R}$, tal que la norma

$$\|\vec{y} - r\vec{x}\|$$

sea lo más pequeña posible. Este problema se ilustra en la Figura 14.4.

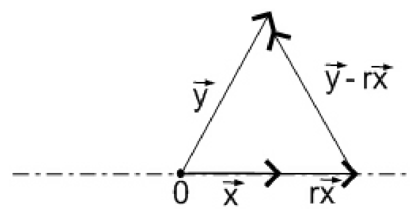

Figura 14-2 Figura 14.4: Vector $\vec{y} - r\vec{x}$.

El problema anterior puede resolverse fácilmente notando que minimizar $\|\vec{y} - r\vec{x}\|$ es equivalente a minimizar $\|\vec{y} - r\vec{x}\|^2$. Se utiliza la definición de la norma junto con las propiedades del producto punto para obtener

$$\begin{aligned} \|\vec{y} - r\vec{x}\|^2 &= (\vec{y} - r\vec{x}) \cdot (\vec{y} - r\vec{x}) \\ &= \vec{y} \cdot \vec{y} - \vec{y} \cdot (r\vec{x}) - (r\vec{x}) \cdot \vec{y} + (r\vec{x}) \cdot (r\vec{x}) \\ &= \|\vec{y}\|^2 - 2r(\vec{x} \cdot \vec{y}) + r^2 \|\vec{x}\|^2 . \end{aligned}$$

Dado que $\|\vec{y} - r\vec{x}\|^2$ es diferenciable como función de r, simplemente derivamos con respecto a r e igualamos a cero para obtener el valor de r que minimiza esta función, dado por

$$r = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}. \tag{14.1}$$

De la discusión anterior se concluye lo siguiente.

Proposición 14.11 *Dados $\vec{x}, \vec{y} \in \mathbb{R}^n$, el punto sobre la recta generada por $\vec{x}$ que está más cercano a $\vec{y}$ es el determinado por el vector*

$$\left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right) \vec{x}.$$

La intuición geométrica nos debe (o debiera) hacer sospechar que los vectores $\vec{x}$ y $\vec{y} - r\vec{x}$ resultan ser ortogonales como se ilustra en la Figura 14.5.

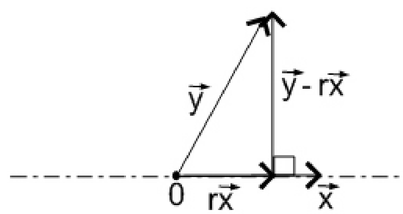

Figura 14-3 Figura 14.5: Vectores ortogonales

A continuación vemos que nuestra intuición es correcta.

Proposición 14.12 *Sean $\vec{x}, \vec{y} \in \mathbb{R}^n$, si $r = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2} \in \mathbb{R}$, entonces*

$$\vec{x} \cdot (\vec{y} - r\vec{x}) = 0.$$

Demostración. Utilizando las propiedades del producto punto se obtiene

$$\begin{aligned} \vec{x} \cdot (\vec{y} - r\vec{x}) &= \vec{x} \cdot \vec{y} - \vec{x} \cdot (r\vec{x}) \\ &= \vec{x} \cdot \vec{y} - r(\vec{x} \cdot \vec{x}) = \vec{x} \cdot \vec{y} - r \|\vec{x}\|^2 . \end{aligned}$$

Sustituyendo el valor de r llegamos a

$$\vec{x} \cdot (\vec{y} - r\vec{x}) = \vec{x} \cdot \vec{y} - \left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right) \|\vec{x}\|^2 = 0.$$

■

Ejemplo 14.13 *Sean*

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

y observemos que

$$\vec{x} \cdot \vec{y} = 1, \ \ \|\vec{x}\| = \sqrt{1} \ \ \text{y} \ \ \|\vec{y}\| = 1.$$

Deseamos encontrar $r \in \mathbb{R}$ tal que la distancia $\|\vec{y} - r\vec{x}\|$ sea lo más pequeña posible. Por la Proposición 14.11 sabemos que

$$r = \left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right) = \frac{1}{2},$$

de manera que el punto más cercano a $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *sobre la recta determinada por el vector* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *es el* $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$. *Adicionalmente,*

$$\vec{y} - r\vec{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

es un vector ortogonal a $\vec{x}$ puesto que

$$\vec{x} \cdot (\vec{y} - r\vec{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = -\frac{1}{2} + \frac{1}{2} = 0.$$

Dos desigualdades

A continuación probamos dos resultados clásicos para la norma y el producto punto.

Proposición 14.14 (Desigualdad de Bunyakovsky-Cauchy-Schwarz) *Si $\vec{x}$ y $\vec{y}$ son vectores de $\mathbb{R}^n$, entonces*

$$|\vec{x} \cdot \vec{y}| \leq \|\vec{x}\| \, \|\vec{y}\| \, .$$

Demostración. Para cualquier escalar r se tiene que

$$\|\vec{y} - r\vec{x}\| \geq 0.$$

Por lo tanto

$$\begin{aligned} \|\vec{y} - r\vec{x}\|^2 &= (\vec{y} - r\vec{x}) \cdot (\vec{y} - r\vec{x}) \\ &= \vec{y} \cdot \vec{y} - \vec{y} \cdot (r\vec{x}) - (r\vec{x}) \cdot \vec{y} + (r\vec{x}) \cdot (r\vec{x}) \\ &= \|\vec{y}\|^2 - 2r(\vec{x} \cdot \vec{y}) + r^2 \|\vec{x}\|^2 \geq 0. \end{aligned}$$

En particular, si $r = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}$ se tiene

$$\|\vec{y}\|^2 - 2\left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right)(\vec{x} \cdot \vec{y}) + \left(\frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|^2}\right)^2 \|\vec{x}\|^2 \geq 0,$$

$$\text{o bien } \|\vec{y}\|^2 - \frac{(\vec{x} \cdot \vec{y})^2}{\|\vec{x}\|^2} \geq 0.$$

De esta última desigualdad obtenemos,

$$\|\vec{x}\|^2 \|\vec{y}\|^2 \geq (\vec{x}\cdot\vec{y})^2$$

y, aplicando la raiz cuadrada en ambos lados de la desigualdad, se tiene

$$|\vec{x}\cdot\vec{y}| \leq \|\vec{x}\| \|\vec{y}\| .$$

■

Proposición 14.15 (Desigualdad del Triángulo) *Si $\vec{x}$ y $\vec{y}$ son vectores de $\mathbb{R}^n$, entonces*

$$||\vec{x}+\vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||.$$

Demostración. Utilizando la definición de la norma, las propiedades del producto punto y la Proposición 14.14 tenemos

$$\begin{aligned} \|\vec{x}+\vec{y}\|^2 &= (\vec{x}+\vec{y})\cdot(\vec{x}+\vec{y}) \\ &= \vec{x}\cdot\vec{x} + 2(\vec{x}\cdot\vec{y}) + \vec{y}\cdot\vec{y} = \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2(\vec{x}\cdot\vec{y}) \\ &\leq \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2|\vec{x}\cdot\vec{y}| \leq \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\|\vec{x}\|\|\vec{y}\| \\ &= (\|\vec{x}\| + \|\vec{y}\|)^2 . \end{aligned}$$

Tomando la raiz cuadrada de ambos lados de la desigualdad concluimos que

$$\|\vec{x}+\vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\| .$$

■

La interpretación geométrica de la desigualdad del triángulo es que en un triángulo, la suma de las longitudes de dos de sus lados es mayor o igual a la longitud del tercer lado. Esto se muestra en la Figura 14.6.

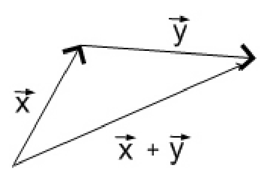

Figura 14.6: Desigualdad del triángulo.

