Álgebra Lineal - Grado de Estadística

Departamento de Álgebra
Universidad de Sevilla

Índice general

* *Nota:* los capítulos y secciones con asterisco son opcionales.

Capítulo 0

* Lenguaje

El objetivo principal en este tema es desarrollar el uso del lenguaje en el contexto de las matemáticas.

0.1. Lógica proposicional

0.1.1. Expresiones

La primera cuestión sobre la que trataremos es qué clase de expresiones se usan en matemáticas como ladrillos para construir. Recordemos de la enseñanza primaria que tenemos oraciones declarativas, imperativas, interrogativas, y exclamaciones. En matemáticas se usan las declarativas, pero tenemos que precisar un poco más. Definimos una *expresión* de forma intuitiva como un enunciado que puede ser asignado a la clase de cosas que llamamos VERDAD o a la clase de cosas que llamamos FALSO. Sin embargo, pronto encontramos problemas.

En primer lugar, tenemos las paradojas. Por ejemplo, “Esta frase es falsa", no puede ser ni verdadera ni falsa. Si decimos que es verdad, ella misma dice que es falsa. Y si suponemos que es falsa, nos dice que es verdadera. No queremos tener paradojas dentro de las sentencias. Una de las más famosas se puede encontrar en el capítulo LI de la segunda parte de 'El Quijote'.

En segundo lugar, algunas expresiones contienen lo que los lógicos llaman una *indeterminada*. La presencia de una indeterminada en una expresión hace que no sea una oración. Por ejemplo, si decimos

x se puede escribir como la suma de dos números primos

no la consideramos una oración. El uso de la indeterminada x es como dejar un hueco sin rellenar, por lo que la expresión no se puede calificar como verdadera o falsa. Sin embargo, si decimos

Todo entero entre 3 y 20 se puede escribir como la suma de dos números primos,

entonces sí tenemos una oración.

Ahora imaginemos el conjunto $\mathscr{S}$ de todas las posibles oraciones. Este conjunto es enorme y complicado, pero una importante característica es que cada oración se puede poner en exactamente uno de dos subconjuntos: $\mathscr{V}$ (oraciones verdaderas) y $\mathscr{F}$ (oraciones falsas). Queremos ver las relaciones entre estas expresiones. En concreto, extraeremos elementos de $\mathscr{S}$ y los combinaremos para generar otros elementos de $\mathscr{S}$, y veremos cómo el carácter de verdadero o falso de las expresiones elegidas determinan el carácter de la nueva. Vamos a estudiar tres formas de combinarlas: negación, conjunción y disyunción.

0.1.2. Negación

En general, usaremos las letras p, q, r y demás para representar frases de manera simbólica. Por ejemplo, definimos una frase p como sigue:

p: Marta ha alquilado un coche hoy.

Ahora consideramos la *negación* de p, que la notaremos por $\neg p$. En el ejemplo se traduce como

$\neg p$: Marta no ha alquilado un coche hoy.

Si p es verdad, entonces $\neg p$ es falsa, y al revés. Esto lo escribiremos mediante una *tabla de verdad*.

p	$\neg p$
V	F
F	V

0.1.3. Conjunción

Cuando dos expresiones se unen mediante la conjunción Y para producir una oración compuesta, necesitamos una forma de distinguir si la oración compuesta es verdadera o falsa, según los sean las oraciones de partida. La notaremos por $p \wedge q$ (leído p y q), y su tabla de verdad es la siguiente:

p	q	$p \wedge q$
V	V	V
V	F	F
F	V	F
F	F	F

0.1.4. Disyunción

La sentencia "$p \vee q$", que se lee "p o q", es verdadera cuando una de las dos es verdadera. Su tabla de verdad es la siguiente:

p	q	$p \vee q$
V	V	V
V	F	V
F	V	V
F	F	F

Es diferente a otra sentencia relacionada, que es la disyunción exclusiva (notada XOR en informática), que es verdadera cuando una, y solamente una, de las expresiones es verdadera. La tabla de verdad asociada es

p	q	p XOR q
V	V	F
V	F	V
F	V	V
F	F	F

0.1.5. Equivalencia lógica

Existen diferentes formas de decir lo mismo. Queremos saber cuándo dos construcciones diferentes con expresiones quieren decir lo mismo, es decir, sean *lógicamente equivalentes*.

Equivalencia lógica

Dos expresiones U y V se dicen **lógicamente equivalente** si tienen la misma tabla de verdad. Lo notaremos por $U \equiv V$.

Por ejemplo, es fácil ver que $p \wedge q$ y $q \wedge p$ son lógicamente equivalentes. De igual forma se puede comprobar la propiedad asociativa de la conjunción y la disyunción, es decir:

$$p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r,$$
$$p \vee (q \vee r) \equiv (p \vee q) \vee r.$$

0.1.6. Tautologías y contradicciones

En ocasiones encontramos que la tabla de verdad de una expresión es siempre verdadera para todos los valores de los elementos que la componen. Por ejemplo, esto ocurre en $p \vee \neg p$ o en $(p \wedge q) \vee (\neg p \vee \neg q)$. Estas expresiones se denominan *tautologías*. La negación de una tautología es una contradicción, y su tabla de verdad siempre contiene el valor falso.

0.2. Sentencias condicionales

En esta sección consideraremos la estructura lógica de las expresiones p *implica* o *necesita* q, y las diferentes forma de expresarlas.

0.2.1. Expresiones si ... entonces ...

En los primeros días de clase será habitual encontrar expresiones del siguiente tipo:

O estudias o suspenderás los exámenes.

Como ya estamos habituados a pensar según los esquemas del cálculo proposicional, podemos formar las siguiente expresiones:

p: No estudias.
q: Suspendes el examen.

La expresión inicial se convierte en $\neg p \vee q$, y podemos razonar de la siguiente forma:“Si no estudio tengo garantizado un suspenso, pero incluso aunque estudie, puedo suspender". En lenguaje natural estamos acostumbrados a decirlo como

Si no estudio, suspenderé el examen.

Esta expresión es de tipo condicional, y la notaremos por $p \to q$ (p implica q). Su tabla de verdad es

p	q	$p \to q$
V	V	V
V	F	F
F	V	V
F	F	V

Es equivalente a la expresión $\neg p \vee q$. A la expresión p se la llama *hipótesis* y a q *conclusión*.

0.2.2. Variaciones sobre $p \to q$

Dadas dos expresiones p y q, queremos analizar otras combinaciones de expresiones " si ... entonces ...". Consideremos las siguientes frases:

p: Ana tiene tarea de álgebra lineal.
q: Ana va a la biblioteca.

Escribamos $p \to q$ como nuestra expresión principal, y hagamos otras combinaciones.

$p \to q$	Si Ana tiene tarea de álgebra lineal, entonces ella va a la biblioteca.	Primaria
$q \to p$	Si Ana va a la biblioteca, entonces ella tiene tarea de álgebra lineal.	Recíproca
$\neg p \to \neg q$	Si Ana no tiene tarea de álgebra lineal, entonces ella no va a la biblioteca.	Inversa
$\neg q \to \neg p$	Si Ana no va a la biblioteca, entonces ella no tiene tarea de álgebra lineal.	Contrarrecíproca

Podemos discutir a partir de este ejemplo cuáles de las anteriores son equivalentes entre sí, pero dejemos que sean las tablas de verdad quienes nos lo indiquen.

p	q	$\neg p$	$\neg q$	$p \to q$	$q \to p$	$\neg p \to \neg q$	$\neg q \to \neg p$
V	V	F	F	V	V	V	V
V	F	F	V	F	V	V	F
F	V	V	F	V	F	F	V
F	F	V	V	V	V	V	V

Observamos que $p \to q \equiv \neg q \to \neg p$, y, de forma simétrica, que $q \to p \equiv \neg p \to \neg q$.

Hay una construcción muy importante con las expresiones condicionales que es verdadera cuando p y q son ambas verdaderas o ambas falsas. Dadas dos expresiones p y q, escribiremos $p \leftrightarrow q$, leído " p si y solamente si q", cuando es verdadera para p y q ambas verdaderas o ambas falsas. Es fácil ver que es una forma abreviada de la expresión $(p \to q) \wedge (q \to p)$.