Medición de ángulos

Dados $\vec{x}, \vec{y} \in \mathbb{R}^n$ queremos medir el ángulo entre estos vectores. Para este efecto, examinando los dos diagramas de la Figura 14.7, vemos que, utilizando el triángulo rectángulo formado por los vectores $r\vec{x}, \vec{y}$ y $\vec{y}-r\vec{x}$

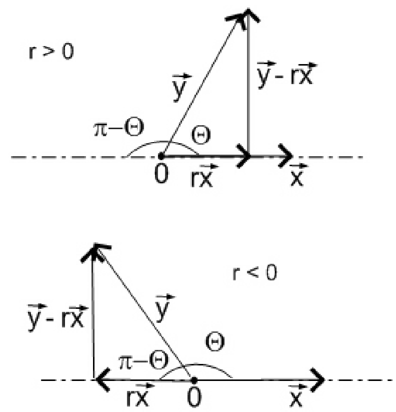

Figura 14-4 Figura 14.7: $\cos\Theta = -\cos(\pi - \Theta)$

(en donde $r = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|^2}$), es posible definir las funciones trigonométricas usuales.

En nuestro caso, definimos el coseno del ángulo Θ sustentado entre los vectores $r\vec{x}$ y $\vec{y}$ en la forma usual, es decir

$$\cos\Theta = \frac{\text{longitud dirigida de } r\vec{x}}{\text{longitud de } \vec{y}} = \frac{r\,(\text{norma de } \vec{x})}{\text{norma de } \vec{y}} = \frac{r\,\|\vec{x}\|}{\|\,\vec{y}\|}.$$

La **longitud dirigida** $r\,\|\vec{x}\|$ toma en cuenta el signo del escalar r, lo cual es relevante para determinar el signo de la función coseno, es decir, si el ángulo Θ es agudo u obtuso. Sustituyendo el valor $r = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|^2}$ (que determina el triángulo rectángulo), obtenemos que $\cos\Theta$ queda definido como

$$\cos\Theta = \left(\frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|^2}\right)\frac{\|\vec{x}\|}{\|\,\vec{y}\|} = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\,\|\,\vec{y}\|}. \tag{14.2}$$

Observemos que la desigualdad dada en la Proposición 14.14 nos garantiza que

$$-1 \leq \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\,\|\,\vec{y}\|} \leq 1$$

y, por lo tanto, $\cos\Theta \in [-1, 1]$.

La definición de $\cos\Theta$ dada por la Expresión (14.2) nos proporciona la siguiente definición alternativa del producto punto como sigue.

$$\vec{x}\cdot\vec{y} = \|\vec{x}\|\,\|\vec{y}\|\cos\Theta. \tag{14.3}$$

Ejemplo 14.16 *Dados*

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad y \quad \vec{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \in \mathbb{R}^5,$$

podemos calcular el ángulo Θ *entre ellos mediante (14.2). Debido a que*

$$\|\vec{x}\| = \sqrt{2}, \ \|\vec{y}\| = 1 \ y \ \ \vec{x} \cdot \vec{y} = -1,$$

se tiene que

$$\cos \Theta = \frac{-1}{\sqrt{2} \times 1} = \frac{-1}{\sqrt{2}}$$

y $\Theta = \frac{3\pi}{4}$.

Paralelismo y vectores unitarios

Recordando que

$$\cos(0) = 1, \ \ \cos \frac{\pi}{2} = 0, \ \ \cos \pi = -1,$$

las siguientes observaciones son inmediatas de la interpretación geométrica del producto punto dada por (14.3).

- $\vec{x}$ y $\vec{y}$ son ortogonales (perpendiculares) syss

$$\cos \Theta = 0 \ \ syss \ \ \vec{x} \cdot \vec{y} = 0.$$

- $\vec{x}$ y $\vec{y}$ son paralelos syss $\cos \Theta = \pm 1$ syss

$$\vec{x} \cdot \vec{y} = \pm ||\vec{x}|| ||\vec{y}||.$$

De lo anterior, resulta natural generalizar la noción de paralelismo como sigue.

Definición 14.17 *Sean* $\vec{x}$ *y* $\vec{y}$ *vectores de* $\mathbb{R}^n$. *Decimos que* $\vec{x}$ *y* $\vec{y}$ *son paralelos syss*

$$\vec{x} \cdot \vec{y} = \pm ||\vec{x}|| || ||\vec{y}||.$$

Ejemplo 14.18 *Los vectores* $\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *y* $\vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *son ortogonales ya que* $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$. *Análogamente, en el contexto de* $\mathbb{R}^n$, *los vectores* $\vec{e}_1, \vec{e}_2, \ldots \vec{e}_n$ *son* ***ortogonales dos a dos*** *ya que el producto punto de cualesquiera dos vectores distintos es igual a cero.*

Ejemplo 14.19 *Los vectores* $\vec{u} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ *y* $\vec{v} = \begin{bmatrix} -2 \\ -10 \end{bmatrix}$ *son paralelos pues* $\vec{u} \cdot \vec{v} = -52$ *y* $||\vec{u}||||\vec{v}|| = \sqrt{26}\sqrt{104} = \sqrt{2704} = 52.$

Una forma mas simple de comprobar que los vectores $\vec{u}$ y $\vec{v}$ son paralelos es apelando al siguiente resultado y al hecho de que $\vec{v} = -2\vec{u}$.

Proposición 14.20 *Sea* $\vec{u} \in \mathbb{R}^n$ *y sea* $c \in \mathbb{R}$ *con* $c \neq 0$. *entonces* $\vec{u}$ *y* $c\vec{u}$ *son paralelos.*

Demostración. Utilizando (14.3) y las propiedades del producto punto, se tiene que si θ es el ángulo entre $\vec{u}$ y $c\vec{u}$, entonces

$$\begin{aligned} \cos\theta &= \frac{\vec{u} \cdot c\vec{u}}{\|\vec{u}\| \, \|c\vec{u}\|} = \frac{c\,(\vec{u} \cdot \vec{u})}{\|\vec{u}\| \, |c| \, \|\vec{u}\|} = \frac{c\,\|\vec{u}\|^2}{|c| \, \|\vec{u}\|^2} \\ &= \frac{c}{|c|} = \begin{cases} 1 \text{ si } c > 0, \\ -1 \text{ si } c < 0, \end{cases} \end{aligned}$$

con lo cual queda demostrada la proposición. ■

Definición 14.21 *Sea* $\vec{u} \in \mathbb{R}^n$. *Decimos que* $\vec{u}$ *es un vector unitario syss* $||\vec{u}|| = 1$.

Ejemplo 14.22 *Los vectores* $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *y* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *son claramente unitarios. También el vector* $\frac{\sqrt{2}}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *es unitario pues, de acuerdo al Corolario 14.6, su norma es igual a* $\frac{\sqrt{2}}{2}\sqrt{2} = 1$.

La siguiente proposición, cuya demostración se deja como ejercicio para el lector, dice que todo vector distinto de cero tiene asociado, de forma natural, un vector unitario.