Recordemos que la equivalencia lógica representaba la igualdad de las tablas de verdad entre las expresiones. Lo anterior indica que si U y V son expresiones equivalentes, podemos escribirlo también en la forma $U \leftrightarrow V$.

0.3. Cuantificadores

En las expresiones matemáticas aparecen con frecuencia indeterminadas, y esto lleva a lo que llamamos los cuantificadores universal y de existencia. Consideremos una frase como la siguiente:

Todos los cuadrados son rectángulos.

La podemos escribir en forma condicional como

Si x es un cuadrado, entonces x es un rectángulo,

pero desde un punto de vista más formal se puede expresar como

Para todo x, si x es un cuadrado, entonces x es un rectángulo.

Esta última expresión puede sonar extraña, pero más adelante veremos la importancia de comenzar la frase con “para todo".

0.3.1. El cuantificador universal

La desigualdad $n^2 < 2^n$ es verdad para todos los números naturales $n \geq 5$ (lo probaremos más adelante). Esto lo podemos expresar de diferentes formas:

Para todos los números naturales n, si $n \geq 5$, entonces $n^2 < 2^n$.
Para todo n, si $n \in \mathbb{N}$ y $n \geq 5$, entonces $n^2 < 2^n$.
$(\forall n)[(n \in \mathbb{N} \wedge n \geq 5) \rightarrow (n^2 < 2^n)]$.

Las hemos expresado en nivel creciente de formalización. En la última aparece el símbolo $\forall$ (para todo), que se denomina *cuantificador universal*.

Ejemplo 0.3.1. Las siguientes expresiones son todas aceptables a la hora de usar el cuantificador universal:

Todo elemento del conjunto B es negativo.
Para todo $x \in B$, $x < 0$.
$(\forall x \in B)(x < 0)$.
Para todo x, si $x \in B$, entonces $x < 0$.
$(\forall x)(x \in B \rightarrow x < 0)$.

Todas estas expresiones tiene su lugar en el discurso matemático, desde la primera, más informal, hasta la última. Es mejor pensar que la primera es una forma conversacional de la última, que no ésta como una formalización de la primera.

Ejemplo 0.3.2. Supongamos que f es una función. Las siguientes expresiones son diferentes formas de decir lo mismo:

La gráfica de f no corta al eje de abscisas.
Para todos los números reales x, $f(x) \neq 0$.
$(\forall x \in \mathbb{R})(f(x) \neq 0)$.
Para todo x, si $x \in \mathbb{R}$, entonces $f(x) \neq 0$.
$(\forall x)(x \in \mathbb{R} \rightarrow f(x) \neq 0)$.

0.3.2. El cuantificador existencial

La desigualdad $n^2 < 2^n$ es cierta para $n \geq 5$, pero no si $1 \leq n \leq 4$. En otras palabras, existen números naturales n (al menos uno) tal que $n^2 \geq 2^n$. Otras formas de expresarlo son:

Existe un número natural n tal que $n^2 \geq 2^n$.
Existe n tal que $n \in \mathbb{N}$ y $n^2 \geq 2^n$.
$(\exists n)(n \in \mathbb{N} \wedge n^2 \geq 2^n)$.

La expresión “existe", que en matemáticas se escribe como $\exists$, se denomina el *cuantificador existencial*.

Ejemplo 0.3.3. Las siguientes son formas aceptables de decir el mismo enunciado:

Algunos elementos del conjunto B son positivos.
Existe $x \in B$ tal que $x > 0$.
$(\exists x \in B)(x > 0)$.
Existe x tal que $x \in B$ y $x > 0$.
$(\exists x)(x \in B \wedge x > 0)$.

Ejemplo 0.3.4. Vamos a ver ahora una expresión algo más compleja. Supongamos que F es un conjunto de funciones. Aquí tenemos tres formas de decir lo mismo:

La gráfica de toda función de F corta al eje de abscisas al menos una vez.
Para toda función $f \in F$, existe $x \in \mathbb{R}$ tal que $f(x) = 0$.
$(\forall f \in F)(\exists x \in \mathbb{R})(f(x) = 0)$.
$(\forall f)[(f \in F) \rightarrow [(\exists x)(x \in \mathbb{R} \wedge f(x) = 0)]$.

Algunas veces expresamos, de manera informal, la condición “para todo" detrás de la propiedad que es universal. Es una forma de hacer que la expresión suene más natural. Por ejemplo,

Existe un elemento del conjunto A que es menor que todo elemento del conjunto B.
Existe $x \in A$ tal que $x < y$ para todo $y \in B$.
Existe $x \in A$ tal que, para todo $y \in B$, $x < y$.
$(\exists x \in A)(\forall y \in B)(x < y)$.

0.4. Negación de expresiones

La característica que define la negación de una expresión es que los valores de la tabla de verdad son los opuestos. En esta sección vamos a construir las negaciones de expresiones compuestas. Por ejemplo, si alguien dice una expresión de la forma

$$(p \wedge q) \rightarrow (q \vee r),$$

y queremos decir que no es cierto, tendremos que escribir

$$\neg[(p \wedge q) \rightarrow (q \vee r)]$$

de una forma más legible.

0.4.1. Negación de $\wedge$ y $\vee$

La forma de negar las expresiones de conjunción y disyunción es mediante las leyes de Morgan:

$$\neg(p \wedge q) \equiv \neg p \vee \neg q, \neg(p \vee q) \equiv \neg p \wedge \neg q.$$

Ejemplo 0.4.1. Vamos a usar las leyes de Morgan para expresar la negación de los siguientes enunciados.

1. Juan tiene ojos azules y pelo castaño.
2. O estoy loco o hay un elefante rosa volando por aquí.
3. Marta tiene al menos 25 años, tiene carnet de conducir, y, o bien ella tiene su propio seguro, o bien ha comprado la asistencia de la compañía de alquiler de coches.

1. O Juan no tiene ojos azules o bien no tiene el pelo castaño.
2. No estoy loco y no hay un elefante rosa volando por aquí.
3. O Marta es menor de 25 años, o no tiene carnet de conducir, o ella no tiene su propio seguro y no ha comprado la asistencia de la compañía de alquiler de coches.

0.4.2. Negación de la implicación

Recordemos que $p \to q$ es equivalente a $\neg p \vee q$. Entonces, con las leyes de Morgan,

$$\neg(p \to q) \equiv \neg(\neg p \vee q) \equiv p \wedge \neg q.$$

Esto puede resultar un poco confuso al principio, pero lo podemos aclarar si pensamos de la siguiente forma. Si tenemos un enunciado que nos dice que p implica a q, entonces nos está diciendo que la verdad de p viene acompañada de la verdad de q. Si negamos eso, decimos que p puede ser verdad, mientras que q es falsa. Este es el fundamento de una demostración por reducción al absurdo.

0.4.3. Negación del cuantificador universal

Supongamos que hacemos la siguiente afirmación:

Todas las personas de la clase aprobaron el primer parcial.

La negación de esto sería algo de la forma

Al menos una persona de la clase no aprobó el primer parcial.

Vamos a ponerlo más formal. Sea C el conjunto de estudiantes de la clase, y P el conjunto de estudiantes de la clase que han aprobado el primer parcial. La afirmación original es

$$(\forall x)(x \in C \to x \in P).$$

La negación es

$$(\exists x)(x \in C \wedge x \notin P).$$

Si notamos por $P(x)$ una propiedad del elemento x, podemos decir lo siguiente:

$$\neg[(\forall x)(P(x))] \equiv (\exists x)(\neg P(x)).$$

Por tanto, el truco para negar un cuantificador universal es que el símbolo $\neg$ cambia el $\forall x$ en $\exists x$ y pasa a negar la propiedad.

Ejemplo 0.4.2. Una expresión matemática como

$$\text{Si } x > 1, \text{ entonces } x^3 - x > 0$$

es una simplificación aceptable de la más formal

$$\text{Para todo } x, \text{ si } x \in \mathbb{R} \text{ y } x > 1, \text{ entonces } x^3 - x > 0.$$

Esta versión nos ayuda a construir la negación:

$$\text{Existe } x \text{ un número real tal que } x > 1 \text{ y } x^3 - x \leq 0.$$

0.4.4. Negación del cuantificador existencial

El tratamiento es muy similar al anterior, y la regla general es

$$\neg[(\exists x)(P(x))] \equiv (\forall x)(\neg P(x)).$$

Ejemplo 0.4.3. Construyamos las negaciones de las siguientes expresiones.

1. $(\exists x \in \mathbb{N})(x \leq 0)$.
2. $(\forall \epsilon > 0)(\exists n \in \mathbb{N})(1/n < \epsilon)$.

Con las reglas anteriores, obtenemos

1. $(\forall x \in \mathbb{N})(x > 0)$.
2. $(\exists \epsilon > 0)(\forall n \in \mathbb{N})(1/n \geq \epsilon)$.

0.5. Inducción matemática

En el conjunto $\mathbb{N}$ de los números naturales se tiene el principio de buena ordenación: todo subconjunto no vacío S de $\mathbb{N}$ tiene un primer elemento. A partir de aquí deduciremos el principio de inducción matemática.

Consideremos el siguiente ejemplo de suma de una progresión aritmética. Queremos calcular la suma de los 100 primeros números naturales. Una forma, claro está, es realizar la suma con paciencia. Otro método es el siguiente. Escribamos la suma dos veces, pero la segunda en orden inverso, y sumamos verticalmente:

$$\begin{array}{rcrcrcrcrcrcr}
1 & + & 2 & + & 3 & + & 4 & + & \ldots & + & 99 & + & 100 \\
100 & + & 99 & + & 98 & + & 97 & + & \ldots & + & 2 & + & 1 \\
\hline
101 & + & 101 & + & 101 & + & 101 & + & \ldots & + & 101 & + & 101
\end{array}$$

Hay 100 términos, por lo que el doble de la suma es $100 \times 101 = 10100$, y la suma es 5050.

Lo anterior es un buen método para obtener incluso una fórmula general, pero hay algo en la prueba que nos puede dejar en duda. Se refiere a los puntos suspensivos que colocamos para representar la suma. El principio de inducción matemática elimina esta ambigüedad.

Supongamos que tenemos un conjunto S, subconjunto de $\mathbb{N}$, y que tiene las siguientes propiedades:

- $1 \in S$.

- Si $n \geq 1$ y $n \in S$, entonces $n+1 \in S$.

¿Qué es S en realidad? La primera propiedad nos dice que $1 \in S$, pero la aplicación de la segunda nos garantiza que $2 = 1+1 \in S$. Si aplicamos de nuevo la segunda propiedad tenemos que $3 = 2+1 \in S$, y así sucesivamente. Como $S \subset \mathbb{N}$, y parece que todo elemento de $\mathbb{N}$ se puede alcanzar por aplicación reiterada de la segunda propiedad, tenemos que $\mathbb{N} \subset S$. Parece claro, pero de nuevo repetimos el argumento de los puntos suspensivos. El principio de buena ordenación nos permite probar la inducción matemática.

Principio de inducción

Supongamos que S es un subconjunto de $\mathbb{N}$ con las siguientes propiedades:

- $1 \in S$.
- Si $n \geq 1$ y $n \in S$, entonces $n+1 \in S$.

Entonces $S = \mathbb{N}$.

PRUEBA: La prueba es por reducción al absurdo, y nos va a servir para poner en funcionamiento los métodos que hemos estudiado. Supongamos que $S \neq \mathbb{N}$. Como S es un subconjunto de $\mathbb{N}$, esto quiere decir que existe $n \in \mathbb{N}$ tal que $n \notin S$. Sea $T = \mathbb{N} - S$, el conjunto de elementos de $\mathbb{N}$ que no están en S. Por lo que estamos suponiendo, T es un conjunto no vacío, subconjunto de $\mathbb{N}$. Por el principio de buena ordenación, contiene un primer elemento, que llamaremos a.

Sabemos que no es posible que $a = 1$, por la primera propiedad. Así, $a > 1$. Entonces $a-1 \in \mathbb{N}$, y como a es el primer elemento de T, se tiene que $a-1 \notin T$. Por tanto, $a-1 \in S$. Pero si aplicamos la segunda propiedad de S, resulta que $(a-1)+1 \in S$, luego $a \in S$. Esto es una contradicción, por lo que T es un conjunto vacío. □

Vamos a ver un ejemplo donde aplicar el principio de inducción matemática.

Ejemplo 0.5.1. Para todo $n \geq 1$,

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

En primer lugar, vamos a ver que se verifica para $n = 1$. En efecto,

$$\sum_{k=1}^{1} k^2 = 1 = 1(1+1)(2\cdot 1+1)/6,$$

y tenemos el primer caso. Supongamos que $n \geq 1$ y que se tiene la propiedad para n, esto es,

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Vamos a probar el resultado para $n+1$. Tenemos que

$$\begin{aligned}\sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n+1)^2 = \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \\ &= \frac{2n^3+3n^2+n+6(n^2+2n+1)}{6} = \frac{2n^3+9n^2+13n+6}{6} \\ &= \frac{(n+1)(n+2)(2n+3)}{6} = \frac{(n+1)(n+2)[2(n+1)+1]}{6}.\end{aligned}$$

La propiedad es cierta para $n+1$, y entonces se verifica para todo $n \geq 1$.

Ejemplo 0.5.2. Supongamos que hay n personas en una habitación, y cada una saluda a otra persona una sola vez. Entonces el número de saludos en total es $n(n-1)/2$.

Si hay una sola persona en la habitación, no saluda a nadie. Se verifica entonces la fórmula para $n = 1$, pues $0 = 1(1-1)/2$. Sea entonces $n \geq 1$, y supongamos que hay $n+1$ personas en una habitación. Saquemos a una persona del lugar. Si se saludan las que quedan, habrá $n(n-1)/2$ saludos. Ahora traemos a la persona que habíamos sacado. Cuando salude a las que están dentro, contaremos n saludos más. Por tanto, el número total de saludos entre las $n+1$ personas es

$$\frac{n(n-1)}{2} + n = \frac{n^2-n+2n}{2} = \frac{n^2+n}{2} = \frac{n(n+1)}{2},$$

que es la fórmula para $n+1$.

Hay diferentes variaciones de la inducción estándar.

Variaciones sobre la inducción

Supongamos que S es un subconjunto de $\mathbb{Z}$ que verifica

- Existe un entero $j \in S$.
- Si $n \geq j$ y $n \in S$, entonces $n+1 \in S$.

Entonces $S = \{n \in \mathbb{Z} : n \geq j\}$.

La prueba es exactamente igual que la del principio de inducción. Esta forma nos permite variar el punto de partida, y no estar anclados al valor 1. Otra forma de inducción es la que se denomina inducción fuerte.

Principio de inducción fuerte

Supongamos que $S \subset \mathbb{N}$ es un conjunto que verifica

- $1 \in S$.
- Si $n \geq 2$ y $1, 2, \ldots, n-1 \in S$, entonces $n \in S$.

Entonces $S = \mathbb{N}$.

Si tenemos que probar algo por inducción, nos puede pasar que la inducción normal no funcione bien al hacer el paso de n a $n+1$, y lo que funcione sean las condiciones que nos indican que $1, 2, \ldots, n$ son elementos de S.

0.6. Funciones

0.6.1. Definición

Dados dos conjuntos A y B, una función f es una regla o conjunto de instrucciones mediante la cual cada elemento de A es emparejado con *exactamente* un elemento de B. El conjunto A se denomina dominio de la función, y a B el codominio. Lo escribiremos como $f : A \to B$. Si $x \in A$ se empareja con $y \in B$ lo notaremos por $y = f(x)$ o $x \mapsto y$. Decimos que y es la *imagen* de x, o que x se aplica en y, y que x es una pre-imagen de y. El subconjunto de B que consiste de las imágenes de los elementos de A se denomina imagen de f.