Proposición 14.23 *Si* $\vec{v} \in \mathbb{R}^n$ *es un vector distinto de cero, entonces el vector* $\frac{1}{||\vec{v}||}\vec{v}$ *es unitario y satisface que* $\vec{v}$ *y* $\frac{1}{||\vec{v}||}\vec{v}$ *son paralelos en la misma dirección. Al vector* $\frac{1}{||\vec{v}||}\vec{v}$ *se le conoce como la* ***normalización*** *de* $\vec{v}$ *y se ilustra en la Figura 14.8 para el caso* $||\vec{v}|| > 1$

Ejemplo 14.24 *Dado el vector*

$$\vec{v} = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix},$$

como

$$||\vec{v}|| = \sqrt{9 + 16} = 5,$$

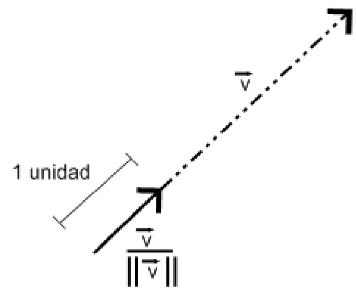

Figura 14-5 Figura 14.8: Normalización de un vector.

su normalización es el vector,

$$\begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \\ 0 \end{bmatrix}.$$

Observación 14.25 *Observemos que la Expresión 14.2 para* $\cos\Theta$ *puede reescribirse como*

$$\cos\Theta = \frac{1}{\|\vec{x}\|}\vec{x}\cdot\frac{1}{\|\,\vec{y}\|}\vec{y},$$

es decir, como el producto punto de los vectores unitarios $\frac{\vec{x}}{\|\vec{x}\|}$ *y* $\frac{\vec{y}}{\|\,\vec{y}\|}$.

Conjuntos y bases ortogonales

Definición 14.26 *Sea C un subconjunto de vectores de $\mathbb{R}^n$. Decimos que C es un **conjunto ortogonal** syss sus elementos son ortogonales dos a dos, es decir, para cualesquiera dos vectores distintos $\vec{u}$ y $\vec{v}$ en C se cumple que $\vec{u}\cdot\vec{v}=\vec{0}$.*

Como ya habíamos notado previamente en el Ejemplo 14.18, la base estándar $\{\vec{e}_1,\ldots,\vec{e}_n\}$ de $\mathbb{R}^n$ es un conjunto ortogonal. En el siguiente ejemplo analizaremos otros conjuntos ortogonales más generales.

Ejemplo 14.27 *Si*

$$\vec{u}_1=\begin{bmatrix}1\\1\\0\end{bmatrix},\vec{u}_2=\begin{bmatrix}-1\\1\\0\end{bmatrix},\vec{u}_3=\begin{bmatrix}0\\0\\1\end{bmatrix}\ y\ \vec{u}_4=\begin{bmatrix}1\\1\\1\end{bmatrix},$$

entonces el conjunto $\{\vec{u}_1,\vec{u}_2,\vec{u}_3\}$ es ortogonal pues

$$\vec{u}_1\cdot\vec{u}_2=\vec{u}_1\cdot\vec{u}_3=\vec{u}_2\cdot\vec{u}_3=0.$$

Pero, el conjunto $\{\vec{u}_1,\vec{u}_2,\vec{u}_4\}$ no es ortogonal pues $\vec{u}_1\cdot\vec{u}_4=2\neq 0$.

Proposición 14.28 *Si* $\{\vec{v}_1, \ldots, \vec{v}_m\}$ *es un subconjunto ortogonal de* $\mathbb{R}^n$ *cuyos elementos no son el vector cero, entonces se cumplen las siguientes afirmaciones.*

(a) $\{\vec{v}_1, \ldots, \vec{v}_m\}$ *es LI.*

(b) Si $\vec{v} \in span\{\vec{v}_1, \ldots, \vec{v}_m\}$, *entonces*

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m$$

donde

$$c_1 = \frac{\vec{v}\cdot\vec{v}_1}{||\vec{v}_1||^2},\ c_2 = \frac{\vec{v}\cdot\vec{v}_2}{||\vec{v}_2||^2}, \ldots,\ c_m = \frac{\vec{v}\cdot\vec{v}_m}{||\vec{v}_m||^2},$$

es decir, de acuerdo a la Proposición 14.12, los escalares c_i *son tales que los vectores* $\vec{v} - c_i\vec{v}_i$ *son ortogonales a* $\vec{v}_i$ *y* $c_i\vec{v}_i$ *es la "componente de* $\vec{v}$ *a lo largo de* $\vec{v}_i$*", noción que vamos a precisar más adelante. Las consideraciones anteriores se muestran en la Figura 14.9.*

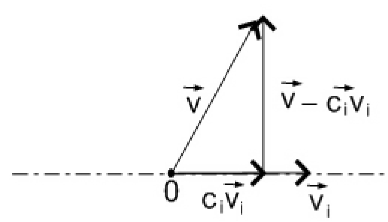

Figura 14-6 Figura 14.9: $\vec{v} = c_i\vec{v}_i + (\vec{v} - c_i\vec{v}_i)$.

Demostración. En el caso de (a) supongamos que

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_m\vec{v}_m = \vec{0}.$$

Debemos probar que todos los α's están forzados a ser cero. Nuestra hipótesis implica que

$$(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_m\vec{v}_m)\cdot\vec{v}_1 = \vec{0}\cdot\vec{v}_1 = 0.$$

y como

$$\vec{v}_1\cdot\vec{v}_2 = \ldots = \vec{v}_1\cdot\vec{v}_m = 0,$$

se cumple

$$\begin{aligned}&(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_m\vec{v}_m)\cdot\vec{v}_1\\ &= \alpha_1\vec{v}_1\cdot\vec{v}_1 + \alpha_2\vec{v}_2\cdot\vec{v}_1 + \ldots + \alpha_m\vec{v}_m\cdot\vec{v}_1 = \alpha_1||\vec{v}_1||^2,\end{aligned}$$

obteniéndose

$$\alpha_1||\vec{v}_1||^2 = 0.$$

Ahora bien, como $||\vec{v}_1||^2 > 0$ (pues $\vec{v}_1$ no es el vector cero), necesariamente $\alpha_1 = 0$. Análogamente puede verificarse que los otros α_i's también son iguales a cero.

Para el inciso (b) notamos que, gracias a lo que acabamos de demostrar en el inciso anterior, $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es una base para $span\{\vec{v}_1, \ldots, \vec{v}_m\}$. Por lo tanto, si $\vec{v} \in span\{\vec{v}_1, \ldots, \vec{v}_m\}$, entonces existen escalares únicos $c_1, c_2, \ldots, c_m$ tales que

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m.$$

Ahora bien, si realizamos el producto punto con $\vec{v}_1$ en ambos lados de esta iguladad, obtenemos

$$\begin{aligned}\vec{v} \cdot \vec{v}_1 &= (c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m) \cdot \vec{v}_1 \\ &= c_1\vec{v}_1 \cdot \vec{v}_1 + 0 + \ldots 0 = c_1||\vec{v}_1||^2,\end{aligned}$$

en donde la penúltima igualdad se sigue del hecho que $\vec{v}_1 \perp \vec{v}_2, \ldots, \vec{v}_1 \perp \vec{v}_m$. De aquí, $c_1 = \frac{\vec{v} \cdot \vec{v}_1}{||\vec{v}_1||^2}$. Análogamente, puede verificarse que para toda $i \in \{1, \ldots, m\}$ se tiene que $c_i = \frac{\vec{v} \cdot \vec{v}_i}{||\vec{v}_i||^2}$. ■