Veamos los detalles de esta definición. En primer lugar, para que una correspondencia $f\colon A \to B$ sea una función, todo elemento $x \in A$ debe tener alguna imagen $y \in B$. En otras palabras,

(F1) Para cada $x \in A$, existe $y \in B$ tal que $f(x) = y$.

Además, la regla que define a f debe producir un único $f(x)$ para todo $x \in A$. Decimos que f está bien definida si $f(x)$ es único para todo $x \in A$. Esto lo expresamos así:

(F2) Si $y_1, y_2 \in B$ son tales que $f(x) = y_1$ y $f(x) = y_2$, entonces $y_1 = y_2$.

Otra forma de formular esta condición es

(F2)' Si $x_1, x_2 \in A$ y $x_1 = x_2$, entonces $f(x_1) = f(x_2)$.

0.6.2. Aplicaciones inyectivas y sobreyectivas

Tipos de funciones

Sea $f\colon A \to B$ una función.

- Decimos que f es ***sobreyectiva*** si para cada $b \in B$ existe $a \in A$ tal que $f(a) = b$. A veces abreviamos diciendo que f es sobre.
- Decimos que f es ***inyectiva*** si dado $f(x_1) = f(x_2)$ implica que $x_1 = x_2$, para $x_1, x_2 \in A$. También es posible usar la forma negativa de esta condición: si $x_1 \neq x_2$ entonces $f(x_1) \neq f(x_2)$.

0.6.3. Imagen y pre-imagen

Imagen de un conjunto

Sea $f\colon A \to B$ una función, y $A_1 \subset A$. Entonces la *imagen* de A_1 es

$$f(A_1) = \{y \in B \mid (\exists x \in A_1)(y = f(x))\} = \{f(x) \mid x \in A_1\}.$$

Para el caso $A_1 = A$ escribiremos $f(A) = \mathrm{im}(f)$.

Preimagen o imagen inversa

Sea $f\colon A \to B$ una función, y $B_1 \subset B$. Definimos la *preimagen* o imagen inversa de B_1 como el conjunto

$$f^{-1}(B_1) = \{x \in A \mid f(x) \in B_1\}.$$

Si $B_1 = \{y\}$ contiene un solo elemento, escribiremos $f^{-1}(y)$ en lugar de $f^{-1}(\{y\})$.

0.6.4. Composición y función inversa

Supongamos que $f\colon A \to B$ y $g\colon B \to C$ son funciones. Definimos la composición $g \circ f\colon A \to C$ como la aplicación definida por $(g \circ f)(x) = g(f(x))$ para todo $x \in A$.

Es fácil ver que $g \circ f$ es una función (verifica las condiciones F1 y F2). Observemos que $f \circ g$ puede que no tenga sentido, pues C no tiene que coincidir con A. Aunque los conjuntos permitieran la composición, en general $g \circ f \neq f \circ g$.

Si $f\colon A \to B$ es una función, entonces la regla que une estos dos conjuntos está bien definida (regla F2) sobre todo A (regla F1). Vemos ahora la cuestión desde el punto de vista de B. ¿Podemos encontrar una aplicación $g\colon B \to A$ cuya regla de asignación sea la opuesta a la de f? Por ejemplo, si $f(8) = 2$, queremos que $g(2) = 8$.

Función inversa

Sea $f\colon A \to B$ una función. Decimos que $g\colon B \to A$ es una inversa de f si $(g \circ f)(x) = x$ para todo $x \in A$, y $(f \circ g)(y) = y$ para todo $y \in B$. A tal función g la notamos por f^{-1}.

Debemos tener mucho cuidado con la notación. Si $B_1 \subset B$, representamos por $f^{-1}(B_1)$ el conjunto pre-imagen de B_1. Esto siempre existe, y puede ser el conjunto vacío. Esta notación no hay que confundirla con la función f^{-1}, que puede que no exista.

Existencia de la función inversa

Supongamos que $f\colon A \to B$ es una función inyectiva y sobreyectiva. Entonces existe una única función inversa $f^{-1} : B \to A$, que también es inyectiva y sobreyectiva.

Demostración. Lo primero que tenemos que hacer es definir cómo actúa g. Sea $y \in B$. Por el carácter sobreyectivo de f, existe $x \in A$ tal que $f(x) = y$. Pero por el carácter inyectivo de f, sabemos que tal x es único. Definimos entonces $g(y) = x$, y vamos a comprobar que es una función.

(F1) Para cada $y \in B$, existe $x \in A$ tal que $g(y) = x$, o lo que es lo mismo, que g está definida en todo B. Esto lo tenemos por la construcción de g.

(F2) Si $x_1, x_2 \in A$ son tales que $g(y) = x_1$ y $g(y) = x_2$, entonces $x_1 = x_2$. Si $g(y) = x_1$, entonces $f(x_1) = y$, y, análogamente, $f(x_2) = y$. Entonces $f(x_1) = f(x_2)$, y por la inyectividad de f se deduce que $x_1 = x_2$.

En consecuencia, g es una función. Vamos a probar ahora que es inyectiva y sobre.

- g es inyectiva. Sean $y_1, y_2 \in B$ tales que $g(y_1) = g(y_2) = x$. Entonces $f(x) = y_1$ y $f(x) = y_2$. Como f es función, se tiene que $y_1 = y_2$.
- g es sobre. Sea $x \in A$. Entonces existe $y \in B$ tal que $f(x) = y$, pues f está definida en todo el conjunto A. Por la construcción de g, se verifica que $g(y) = x$.

Ahora queda ver que g cumple las condiciones de función inversa.

- $(g \circ f)(x) = x$ para todo $x \in A$. Sea $y = f(x)$. Entonces, por la definición de g, sabemos que $(g(f(x)) = g(y) = x$, que es el resultado.
- $(f \circ g)(y) = y$ para todo $y \in B$. Sea $x = g(y)$. Entonces $f(x) = y$, o bien, $f(g(y)) = y$.

Lo único que nos queda probar es la unicidad de la función g. Para ello, supongamos que $h : B \to A$ es una función que satisface las condiciones $(h \circ f)(x) = x$ para todo $x \in A$, y $(f \circ h)(y) = y$ para todo $y \in B$. Como

$$(f \circ g)(y) = y = (f \circ h)(y) \text{ para todo } y \in B,$$

la inyectividad de f nos dice que $g(y) = h(y)$ para todo $y \in B$, o lo que es lo mismo, que $g = h$. □

0.7. Números complejos

Los números reales tienen una gran deficiencia: la de que no toda función polinómica tiene una raíz. El ejemplo más sencillo y notable es el hecho de que no existe ningún número real x tal que $x^2+1=0$. Desde hace mucho tiempo se inventó un número i con la propiedad de que $i^2+1=0$. La admisión de este número parecía simplificar muchos cálculos algebraicos, especialmente cuando se admitían los "números complejos" $a+bi$, para $a,b\in\mathbb{R}$, y se suponían válidas todas las leyes del cálculo aritmético.

Por ejemplo, la ecuación

$$x^2+x+1=0$$

carece de raíces reales, puesto que

$$x^2+x+1=(x+\frac{1}{2})^2+\frac{3}{4}>0, \text{ para todo } x.$$

Pero la fórmula de resolución de ecuaciones cuadráticas sugiere las "soluciones"

$$x=\frac{-1+\sqrt{-3}}{2}, x=\frac{-1-\sqrt{-3}}{2}.$$

Si interpretamos $\sqrt{-3}$ como $\sqrt{3\cdot(-1)}=\sqrt{3}\sqrt{-1}=3i$, entonces estos números serían

$$-\frac{1}{2}+\frac{\sqrt{3}}{2}i \text{ y } -\frac{1}{2}-\frac{\sqrt{3}}{2}i.$$

Es incluso posible "resolver" ecuaciones cuadráticas cuyos coeficientes son a su vez números complejos. Por ejemplo, la ecuación

$$x^2+x+1+i=0$$

admite las soluciones

$$x=\frac{-1\pm\sqrt{1-4(1+i)}}{2}=\frac{-1\pm\sqrt{-3-4i}}{2},$$

donde el símbolo $\sqrt{-3-4i}$ significa un número $\alpha+\beta i$ cuyo cuadrado es $-3-4i$.