Ejemplo 14.29 *En el Ejemplo 14.27 ya hemos notado que si*

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \ y \ \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

entonces $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es un conjunto ortogonal de vectores distintos de cero. El teorema anterior asegura que $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ es LI y, por lo tanto una base para $\mathbb{R}^3$. Supongamos ahora que deseamos encontrar escalares c_1, c_2 y c_3 tales que si

$$\vec{u} = \begin{bmatrix} 2 \\ -2 \\ 0 \end{bmatrix},$$

entonces

$$\vec{u} = c_1\vec{u}_1 + c_2\vec{u}_2 + c_3\vec{u}_3.$$

Por supuesto que una forma de hacerlo es resolviendo el sistema cuya matriz de coeficientes tiene por columnas a $\vec{u}_1$, $\vec{u}_2$ y $\vec{u}_3$ y cuyo vector de términos independientes es el vector $\vec{u}$. Otra forma de hacerlo es invocando la Proposición 14.28, obteniéndose

$$c_1 = \frac{\vec{u} \cdot \vec{u}_1}{||\vec{u}_1||^2} = 0, \ c_2 = \frac{\vec{u} \cdot \vec{u}_2}{||\vec{u}_2||^2} = -\frac{4}{2} = -2 \ y \ c_3 = \frac{\vec{u} \cdot \vec{u}_3}{||\vec{u}_3||^2} = \frac{1}{1} = 1.$$

Ejemplo 14.30 *Es sencillo verificar que si*

$$\vec{v}_1 = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} -1 \\ 4 \\ 0 \\ 0 \end{bmatrix} \text{ y } \vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \end{bmatrix},$$

entonces

$$\vec{v}_1 \cdot \vec{v}_2 = \vec{v}_1 \cdot \vec{v}_3 = \vec{v}_2 \cdot \vec{v}_3 = 0,$$

por lo que $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ es un conjunto ortogonal (de vectores distintos de cero). La Proposición 14.28 asegura que $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ es LI y, por lo tanto, una base para

$$S = span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}.$$

Por supuesto, S es igual al subespacio que consiste de aquellos vectores de $\mathbb{R}^4$ cuya cuarta coordenada es igual a cero.

Supongamos ahora que deseamos encontrar escalares c_1, c_2 y c_3 tales que si

$$\vec{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \\ 0 \end{bmatrix},$$

entonces $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3$. Una forma de hacerlo es resolviendo el sistema cuya matriz de coeficientes tiene por columnas a $\vec{v}_1$, $\vec{v}_2$ y $\vec{v}_3$ y cuyo vector de términos independientes es el vector $\vec{v}$. Otra forma de hacerlo es invocando la Proposición 14.28, obteniéndose

$$c_1 = \frac{\vec{v} \cdot \vec{v}_1}{||\vec{v}_1||^2} = \frac{11}{17},\ c_2 = \frac{\vec{v} \cdot \vec{v}_2}{||\vec{v}_2||^2} = \frac{-7}{17} \text{ y } c_3 = \frac{\vec{v} \cdot \vec{v}_3}{||\vec{v}_3||^2} = \frac{4}{4} = 1.$$

Definición 14.31 *Sea S un subespacio de $\mathbb{R}^n$ y sea $\{\vec{v}_1, \ldots, \vec{v}_m\}$ un subconjunto de S. Decimos que $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es una* ***base ortogonal*** *para S syss $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es un conjunto ortogonal y $span\{\vec{v}_1, \ldots, \vec{v}_m\} = S$.*

Ejemplo 14.32 *La base estándar $\{\vec{e}_1, \ldots, \vec{e}_n\}$ es una base ortogonal para $\mathbb{R}^n$ y el conjunto $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ que definimos previamente en el ejemplo 14.30 es una base ortogonal para el subespacio que consiste de aquellos vectores de $\mathbb{R}^4$ cuya cuarta coordenada es igual a cero.*

Observación 14.33 *Los escalares $c_1, c_2, \ldots, c_n$ de la Proposición 14.28 se conocen como las* ***coordenadas*** *del vector $\vec{v}$ con respecto a la base ortogonal $\{\vec{v}_1, \ldots, \vec{v}_m\}$.*

Capítulo 15

Bases ortogonales y mínimos cuadrados

Proyecciones ortogonales

La Proposición 14.28 vista al final del capítulo anterior nos introduce a las ventajas de contar con bases ortogonales para un subespacio. Recordemos que si $\{\vec{v}_1, \ldots, \vec{v}_m\}$ es una base ortogonal de un subespacio S de $\mathbb{R}^n$, entonces para todo $\vec{v} \in S$ se tiene que

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \ldots + c_m\vec{v}_m,$$

en donde

$$c_1 = \frac{\vec{v} \cdot \vec{v}_1}{||\vec{v}_1||^2},\ c_2 = \frac{\vec{v} \cdot \vec{v}_2}{||\vec{v}_2||^2}, \ldots,\ c_m = \frac{\vec{v} \cdot \vec{v}_m}{||\vec{v}_m||^2}.$$

De acuerdo a la Observación 14.33, nos referimos a los escalares c_i como las coordenadas del vector $\vec{v}$ con respecto a la base $\{\vec{v}_1, \ldots, \vec{v}_m\}$. En el caso de bases ortogonales estas coordenadas son muy simples de calcular.

Observando la Figura 15-1, vemos que para toda $i = 1, 2, \ldots, m$, los vectores $c_i\vec{v}_i$ son aquellos para los cuales la distancia

$$\|\vec{v} - c_i\vec{v}_i\|$$

es mínima y adicionalmente,

$$(\vec{v} - c_i\vec{v}_i) \perp \vec{v}.$$

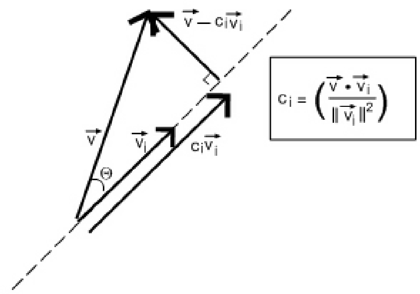

Figura 15-1 Proyección ortogonal de $\vec{v}$ sobre $\vec{v}_i$.

Dadas las consideraciones anteriores, la siguiente definición de la proyección ortogonal de un vector sobre otro resulta natural.

Definición 15.1 *Sean $\vec{u}$ y $\vec{v}$ vectores en $\mathbb{R}^n$ con $\vec{v} \neq \vec{0}$. La **proyección ortogonal** de $\vec{v}$ sobre $\vec{u}$, se denota por $proj_{\vec{u}}\vec{v}$ y se define como el vector*

$$proj_{\vec{u}}\vec{v} = \left(\frac{\vec{v} \cdot \vec{u}}{||\vec{u}||^2}\right)\vec{u}. \tag{15.1}$$

Observación 15.2 *En algunos textos, $proj_{\vec{u}}\vec{v}$ se conoce como "la componente de $\vec{v}$ a lo largo de $\vec{u}$".*

Ejemplo 15.3 *Si* $\vec{u} = \begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix}$ *y* $\vec{v} = \begin{bmatrix} 7 \\ 14 \\ 4 \end{bmatrix}$, *entonces*

$$proj_{\vec{u}}\vec{v} = \left(\frac{\vec{v} \cdot \vec{u}}{||\vec{u}||^2}\right)\vec{u} = \left(\frac{-14 + 70 + 4}{30}\right)\begin{bmatrix} -2 \\ 5 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 10 \\ 2 \end{bmatrix}.$$

Se tiene así que la distancia mínima entre $\vec{v}$ y el subespacio generado por $\vec{u}$ (una recta) es igual a

$$\begin{aligned} \|\vec{v} - proj_{\vec{u}}\vec{v}\| &= \left\| \begin{bmatrix} 7 \\ 14 \\ 4 \end{bmatrix} - \begin{bmatrix} -4 \\ 10 \\ 2 \end{bmatrix} \right\| \\ &= \left\| \begin{bmatrix} 11 \\ 4 \\ 2 \end{bmatrix} \right\| = \sqrt{121 + 16 + 4} = \sqrt{141}. \end{aligned}$$