Para nosotros, los números complejos son símbolos de la forma $a+bi$, a y b reales, con las siguientes operaciones:

1. $(a+bi)+(c+di)=(a+c)+(b+d)i$,

2. $(a+bi)(c+di)=(ac-bd)+(ad+bc)i$.

Vamos a dar nombres a cada parte. Si $z = a + bi$, decimos que a es la parte real de z, y lo notaremos por $a = \mathrm{Re}(z)$, y b es la parte imaginaria de z: $b = \mathrm{Im}(z)$. Al conjunto de números complejos lo llamaremos $\mathbb{C}$.

Los números reales son un subconjunto de los números complejos: son aquellos que tienen parte imaginaria igual a cero. Si $z = a + bi$ es un número complejo, entonces el conjugado $\bar{z}$ de z se define como

$$\bar{z} = a - bi,$$

y el módulo de z está definido por

$$|z| = \sqrt{a^2 + b^2}.$$

Los números complejos también se pueden dividir. Por ejemplo,

$$\frac{15+i}{6+3i} = \frac{(15+i)(6-3i)}{(6+3i)(6-3i)} = \frac{(90+3)+(-45+6)i}{36+9}$$
$$= \frac{93-39i}{45} = \frac{31}{15} - \frac{13}{15}i.$$

Se tienen las siguientes propiedades fundamentales:

Propiedades de los números complejos

Sean z y w números complejos. Entonces

1. $\bar{\bar{z}} = z$.
2. $\bar{z} = z$ si y solamente si z es un número real.
3. $\overline{z+w} = \overline{z} + \overline{w}$.
4. $\overline{-z} = -\overline{z}$.
5. $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$.
6. $\overline{z^{-1}} = \overline{z}^{-1}$, si $z \neq 0$.
7. $|z|^2 = z \cdot \overline{z}$.
8. $|z \cdot w| = |z| \cdot |w|$.

Demostración. Las demostraciones son muy simples. Por ejemplo, sean $z = a + bi$, $w = c + di$. Entonces

$$\overline{z \cdot w} = \overline{(a+bi)(c+di)} = \overline{(ac-bd)+(ad+bc)i}$$
$$= (ac-bd)-(ad+bc)i = (a-bi)(c-di) = \bar{z} \cdot \bar{w}.$$

De aquí se deduce fácilmente la última:

$$|z \cdot w|^2 = (z \cdot w)\overline{z \cdot w} = z \cdot w \cdot \bar{z} \cdot \bar{w} = (z \cdot \bar{z})(w \cdot \bar{w}) = |z|^2 |w|^2,$$

y como los módulos de números complejos son números no negativos, se sigue que $|z \cdot w| = |z| \cdot |w|$. □

Todo número complejo $z \neq 0$ se puede escribir como

$$z = |z| \frac{z}{|z|} = |z| \, u.$$

Es inmediato que $|u| = 1$, y si $u = \alpha + i\beta$, entonces $\alpha^2 + \beta^2 = 1$. Esto lo podemos escribir como

$$u = \alpha + i\beta = \cos\theta + i \sin\theta$$

para algún número θ. Así, todo número complejo z no nulo se puede escribir

$$z = r(\cos\theta + i \sin\theta)$$

para algún $r > 0$. El número r es único (igual a $|z|$), pero θ no es único. Si una posibilidad es θ_0, las demás son $\theta_0 + 2k\pi$, para $k \in \mathbb{Z}$. A este ángulo se le llama argumento de z.

En principio se había introducido el número i para resolver la ecuación $x^2 + 1 = 0$. El teorema fundamental del Álgebra afirma que con estos números podemos resolver cualquier ecuación polinómica: toda ecuación

$$z^n + a_{n-1} z^{n-1} + \cdots + a_0 = 0, a_0, \ldots, a_{n-1} \in \mathbb{C}$$

tiene una raíz compleja.

Un hecho que usaremos más adelante es que si $a_0, \ldots, a_{n-1}$ son reales, y $a + bi$, con $a, b \in \mathbb{R}$, satisface la ecuación

$$z^n + a_{n-1} z^{n-1} + \cdots + a_0 = 0,$$

entonces $a - ib$ satisface también esta ecuación. En efecto, sea $w = a + bi$. Entonces

$$w^n + a_{n-1} w^{n-1} + \cdots + a_0 = 0.$$

Si aplicamos conjugación a ambos lados de la igualdad, obtenemos

$$\overline{w^n} + \overline{a_{n-1} w^{n-1}} + \cdots + \overline{a_0} = 0,$$

pues el conjugado de la suma es la suma de conjugados. Lo mismo se aplica al producto, y recordemos que los números a_j son reales. Entonces $\overline{a_j} = a_j$, de donde queda

$$\overline{w}^n + a_{n-1} \overline{w}^{n-1} + \cdots + a_0 = 0.$$

Así, el conjugado de w también es raíz del polinomio. Esto significa que las raíces no reales de un polinomio con coeficientes reales se presentan siempre por pares, una raíz y su conjugada.

Capítulo 1

Sistemas de ecuaciones lineales

1.1. Introducción

Un problema fundamental que aparece en matemáticas es el análisis y resolución de m ecuaciones algebraicas con n incógnitas. El estudio de un sistema de ecuaciones lineales simultáneas está íntimamente ligado al estudio de una matriz rectangular de números definida por los coeficientes de las ecuaciones. Esta abstracción aparece desde el momento en que se trataron estos problemas.

El primer análisis registrado de ecuaciones simultáneas lo encontramos en el libro chino *Jiu zhang Suan-shu* (*Nueve Capítulos sobre las artes matemáticas*),(véase Carlos Maza y McTutor) escrito alrededor del 200 a.C. Al comienzo del capítulo VIII, aparece un problema de la siguiente forma:

Tres gavillas de buen cereal, dos gavillas de cereal mediocre y una gavilla de cereal malo se venden por 39 dou. Dos gavillas de bueno, tres mediocres y una mala se venden por 34 dou. Y una buena, dos mediocres y tres malas se venden por 26 dou. ¿Cuál es el precio recibido por cada gavilla de buen cereal, cada gavilla de cereal mediocre, y cada gavilla de cereal malo?

Hoy en día, este problema lo formularíamos como un sistema de tres ecuaciones con tres incógnitas:

$$\begin{aligned} 3x &+ 2y &+ z &= 39, \\ 2x &+ 3y &+ z &= 34, \\ x &+ 2y &+ 3z &= 26, \end{aligned}$$

donde x, y y z representan el precio de una gavilla de buen, mediocre y mal cereal, respectivamente. Los chinos vieron el problema esencial. Colocaron los coeficientes de este sistema, representados por cañas de bambú de color, como un cuadrado sobre un tablero de contar (similar a un ábaco), y manipulaban las

filas del cuadrado según ciertas reglas establecidas. Su tablero de contar y sus reglas encontraron su camino hacia Japón y finalmente aparecieron en Europa, con las cañas de color sustituidas por números y el tablero reemplazado por tinta y papel.

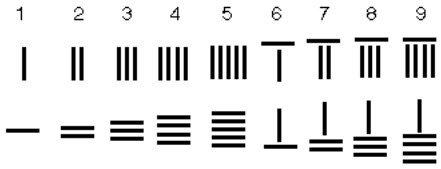

Figura 1.1: Numerales chinos con cañas de bambú

En Europa, esta técnica llegó a ser conocida como *eliminación gaussiana*, en honor del matemático alemán Carl F. Gauss, que popularizó el método.

Figura 1.2: C.F. Gauss (1828)

Como la técnica de eliminación es fundamental, empezamos el estudio de nuestra materia aprendiendo cómo aplicar este método para calcular las soluciones de los sistemas lineales. Después de que los aspectos computacionales se manejen bien, profundizaremos en cuestiones más teóricas.

1.2. Eliminación gaussiana y matrices

El problema es calcular, si es posible, una solución común a un sistema de m ecuaciones y n incógnitas de la forma

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

donde las x_i son las incógnitas y los a_{ij}, b_i son constantes (números reales o incluso complejos). Los números a_{ij} se denominan ***coeficientes*** del sistema, y el conjunto de los b_i ***términos independientes*** del sistema. Para estos sistemas, existen tres posibilidades:

- SOLUCIÓN ÚNICA: Existe uno y sólo un conjunto de valores para las incógnitas x_i que satisfacen las ecuaciones simultáneamente.
- INFINITAS SOLUCIONES: Existen infinitos conjuntos de valores para las incógnitas x_i que satisfacen las ecuaciones simultáneamente. Veremos más adelante que si el sistema tiene más de una solución, entonces tiene infinitas.
- SIN SOLUCIÓN: No hay ningún conjunto de valores para las incógnitas x_i que satisfagan todas las ecuaciones simultáneamente. El conjunto de soluciones es vacío.

Gran parte del trabajo acerca de los sistemas de ecuaciones es decidir cuál de estas tres posibilidades es la que se presenta. La otra parte de la tarea es calcular la solución si es única o describir el conjunto de soluciones si hay más de una. Incluso cuando no hay soluciones se puede hablar de *pseudosoluciones*. Esto lo trataremos cuando estudiemos mínimos cuadrados.

Ejemplo 1.2.1. Los casos anteriores los podemos encontrar ya en el sencillo ejemplo $ax = b$, que es una ecuación de primer grado. La podemos ver como un sistema de ecuaciones de una ecuación y una incógnita.

- La ecuación $3x = 1$ tiene solución única $x = \frac{1}{3}$.
- La ecuación $0 \cdot x = 0$ tiene infinitas soluciones.
- La ecuación $0 \cdot x = 1$ no tiene solución.

Podemos ampliar un poco la variedad con sistemas que resulten sencillos de resolver.

1. El sistema
$$\begin{cases} 2x & = & 1, \\ y & = & -1 \end{cases}$$
consiste en dos ecuaciones de primer grado, con solución $x = \frac{1}{2}, y = -1$.

2. El sistema
$$\begin{cases} 2x & +y & = & 1, \\ x & +\frac{1}{2}y & = & \frac{1}{2} \end{cases}$$
puede parecer más complicado, pero no lo es. Si multiplicamos la segunda ecuación por 2, obtenemos una ecuación igual a la primera, por lo que el conjunto de soluciones son todos los números x, y tales que $2x + y = 1$. Para cada valor de y existe un valor de x que verifica la ecuación.

3. El sistema
$$\begin{cases} 2x & +y & = & 1, \\ x & +\frac{1}{2}y & = & 0 \end{cases}$$
se parece mucho al anterior, pero al multiplicar la segunda ecuación por 2 obtenemos las relaciones
$$\begin{cases} 2x & +y & = & 1, \\ 2x & +y & = & 0 \end{cases}$$
Esto implica que $1 = 0$, por lo que no hay solución; el sistema es incompatible.

La eliminación gaussiana es una herramienta que nos permitirá tratar las diferentes situaciones de una manera ordenada. Es un *algoritmo* que sistemáticamente transforma un sistema en otro más simple, pero ***equivalente***, es decir, que posee el mismo conjunto de soluciones. La idea es llegar a un sistema lo más sencillo posible, eliminando variables, y obtener al final un sistema que sea fácilmente resoluble. Por ejemplo, uno diagonal para el caso $m = n$. El proceso de eliminación descansa sobre tres operaciones simples que transforman un sistema en otro equivalente. Para describir estas operaciones, sea E_k la k-ésima ecuación
$$E_k : a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = b_k$$
y escribamos el sistema como
$$\mathscr{S} \equiv \begin{Bmatrix} E_1 \\ E_2 \\ \vdots \\ E_m \end{Bmatrix}.$$
Dado un sistema lineal $\mathscr{S}$, cada una de las siguientes ***transformaciones elementales*** produce un sistema equivalente $\mathscr{S}'$.

1. Intercambio de las ecuaciones i-ésima y j-ésima. Esto es, si

$$\mathscr{S} \equiv \left\{ \begin{array}{c} E_1 \\ \vdots \\ E_i \\ \vdots \\ E_j \\ \vdots \\ E_m \end{array} \right\}, \text{ entonces } \mathscr{S}' \equiv \left\{ \begin{array}{c} E_1 \\ \vdots \\ E_j \\ \vdots \\ E_i \\ \vdots \\ E_m \end{array} \right\}.$$

2. Reemplaza la i-ésima ecuación por un múltiplo no nulo de ella. Esto es,

$$\mathscr{S}' \equiv \left\{ \begin{array}{c} E_1 \\ \vdots \\ \alpha E_i \\ \vdots \\ E_m \end{array} \right\}, \text{ donde } \alpha \neq 0.$$

3. Reemplaza la j-ésima ecuación por una combinación de ella misma más un múltiplo de la j-ésima ecuación. Esto es,

$$\mathscr{S}' \equiv \left\{ \begin{array}{c} E_1 \\ \vdots \\ E_i \\ \vdots \\ E_j + \alpha E_i \\ \vdots \\ E_m \end{array} \right\}.$$

Es fácil ver que estas operaciones no cambian el conjunto de soluciones.

El problema más común en la práctica es la resolución de un sistema con n ecuaciones y n incógnitas, lo que se conoce como un ***sistema cuadrado***, con solución única. En este caso, la eliminación gaussiana es directa, y más tarde estudiaremos las diferentes posibilidades. Lo que sigue es un ejemplo típico. Consideremos el sistema

$$\begin{array}{rcrcrcr} 2x & + & y & + & z & = & 1, \\ 6x & + & 2y & + & z & = & -1, \\ -2x & + & 2y & + & z & = & 7. \end{array} \tag{1.2.1}$$

En cada paso, la estrategia es centrarse en una posición, llamada ***posición pivote***, y eliminar todos los términos por debajo de la posición usando las tres operaciones elementales. El coeficiente en la posición pivote se denomina ***pivote***, mientras que la ecuación en donde se encuentra el pivote se llama ***ecuación pivotal***. Solamente se permiten números no nulos como pivotes. Si un coeficiente en una posición pivote es cero, entonces la ecuación pivotal se intercambia con una ecuación por *debajo* para producir un pivote no nulo. Esto siempre es posible para sistemas cuadrados con solución única. A menos que sea cero, el primer coeficiente de la primera ecuación se toma como el primer pivote. Por ejemplo, el elemento $\boxed{2}$ del sistema es el pivote del primer paso:

$$\begin{aligned} \boxed{2}x &+ y + z = 1, \\ 6x &+ 2y + z = -1, \\ -2x &+ 2y + z = 7. \end{aligned}$$

Paso 1. Elimina todos los términos por debajo del pivote.

- Resta tres veces la primera ecuación de la segunda para generar el sistema equivalente

$$\begin{aligned} \boxed{2}x &+ y + z = 1, \\ -y &- 2z = -4, \quad (E_2 - 3E_1) \\ -2x &+ 2y + z = 7. \end{aligned}$$

- Suma la primera ecuación a la tercera para formar el sistema equivalente

$$\begin{aligned} \boxed{2}x &+ y + z = 1, \\ -y &- 2z = -4, \\ 3y &+ 2z = 8 \quad (E_3 + E_1). \end{aligned}$$

Paso 2. Selecciona un nuevo pivote.

- De momento, seleccionamos un nuevo pivote buscando para abajo y a la derecha. Más adelante veremos una mejor estrategia. Si este coeficiente no es cero, entonces es nuestro pivote. En otro caso, intercambiamos con una ecuación que esté por *debajo* de esta posición para colocar el elemento no nulo en la posición pivote. En nuestro ejemplo, -1 es el segundo pivote:

$$\begin{aligned} 2x &+ y + z = 1, \\ \boxed{-1}y &- 2z = -4, \\ 3y &+ 2z = 8. \end{aligned}$$

Paso 3. Elimina todos los términos por debajo del pivote.