Observación 15.4 *Si $\{\vec{v}_1, \ldots \vec{v}_m\}$ es una base ortogonal para S y $\vec{v} \in S$, entonces*

$$\vec{v} = \frac{\vec{v}\cdot\vec{v}_1}{||\vec{v}_1||^2}\vec{v}_1 + \frac{\vec{v}\cdot\vec{v}_2}{||\vec{v}_2||^2}\vec{v}_2 + \ldots + \frac{\vec{v}\cdot\vec{v}_m}{||\vec{v}_m||^2}\vec{v}_m$$

$$= proj_{\vec{v}_1}\vec{v} + proj_{\vec{v}_2}\vec{v} + \ldots + proj_{\vec{v}_m}\vec{v}.$$

Por ejemplo, en el contexto de $\mathbb{R}^3$, tenemos que la base estándar $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ es una base ortogonal para $\mathbb{R}^3$ y cualquier vector $\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ puede expresarse como

$$\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = proj_{\vec{e}_1}\vec{v} + proj_{\vec{e}_2}\vec{v} + proj_{\vec{e}_3}\vec{v} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix}.$$

Proyección ortogonal sobre subespacios

Hasta el momento hemos visto como obtener la proyección (ortogonal) de un vector $\vec{y} \in \mathbb{R}^n$ sobre una recta; es decir sobre un subespacio de la forma $span\{\vec{x}\}$ para $\vec{x} \in \mathbb{R}^n, \vec{x} \neq \vec{0}$. En otras palabras, podemos obtener proyecciones sobre subespacios de dimensión uno. En el caso de proyecciones sobre espacios de cualquier dimensión tenemos la siguiente definición.

Definición 15.5 *Sea S un subespacio de $\mathbb{R}^n$ y supongamos que el conjunto $\{\vec{v}_1, \ldots \vec{v}_m\}$ es una base ortogonal para S. Si $\vec{v} \in \mathbb{R}^n$ (independientemente de si $\vec{v} \in S$ o $\vec{v} \notin S$), entonces definimos la proyección ortogonal de $\vec{v}$ sobre S como*

$$proj_S\vec{v} = \frac{\vec{v}\cdot\vec{v}_1}{||\vec{v}_1||^2}\vec{v}_1 + \frac{\vec{v}\cdot\vec{v}_2}{||\vec{v}_2||^2}\vec{v}_2 + \ldots + \frac{\vec{v}\cdot\vec{v}_m}{||\vec{v}_m||^2}\vec{v}_m$$

$$= proj_{\vec{v}_1}\vec{v} + proj_{\vec{v}_2}\vec{v} + \ldots + proj_{\vec{v}_m}\vec{v}$$

Gracias a la Observación 15.4, es claro que si $\vec{v} \in S$, entonces

$$\vec{v} = proj_S\vec{v}.$$

En la figura 15-2 se ilustra el caso en el cual el subespacio es un plano.

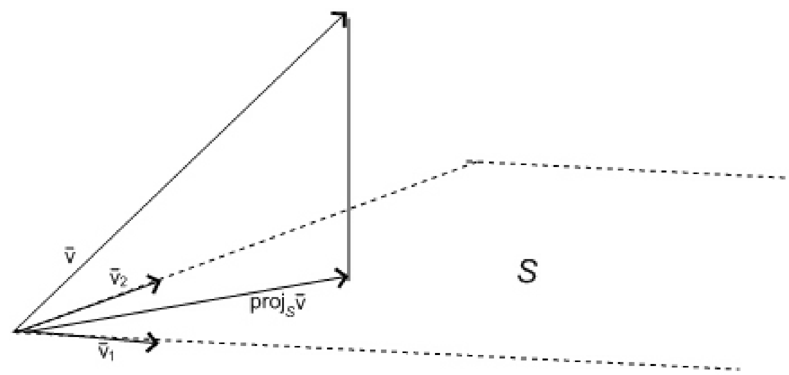

Figura 15-2 Proyección ortogonal de $\vec{v}$ sobre un plano S.

Observación 15.6 *Evidentemente, si* $m = 1$ *se tiene que* $proj_S \vec{v} = proj_{\vec{v}_1} \vec{v}$.

Proposición 15.7 *Con la notación anterior, si* $\vec{s} \in S$, *entonces* $\vec{s} \perp (\vec{v} - proj_S \vec{v})$, *es decir,* $\vec{v} - proj_S \vec{v}$ *es ortogonal a todos los vectores de* S.

Demostración. *Dado que* $\{\vec{v}_1, \dots \vec{v}_m\}$ *es una base ortogonal para* S, *es suficiente probar que para toda* $i = 1, 2, \dots, m$ *se cumple*

$$\vec{v}_i \perp (\vec{v} - proj_S \vec{v}).$$

Observando que

$$\begin{aligned} \vec{v}_i \cdot (\vec{v} - proj_S \vec{v}) &= \vec{v}_i \cdot \vec{v} - \vec{v}_i \cdot proj_S \vec{v} \\ &= \vec{v}_i \cdot \vec{v} - \vec{v}_i \cdot \left(\frac{\vec{v} \cdot \vec{v}_1}{||\vec{v}_1||^2} \vec{v}_1 + \cdots + \frac{\vec{v} \cdot \vec{v}_m}{||\vec{v}_m||^2} \vec{v}_m \right) \\ &= \vec{v}_i \cdot \vec{v} - \left(\frac{\vec{v} \cdot \vec{v}_i}{||\vec{v}_i||^2} (\vec{v}_i \cdot \vec{v}_i) \right) = \vec{v}_i \cdot \vec{v} - \vec{v}_i \cdot \vec{v} = 0, \end{aligned}$$

por lo que $\vec{v}_i \perp (\vec{v} - proj_S \vec{v})$. ■

Ejemplo 15.8 *Sea*

$$span \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\} = S \subset \mathbb{R}^3.$$

Si $\vec{v} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$, *encontramos* $proj_S \vec{v}$ *como sigue. Primero encontramos* *una base ortogonal para S. Sea* $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ *y*

$$\vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - proj_{\vec{v}_1} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}.$$

Como $\|\vec{v}_1\|^2 = 3$ *y*

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

se tiene que

$$\vec{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} -4 \\ -1 \\ 5 \end{bmatrix}.$$

Como cualquier múltiplo no nulo de $\vec{v}_2$ *seguirá siendo ortogonal a* $\vec{v}_1$, *sin pérdida de generalidad tomamos el vector* $3\vec{v}_2$ *por lo que*

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -4 \\ -1 \\ 5 \end{bmatrix} \right\}$$

es una base ortogonal de S. Observando que

$$\begin{aligned} &\|\vec{v}_1\|^2 = 3, \|3\vec{v}_2\|^2 = 42, \\ &\vec{v} \cdot \vec{v}_1 = 6 \quad y \quad \vec{v} \cdot 3\vec{v}_2 = 18, \end{aligned}$$

calculamos $proj_S \vec{v}$ *de acuerdo a la Definición 15.5:*

$$\begin{aligned} proj_S \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} &= proj_{\vec{v}_1} \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} + proj_{3\vec{v}_2} \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} \\ &= \frac{6}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{18}{42} \begin{bmatrix} -4 \\ -1 \\ 5 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 2 \\ 11 \\ 29 \end{bmatrix}. \end{aligned}$$

Notamos que

$$\vec{v} - proj_S \vec{v} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix} - \frac{1}{7} \begin{bmatrix} 2 \\ 11 \\ 29 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 12 \\ -18 \\ 6 \end{bmatrix}$$

y se cumplen

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \frac{1}{7} \begin{bmatrix} 12 \\ -18 \\ 6 \end{bmatrix} = 0 = \begin{bmatrix} -4 \\ -1 \\ 5 \end{bmatrix} \cdot \frac{1}{7} \begin{bmatrix} 12 \\ -18 \\ 6 \end{bmatrix}.$$

Ejemplo 15.9 *Sea S el subespacio de $\mathbb{R}^4$ que tiene como base al conjunto $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, donde*

$$\vec{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \ \vec{w}_2 = \begin{bmatrix} -1 \\ 2 \\ -1 \\ 2 \end{bmatrix}, \ \vec{w}_3 = \begin{bmatrix} -3 \\ -3 \\ 1 \\ 2 \end{bmatrix}$$

y supongamos que queremos encontrar la proyección ortogonal de $\vec{u}$ sobre S, donde $\vec{u} = \begin{bmatrix} 6 \\ 6 \\ -2 \\ -4 \end{bmatrix}$.