- Suma tres veces la segunda ecuación a la tercera para llegar al sistema equivalente:

$$\begin{array}{rcrcrcrl} 2x & + & y & + & z & = & 1, & \\ & & \boxed{-1}y & - & 2z & = & -4, & \\ & & & - & 4z & = & -4 & (E_3+3E_2). \end{array}$$

- En general, en cada paso nos movemos abajo y hacia la derecha para seleccionar el nuevo pivote, y entonces eliminar todos los términos por debajo de él hasta que ya no podamos seguir. En este ejemplo, el tercer pivote es -4, pero como ya no hay nada por debajo que eliminar, paramos el proceso.

En este punto, decimos que hemos ***triangularizado*** el sistema. Un sistema triangular se resuelve muy fácilmente mediante el método de ***sustitución hacia atrás***, en el que la última ecuación se resuelve para la última incógnita y se sustituye hacia atrás en la penúltima ecuación, la cual se vuelve a resolver para la penúltima incógnita, y continuamos así hasta llegar a la primera ecuación. En nuestro ejemplo, de la última ecuación obtenemos

$$z = 1.$$

Sustituimos $z = 1$ en la segunda ecuación, y tenemos

$$y = 4 - 2z = 4 - 2(1) = 2.$$

Por último, sustituimos $z = 1$ y $y = 2$ en la primera ecuación para obtener

$$x = \frac{1}{2}(1 - y - z) = \frac{1}{2}(1 - 2 - 1) = -1,$$

que completa la solución.

No hay razón para escribir los símbolos como x, y o z en cada paso, pues lo único que manejamos son los coeficientes. Si descartamos los símbolos, entonces el sistema de ecuaciones se reduce a una matriz rectangular de números en la que cada fila representa una ecuación. Por ejemplo, el sistema 1.2.1 se reduce a la siguiente matriz

$$\left(\begin{array}{rrr|r} 2 & 1 & 1 & 1 \\ 6 & 2 & 1 & -1 \\ -2 & 2 & 1 & 7 \end{array}\right) \text{(las barras indican donde aparece el signo } =).$$

La ***matriz de coeficientes*** está formada por los números a la izquierda de la línea vertical. La matriz completa, matriz de coeficientes aumentada por los términos de la derecha, se denomina ***matriz ampliada*** asociada al sistema. Si la

matriz de coeficientes se nota por A y el lado derecho por $\boldsymbol{b}$, entonces la matriz ampliada del sistema la escribiremos como $(A|\boldsymbol{b})$.

Un ***escalar*** es un número real o un número complejo, y una ***matriz*** es una disposición de escalares en rectángulo. Usaremos letras mayúsculas para las matrices y minúsculas con subíndice para las entradas individuales de la matriz. Así, escribiremos

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = (a_{ij}).$$

El primer subíndice de un elemento de la matriz indica la ***fila***, y el segundo subíndice denota la ***columna*** donde se encuentra. Por ejemplo, si

$$A = \begin{pmatrix} 2 & 1 & 3 & 4 \\ 8 & 6 & 5 & -9 \\ -3 & 8 & 3 & 7 \end{pmatrix}, \text{ entonces } a_{11} = 2, a_{12} = 1, \dots, a_{34} = 7. \tag{1.2.2}$$

Una ***submatriz*** de una matriz dada A es una matriz que se obtiene borrando un conjunto de filas y columnas de A. Por ejemplo,

$$B = \begin{pmatrix} 2 & 4 \\ -3 & 7 \end{pmatrix}$$

es una submatriz de A porque B es el resultado de borrar la segunda fila, y las columnas segunda y tercera de A.

Una matriz A se dice que tiene ***orden*** $m \times n$ cuando A tiene exactamente m filas y n columnas. La matriz A de (1.2.2) es una matriz 3×4. Por convenio, las matrices 1×1 se identifican con escalares, y al revés. Para enfatizar que una matriz A es de orden $m \times n$, usaremos la notación $A_{m\times n}$. Cuando $m = n$, es decir, cuando el número de filas y columnas coincide, diremos que la matriz es ***cuadrada***. En otro caso, la llamamos ***rectangular***. Las matrices que tienen una sola fila o una sola columna las llamaremos, respectivamente, ***vectores fila*** o ***vectores columna***.

El símbolo A_{i*} se usa para notar la fila i-ésima, y A_{*j} para la j-ésima columna. Por ejemplo, si A es la matriz de (1.2.2), entonces

$$A_{2*} = \begin{pmatrix} 8 & 6 & 5 & -9 \end{pmatrix} \text{ y } A_{*2} = \begin{pmatrix} 1 \\ 6 \\ 8 \end{pmatrix}.$$

Más adelante necesitaremos efectuar dos operaciones con las filas y las columnas, derivadas de las operaciones que hemos hecho en las ecuaciones. Una fila

de una matriz $A_{m\times n}$ es un elemento del producto cartesiano $\mathbb{K}^n$, donde $\mathbb{K}$ es el cuerpo base. Una columna de $A_{m\times n}$ es un elemento de $\mathbb{K}^m$. Podemos definir dos operaciones sobre las filas.

- Dado un escalar $\alpha \in \mathbb{K}$ y una fila $\boldsymbol{a} = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}$, definimos

$$\alpha\boldsymbol{a} = \begin{pmatrix} \alpha a_1 & \dots & \alpha a_n \end{pmatrix},$$

  es decir, multiplicamos cada componente de $\boldsymbol{a}$ por el escalar α.

- Dadas dos filas $\boldsymbol{a} = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}, \boldsymbol{a}' = \begin{pmatrix} a'_1 & \dots & a'_n \end{pmatrix}$, definimos

$$\boldsymbol{a} + \boldsymbol{a}' = \begin{pmatrix} a_1 + a'_1 & \dots & a_n + a'_n \end{pmatrix},$$

  esto es, sumamos componente a componente.

Las operaciones análogas se definen también para las columnas.

La eliminación gaussiana se puede realizar sobre la matriz ampliada $(A|\boldsymbol{b})$ mediante operaciones elementales sobre las filas de $(A|\boldsymbol{b})$. Estas operaciones elementales de filas se corresponden a las tres operaciones elementales que hemos visto antes.

Para una matriz de orden $m \times n$ de la forma

$$M = \begin{pmatrix} M_{1*} \\ \vdots \\ M_{i*} \\ \vdots \\ M_{j*} \\ \vdots \\ M_{m*} \end{pmatrix},$$

los tres tipos de ***operaciones elementales de filas*** sobre M son como sigue.

- Tipo I. Intercambio de filas i y j para dar

$$\begin{pmatrix} M_{1*} \\ \vdots \\ M_{j*} \\ \vdots \\ M_{i*} \\ \vdots \\ M_{m*} \end{pmatrix}. \tag{1.2.3}$$

- Tipo II. Reemplazo de la fila i por un múltiplo no nulo de ella para dar

$$\begin{pmatrix} M_{1*} \\ \vdots \\ \alpha M_{i*} \\ \vdots \\ M_{m*} \end{pmatrix}, \text{ donde } \alpha \neq 0. \tag{1.2.4}$$

- Tipo III. Reemplazo de la fila j por una combinación de ella más un múltiplo de la fila i para dar

$$\begin{pmatrix} M_{1*} \\ \vdots \\ M_{i*} \\ \vdots \\ M_{j*} + \alpha M_{i*} \\ \vdots \\ M_{m*} \end{pmatrix}. \tag{1.2.5}$$

Para resolver el sistema 1.2.1 mediante operaciones elementales por fila, partimos de la matriz ampliada $(A|\boldsymbol{b})$ y triangularizamos la matriz de coeficientes A realizando la misma secuencia de operaciones por fila que se corresponden a las operaciones elementales realizadas sobre las ecuaciones.

$$\left(\begin{array}{ccc|c} \boxed{2} & 1 & 1 & 1 \\ 6 & 2 & 1 & -1 \\ -2 & 2 & 1 & 7 \end{array}\right) \xrightarrow[F_3+F_1]{F_2-3F_1} \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & \boxed{-1} & -2 & -4 \\ 0 & 3 & 2 & 8 \end{array}\right)$$

$$\xrightarrow{F_3+3F_2} \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & -1 & -2 & -4 \\ 0 & 0 & -4 & -4 \end{array}\right).$$

La matriz final representa el sistema triangular

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

que se resuelve por sustitución hacia atrás, como explicamos antes. En general, si un sistema $n \times n$ se triangulariza a la forma

$$\left(\begin{array}{cccc|c} t_{11} & t_{12} & \dots & t_{1n} & c_1 \\ 0 & t_{22} & \dots & t_{2n} & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & t_{nn} & c_n \end{array}\right) \tag{1.2.6}$$

en donde cada $t_{ii} \neq 0$ (no hay pivotes nulos), entonces el algoritmo general de sustitución hacia atrás es como sigue.