Claramente $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ es un conjunto ortogonal Como $||\vec{w}_1||^2 = 3$, $||\vec{w}_2||^2 = 10$ y $||\vec{w}_3||^2 = 23$, tenemos que, por la Definición 15.5 $proj_S\vec{u}$ está dada por,

$$\begin{aligned} proj_S\vec{u} &= proj_{\vec{w}_1}\vec{u} + proj_{\vec{w}_2}\vec{u} + proj_{\vec{w}_3}\vec{u} \\ &= \frac{\vec{u}\cdot\vec{w}_1}{||\vec{w}_1||^2}\vec{v}_1 + \frac{\vec{u}\cdot\vec{w}_2}{||\vec{w}_2||^2}\vec{v}_2 + \frac{\vec{u}\cdot\vec{v}_3}{||\vec{w}_3||^2}\vec{w}_3 \\ &= \frac{0}{||\vec{w}_1||^2}\vec{v}_1 + \frac{0}{||\vec{w}_2||^2}\vec{v}_2 + \left(-\frac{46}{23}\right)\vec{w}_3 = -2\begin{bmatrix} -3 \\ -3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ -2 \\ -4 \end{bmatrix}. \end{aligned}$$

Observamos que $proj_S\vec{u} = \vec{u}$, lo cual tiene sentido porque $\vec{u} = -2\vec{w}_3 \in S$.

El siguiente teorema resume algunas propiedades de la proyección ortogonal.

Teorema 15.10 (Propiedades de la proyección ortogonal) *Sea S un subespacio de $\mathbb{R}^n$ con base ortogonal $\{\vec{v}_1, \dots \vec{v}_m\}$ y sea $\vec{v}$ un vector cualquiera en $\mathbb{R}^n$. Entonces:*

1. $proj_S\vec{v} = proj_{\vec{v}_1}\vec{v} + proj_{\vec{v}_2}\vec{v} + \ldots + proj_{\vec{v}_m}\vec{v} \in S$.

2. $proj_S\vec{v} = \vec{v}$ *syss* $\vec{v} \in S$.

3. $\vec{v} - proj_S\vec{v}$ *es ortogonal a todos los elementos de* S.

4. $proj_S\vec{v}$ *es independiente de la elección de la base ortogonal para* S.

5. *$proj_S\vec{v}$ es el único vector en S que está lo más cerca de $\vec{v}$. Esto es,*

$$||\vec{v} - proj_S\vec{v}|| \leq ||\vec{v} - \vec{z}||$$

para toda $\vec{z}$ en S y la igualdad se tiene syss $\vec{z} = proj_S\vec{v}$.

Observación 15.11 *Dados $\vec{v} \in \mathbb{R}^n$ y un subespacio $S \subset \mathbb{R}^n$, podemos pensar en la siguiente transformación lineal.*

$$\begin{gathered} P_S : \mathbb{R}^n \longrightarrow \mathbb{R}^n, \\ P_S(\vec{v}) = proj_S\vec{v}. \end{gathered}$$

Utilizando la Definición 15.5, es posible probar que esta función cumple con las propiedades de una transformación lineal.

Proceso de ortogonalización de Gram-Schmidt

Dado cualquier subespacio S en $\mathbb{R}^n$, es natural preguntarse si siempre existe una base ortogonal para éste. Por ejemplo, supongamos que

$$S = span\{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset \mathbb{R}^4,$$

en donde

$$\vec{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{w}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{w}_3 = \begin{bmatrix} 0 \\ 2 \\ 4 \\ 0 \end{bmatrix}.$$

Estos vectores son LI puesto que claramente $\vec{w}_2 \notin span\{\vec{w}_1\}$ y $\vec{w}_3 \notin span\{\vec{w}_1, \vec{w}_2\}$, no obstante

$$\vec{w}_1 \cdot \vec{w}_2 = -1, \quad \vec{w}_1 \cdot \vec{w}_3 = 2, \quad \vec{w}_2 \cdot \vec{w}_3 = 2,$$

de manera que ninguno de estos vectores es ortogonal a los otros. ¿Cómo encontrar vectores ortogonales que generen al mismo subespacio S? Queremos construir una base ortogonal $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ para S, es decir, se deben cumplir

$$\begin{gathered} v_1 \cdot v_2 = v_1 \cdot v_3 = v_2 \cdot v_3 = 0, \\ span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = S. \end{gathered}$$

Para construir dicha base procedemos como sigue. Sea $\vec{v}_1$ cualquier vector de la base original, digamos $\vec{w}_1$, de manera que $\vec{v}_1 = \vec{w}_1$ y $||v_1||^2 = ||w_1||^2 = 5$. Como $\vec{w}_2$ no es ortogonal a $\vec{w}_1$, lo reemplazamos por

$$\begin{aligned}\vec{v}_2 &= \vec{w}_2 - proj_{\vec{v}_1}\vec{w}_2 = \vec{w}_2 - \frac{\vec{w}_2 \cdot \vec{v}_1}{||\vec{v}_1||^2}\vec{v}_1 \\ &= \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \left(-\frac{1}{5}\right)\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{-3}{5} \\ \frac{6}{5} \\ 0 \\ 0 \end{bmatrix}\end{aligned}$$

y verificamos que

$$\vec{v}_2 \perp \vec{v}_1.$$

En efecto,

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

La intuición geométrica del procedimiento anterior es que $\vec{w}_2$ se descompone como la suma de dos vectores ortogonales, concretamente,

$$\vec{w}_2 = proj_{\vec{v}_1}\vec{w}_2 + \left(\vec{w}_2 - proj_{\vec{v}_1}\vec{w}_2\right),$$

adicionalmente $\vec{v}_2$ es la componente del vector que es ortogonal a $\vec{v}_1$. Finalmente, el vector $\vec{v}_3$ se construye a partir de $\vec{w}_3$ de forma similar. Si $R = span\{\vec{v}_1.\vec{v}_2\}$, entonces $\{\vec{v}_1.\vec{v}_2\}$ es una base ortogonal para R y definimos

$$\begin{aligned}\vec{v}_3 &= \vec{w}_3 - proj_R\vec{w}_3 = \vec{w}_3 - (proj_{\vec{v}_1}\vec{w}_3 + proj_{\vec{v}_2}\vec{w}_3) \\ &= \vec{w}_3 - \left(\frac{\vec{w}_3 \cdot \vec{v}_1}{||\vec{v}_1||^2}\vec{v}_1 + \frac{\vec{w}_3 \cdot \vec{v}_2}{||\vec{v}_2||^2}\vec{v}_2\right) \\ &= \begin{bmatrix} 0 \\ 2 \\ 4 \\ 0 \end{bmatrix} - \left(\frac{2}{5}\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \frac{12}{5} \times \frac{5}{9}\begin{bmatrix} \frac{-3}{5} \\ \frac{6}{5} \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 4 \\ 0 \end{bmatrix}.\end{aligned}$$

Observamos que, por el Teorema 15.10, para toda $\vec{s} \in S$ se cumple

$$\vec{s} \cdot (\vec{w}_3 - proj_R\vec{w}_3) = \vec{s} \cdot \vec{v}_3 = 0;$$

en particular, como $\vec{v}_1, \vec{v}_2 \in S$, se tiene que

$$\vec{v}_3 \cdot \vec{v}_1 = \vec{v}_3 \cdot \vec{v}_2 = 0.$$

Esto último puede verificarse haciendo el cálculo directo con los vectores encontrados. Tenemos así que el conjunto $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ es una base ortogonal para S.

La discusión anterior nos lleva al siguiente teorema.

Teorema 15.12 (Proceso de ortogonalización de Gram - Schmidt) *Supongamos que S es un subespacio con base $\{\vec{w}_1, \ldots, \vec{w}_m\}$. Definimos $\vec{v}_1 = \vec{w}_1$ y*

$$\vec{v}_k = \vec{w}_k - proj_{S_{k-1}}\vec{w}_k = \vec{w}_k - \sum_{t=1}^{k-1} proj_{\vec{v}_t}\vec{w}_k = \vec{w}_k - \sum_{t=1}^{k-1} \frac{\vec{w}_k \cdot \vec{v}_t}{||\vec{v}_t||^2}\vec{v}_t$$

para $k \geq 2$. Entonces $\{\vec{v}_1, \ldots \vec{v}_m\}$ es una base ortogonal para S.

Demostración. La demostración es inmediata por inducción, notando que si $m = 1$, $\{\vec{v}_1\}$ es un conjunto ortogonal. Si $\{\vec{v}_1, \ldots \vec{v}_{k-1}\}$ es un conjunto ortogonal, definimos $S_{k-1} = span\{\vec{v}_1, \ldots \vec{v}_{k-1}\}$. Se tiene que para todo $\vec{v}_j$ con $j \in \{1, 2, \ldots, k-1\}$,

$$\vec{v}_j \cdot \left(\vec{w}_{kj} - proj_{S_{k-1}}\vec{w}_{kj}\right) = 0,$$

por lo tanto,

$$\vec{v}_j \cdot \vec{v}_k = 0.$$

y $\{\vec{v}_1, \ldots \vec{v}_k\}$ es un conjunto ortogonal. Se concluye que para todo $m \in \mathbb{N}$, $\{\vec{v}_1, \ldots \vec{v}_m\}$ es un conjunto ortogonal y, por lo tanto, es un conjunto LI con m elmentos que constituye una base ortogonal para S[1]. ■

Observación 15.13 *Si $\{\vec{v}_1, \ldots \vec{v}_m\}$ es la base ortogonal para S obtenida por el proceso de Gram-Schimdt, normalizando los vectores de esta base obtenemos lo que se conoce como una base ortonormal $\{\vec{u}_1, \ldots \vec{u}_m\}$ para S; es decir, se tiene que para todo $i, j = 1, 2, \ldots, m$, se cumple*

$$\vec{u}_i \cdot \vec{u}_j = \begin{cases} 0 \ si \ i \neq j, \\ 1 \ si \ i = j. \end{cases}$$

Los vectores de esta base ortonormal están dados como

$$\vec{u}_i = \frac{1}{\|\vec{v}_i\|}\vec{v}_i.$$

Mínimos cuadrados

Preliminares

Sea $S = span\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m\} \subset \mathbb{R}^n$. Definimos la matriz

$$A = [\vec{u}_1 \ \vec{u}_2 \cdots \vec{u}_m] \in M_{n \times m},$$

[1] Observemos que los vectores $v_i's$, por construcción, son combinación lineal de los $w_j's$ y, por lo tanto todos pertenecen a S.

cuyas columnas son los vectores que generan a S. Por construcción, el espacio de columnas $\text{col}(A)$ es igual a S. Sea $\vec{v}$ cualquier vector en $\mathbb{R}^n$. Como $proj_S\vec{v} \in S$, existirá alguna $\vec{x} \in \mathbb{R}^m$ tal que

$$A\vec{x} = proj_S\vec{v}.$$

Dado que $\vec{u}_i \in S$, sabemos que para todo $i = 1, 2, \ldots, m$, se cumple

$$\vec{u}_i \cdot (\vec{v} - A\vec{x}) = 0 = \vec{u}_i^T\,(\vec{v} - A\vec{x})\,. \tag{15.2}$$

Como

$$A^T = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix},$$

utilizamos la expresión (15.2) y la definición de producto de matrices para concluir que

$$A^T\,(\vec{v} - A\vec{x}) = 0.$$

Las consideraciones anteriores nos llevan al siguiente resultado.

Proposición 15.14 *Sean $\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m\}$ cualquier conjunto de generadores de $S \subset \mathbb{R}^n$, $\vec{v}$ cualquier vector en $\mathbb{R}^n$ y*

$$A = [\vec{u}_1\ \vec{u}_2 \cdots \vec{u}_m] \in M_{n\times m}.$$

Entonces existe $\vec{x} \in \mathbb{R}^m$ tal que $A\vec{x} = proj_S\vec{v}$ y adicionalmente se tiene que,

$$(\vec{v} - A\vec{x}) \in Null(A^T).$$

Definición 15.15 *Con la misma notación que arriba, el vector $proj_S\vec{v}$ se conoce como la **aproximación de mínimos cuadrados de $\vec{v}$ sobre** S y $\vec{v} - proj_S\vec{v}$ se denomina el **vector residual.***

Aplicaciones

En la siguiente sección analizaremos algunas aplicaciones importantes de los vectores que acabamos de definir.

Supongamos que tenemos m observaciones (correspondientes a millones de desempleados, por ejemplo) a lo largo del tiempo:

$$(t_1, y_1), (t_2, y_2), \ldots (t_m, y_m)$$

con $t_1 < t_2 < \ldots < t_m$. Se conjetura que existe una relación lineal del tipo

$$y = ct + d,$$

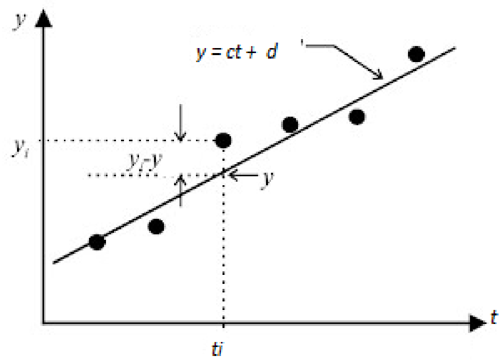

Figura 15-3 Aproximación lineal

entre el tiempo y el número de desempleados. Como se muestra en la Figura 15-3, cualquier recta que se proponga pasará, probablemente, por muy pocos puntos.

En términos algebraicos, tenemos m valores para t y y y queremos encontrar valores para c y d tales que se cumplan las m ecuaciones dadas por

$$ct_i + d = y_i, \quad i = 1, 2, \ldots, m.$$

La matriz de este sistema está dada por

$$A = \begin{bmatrix} t_1 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}$$

y si $m > 2$ el sistema será inconsistente, a menos que todos los puntos estén perfectamente alineados en una recta (poco probable).

Sea $S = \text{col}(A)$, la situación que tenemos es que el vector

$$\vec{y} = \begin{bmatrix} y_i \\ \vdots \\ y_m \end{bmatrix}$$

usualmente no pertenecerá a S, es decir, el sistema

$$A\vec{x} = \vec{y}$$

será inconsistente de manera que se tendrá

$$\|\vec{y} - A\vec{x}\| > 0.$$

Sin embargo, podemos encontrar la solución que mejor aproxime al sistema en el siguiente sentido: se desea encontrar $\vec{x} = \begin{bmatrix} c \\ d \end{bmatrix}$ de manera que

$$\|\vec{y} - A\vec{x}\|$$

sea lo más pequeño posible. De acuerdo a lo visto en la sección anterior, sabemos que $\vec{x}$ debe ser una solución al sistema dado por

$$A\vec{x} = proy_S\vec{y}.$$

Una observación importante es que sabemos que $\vec{x}$ es tal que el vector residual, dado por

$$\vec{y} - proy_S\vec{y} = \vec{y} - A\vec{x},$$

minimiza la distancia

$$\|\vec{y} - A\vec{x}\|,$$

equivalentemente, se minimiza

$$\|\vec{y} - A\vec{x}\|^2.$$

En otras palabras, se minimiza la cantidad

$$E = \sum_{i=1}^{m}(y_i - ct_i - d)^2,$$

de manera que c y d son tales que la suma de los cuadrados de las distancias verticales entre la recta estimada y los datos observados es mínima (de aquí el nombre de mínimos cuadrados).

Utilizando la Proposición 15.14 sabemos que

$$\vec{y} - A\vec{x} \in Null(A^T),$$

por lo tanto $\vec{x}$ es tal que

$$A^T(\vec{y} - A\vec{x}) = 0$$

o reescribiendo,

$$A^TA\vec{x} = A^T\vec{y}. \tag{15.3}$$

Por construcción, sabemos que este sistema es consistente y siempre podremos encontrar el vector $\vec{x}$. En este caso particular, si $\vec{x} = \begin{bmatrix} c \\ d \end{bmatrix}$ es solución del sistema (15.3), entonces $y = ct + d$ será la recta que mejor aproxima a los datos (el conjunto de puntos $\{(t_i, yi)\}$).

Observamos que, para este caso, las columnas de A son LI. Cuando esto sucede el problema se simplifica en virtud del siguiente resultado.

Teorema 15.16 *Si $A \in M_{n\times m}$ y $rank(A) = m$, entonces $A^TA \in M_{m\times m}$ es invertible.*

Demostración. Primero vamos a probar que

$$Null(A) = Null(A^T A).$$

Sea $\vec{x} \in Null(A)$, es decir, se satisface $A\vec{x} = \vec{0}$ y, por lo tanto $A^T A\vec{x} = \vec{0}$ y $\vec{x} \in Null(A^T A)$. Concluimos así, que

$$Null(A) \subset Null(A^T A).$$

Análogamente, si $\vec{x} \in Null(A^T A)$, entonces $A^T A\vec{x} = \vec{0}$ y multiplicando por $\vec{x}^T$, por la izquierda, obtenemos

$$\vec{x}^T A^T A\vec{x} = \vec{x}^T \vec{0} = \vec{0}.$$

Reescribimos esta última igualdad como

$$(A\vec{x})^T (A\vec{x}) = \vec{0} = (A\vec{x}) \cdot (A\vec{x}) = \|A\vec{x}\|^2 = 0.$$

Por lo tanto, necesariamente $A\vec{x} = \vec{0}$ y $\vec{x} \in Null(A)$. De aquí que,

$$Null(A^T A) \subset Null(A)$$

y concluimos que,

$$Null(A) = Null(A^T A).$$

Ahora bien, por el Teorema de la dimensión (Teorema 12.7),

$$Rank(A) + Null(A) = Rank(A^T A) + Null(A^T A) = m,$$

por lo que $Rank(A) = Rank(A^T A) = m$. Como $A^T A \in M_{m\times m}$ y sus columnas son LI, esta matriz es invertible. ■

Observación 15.17 *Podríamos no haber invocado el Teorema de la dimensión, simplemente notando que como todas las columnas de A son LI, entonces $Null(A) = 0 = Null(A^T A)$. De esta forma, $A^T A$ corresponde a una transformación lineal inyectiva entre $\mathbb{R}^m$ y $\mathbb{R}^m$, por lo tanto será biyectiva y $A^T A$ es invertible.*

Corolario 15.18 *Sean $\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_m\}$ cualquier base de $S \subset \mathbb{R}^n$, $\vec{y}$ cualquier vector en $\mathbb{R}^n$ y*

$$A = [\vec{u}_1\ \vec{u}_2 \cdots \vec{u}_m] \in M_{n\times m}.$$

Entonces existe $\vec{x} \in \mathbb{R}^m$ tal que $A\vec{x} = proj_S\vec{y}$. Adicionalmente, se tiene que

$$\vec{x} = \left(A^T A\right)^{-1} A^T \vec{y}.$$

Demostración. Por la proposición anterior, $A^T A$ es invertible. Por lo tanto, multiplicando la expresión (15.3) por $\left(A^T A\right)^{-1}$ por la izquierda, se obtiene

$$\vec{x} = \left(A^T A\right)^{-1} A^T \vec{y}.$$

■

En particular, la solución para el problema de encontrar la recta $y = ct + d$ que mejor aproxime a los puntos dados está dada por

$$\vec{x} = \begin{bmatrix} c \\ d \end{bmatrix} = (A^T A)^{-1} A^T \vec{y},$$

,

donde $A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{bmatrix}$ y $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$.

Ejemplo 15.19 *Supongamos que los datos observados son* $(1, 2)$, $(2, 3)$, $(3, 5)$ *y* $(4, 7)$. *En este caso,*

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix}.$$

Por lo tanto,

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \ y \ (A^T A)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} c \\ d \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 1.7 \\ 0 \end{bmatrix}$$

Así, la recta que mejor predice el desempleo futuro es

$$y = ct + d = 1.7t.$$

Ejemplo 15.20 *Uno de los primeros matemáticos en utilizar el método de mínimos cuadrados fue K.F.Gauss, quien en 1801 aproximó la* *órbita del asteroide Ceres. Las órbitas de los cuerpos celestes del sistema solar siempre siguen una curva cónica. En coordenadas polares, la distancia r y el ángulo θ con respecto al sol satisfacen la ecuación*

$$r = a - e(r\cos\theta),$$

en donde e representa la eccentricidad de la órbita y $a > 0$. Recordemos que $e = 0$ en un círculo, $0 < e < 1$ en una elipse, $e = 1$ en una parábola y $e > 1$ en una hipérbola.

Supongamos que las observaciones de un cometa arrojan el siguiente conjunto de datos para (r, θ) :

$$\{(1.2, 0.3), (2.1, 1.2), (4.1, 2.6), (6.3, 3.8)\}.$$

Para encontrar la órbita aproximada debemos encontrar A y e que mejor aproximen el sistema lineal

$$\begin{aligned} a - e(1.2\cos 0.3) &= 1.2, \\ a - e(2.1\cos 1.2) &= 2.1, \\ a - e(4.1\cos 2.6) &= 4.1, \\ a - e(6.3\cos 3.8) &= 6.3, \end{aligned}$$

mismo que puede reescribirse como

$$\begin{aligned} a - 1.146e &= 1.2, \\ a - 0.761e &= 2.1, \\ a + 3.513e &= 4.1, \\ a + 4.983e &= 6.3. \end{aligned}$$

De aquí

$$A = \begin{bmatrix} 1 & -1.146 \\ 1 & -0.761 \\ 1 & 3.513 \\ 1 & 4.983 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1.2 \\ 2.1 \\ 4.1 \\ 6.3 \end{bmatrix}.$$

Por lo tanto,

$$\vec{x} = \begin{bmatrix} a \\ e \end{bmatrix} = \left(A^T A\right)^{-1} A^T \vec{y} = \begin{bmatrix} 2.242\,3 \\ 0.718\,05 \end{bmatrix}.$$

Concluimos que la órbita cónica que mejor aproxima al conjunto de datos observados es

$$r = 2.242\,3 - 0.718\,05(r\cos\theta)$$

lo cual corresponde a una órbita elíptica.