Algoritmo de sustitución hacia atrás

Determina los x_i de 1.2.6 mediante $x_n = c_n / t_{nn}$ y procede de manera recursiva calculando

$$x_i = \frac{1}{t_{ii}}(c_i - t_{i,i+1}x_{i+1} - t_{i,i+2}x_{i+2} - \cdots - t_{in}x_n)$$

para $i = n-1, n-2, \dots, 2, 1$.

Ejemplo 1.2.2. Consideremos el sistema

$$\begin{array}{rcrcrcr} & & v & - & w & = & 3, \\ -2u & + & 4v & - & w & = & 1, \\ -2u & + & 5v & - & 4w & = & -2. \end{array}$$

La matriz ampliada es

$$\left(\begin{array}{ccc|c} 0 & 1 & -1 & 3 \\ -2 & 4 & -1 & 1 \\ -2 & 5 & -4 & -2 \end{array}\right).$$

Como la posición pivotal contiene el valor cero, intercambiamos las filas una y dos antes de comenzar la eliminación:

$$\left(\begin{array}{ccc|c} \boxed{0} & 1 & -1 & 3 \\ -2 & 4 & -1 & 1 \\ -2 & 5 & -4 & -2 \end{array}\right) \xrightarrow{F_{12}} \left(\begin{array}{ccc|c} \boxed{-2} & 4 & -1 & 1 \\ 0 & 1 & -1 & 3 \\ -2 & 5 & -4 & -2 \end{array}\right)$$

$$\xrightarrow{F_3 - F_1} \left(\begin{array}{ccc|c} -2 & 4 & -1 & 1 \\ 0 & \boxed{1} & -1 & 3 \\ 0 & 1 & -3 & -3 \end{array}\right)$$

$$\xrightarrow{F_3 - F_2} \left(\begin{array}{ccc|c} -2 & 4 & -1 & 1 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & -2 & -6 \end{array}\right).$$

La sustitución hacia atrás nos da

$$\begin{aligned} w &= \tfrac{-6}{-2} = 3, \\ v &= 3 + v = 3 + 3 = 6, \\ u &= \tfrac{1}{-2}(1 - 4v + w) = \tfrac{1}{-2}(1 - 24 + 3) = 10. \end{aligned}$$

Ejemplo 1.2.3. Resolvamos el sistema

$$\left\{ \begin{array}{rcrcrcr} 6x_1 & + & 4x_2 & + & 7x_3 & = & -1, \\ 3x_1 & + & 2x_2 & - & 5x_3 & = & 4, \\ 4x_1 & + & 2x_2 & - & 2x_3 & = & 5, \end{array} \right. .$$

Aplicamos eliminación gaussiana a la matriz ampliada $\left(\begin{array}{cc} A & \boldsymbol{b} \end{array} \right)$. El elemento $(1,1)$ es no nulo, y lo usamos como pivote para obtener valores nulos por debajo de él.

$$\left[\begin{array}{cccc} 6 & 4 & 7 & -1 \\ 3 & 2 & -5 & 4 \\ 4 & 2 & -2 & 5 \end{array} \right] \xrightarrow{F_2 - \frac{1}{2}F_1, F_3 - \frac{2}{3}F_1} \left[\begin{array}{cccc} 6 & 4 & 7 & -1 \\ 0 & 0 & -17/2 & 9/2 \\ 0 & -2/3 & -\frac{20}{3} & \frac{17}{3} \end{array} \right].$$

El elemento $(2,2)$ es nulo, por lo que buscamos un elemento no nulo por debajo de él, y lo encontramos en la posición $(3,2)$. Por ello, efectuamos el intercambio de las filas 2 y 3.

$$\left[\begin{array}{cccc} 6 & 4 & 7 & -1 \\ 0 & 0 & -17/2 & 9/2 \\ 0 & -2/3 & -\frac{20}{3} & \frac{17}{3} \end{array} \right] \overset{F_{23}}{\rightarrow} \left[\begin{array}{cccc} 6 & 4 & 7 & -1 \\ 0 & -2/3 & -\frac{20}{3} & \frac{17}{3} \\ 0 & 0 & -17/2 & 9/2 \end{array} \right] = T.$$

El elemento $(3,2)$ ya es nulo, y de esta forma hemos llegado a una forma triangular en la matriz de coeficientes. Recordemos que la matriz anterior representa el sistema

$$\left\{ \begin{array}{rcrcrcr} 6x_1 & + & 4x_2 & + & 7x_3 & = & -1, \\ & - & \frac{2}{3}x_2 & - & \frac{20}{3}x_3 & = & \frac{17}{3}, \\ & & & - & \frac{17}{2} & = & \frac{9}{2}. \end{array} \right.$$

Ahora aplicamos sustitución hacia atrás:

$$\begin{aligned} x_3 &= \frac{t_{34}}{t_{33}} = \frac{\frac{9}{2}}{-\frac{17}{2}} = -\frac{9}{17}, \\ x_2 &= \frac{1}{t_{22}}(t_{24} - t_{23} \cdot x3) = \frac{1}{\frac{-2}{3}}(\frac{17}{3} + \frac{20}{3}x_3) = -\frac{109}{34}, \\ x_1 &= \frac{1}{t_{11}}(t_{14} - t_{13} \cdot x_3 - t_{12} \cdot x_2) = \frac{1}{6}(-1 - 7x_3 - 4x_2) = \frac{44}{17}. \end{aligned}$$

1.3. Complejidad

Una forma de medir la eficiencia de un algoritmo es contando el número de operaciones aritméticas que se realizan. A veces, este dato no es suficiente para medir la eficiencia de un algoritmo. Hasta ahora, muchos ordenadores ejecutan las instrucciones de forma secuencial, mientras que ya aparecen máquinas capaces de ejecutar instrucciones en paralelo, en donde múltiples tareas numéricas se pueden ejecutar simultáneamente. Un algoritmo que use paralelismo puede tener un mayor número de operaciones que otro secuencial, pero ejecutarse más rápidamente en una máquina que admita la ejecución de instrucciones en diferentes procesadores de manera simultánea.

Por diferentes razones, agrupamos por un lado el número de multiplicaciones/divisiones y por otro sumas/restas. Sin embargo, en muchos casos se agrupan y hablamos de ***número de operaciones en coma flotante*** o ***flops***.

Es posible contar el número de operaciones realizadas en la eliminación gaussiana y sustitución hacia atrás, para después comparar con otros algoritmos.

Comencemos con el cálculo del número de operaciones de la eliminación gaussiana. Escribamos la matriz de partida en la forma

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \dots & a_{2n} & a_{2,n+1} \\ \vdots & & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} & a_{n,n+1} \end{array}\right)$$

para un mejor tratamiento de los índice. En la primera fase hacemos ceros en la primera columna. Para ello, calculamos el multiplicador $m_2 = \frac{a_{21}}{a_{11}}$ (un producto), y luego realizamos las operaciones

$$a_{2j} - m_2 a_{1j}, j = 2, \dots, n+1,$$

que contienen n productos y n sumas. Observemos que empezamos con $j = 2$ porque para $j = 1$ sabemos que se va a obtener un cero en la posición $(2,1)$. Por tanto, en este paso tenemos $1+n$ productos y n sumas.

Lo mismo lo vamos a tener al procesar las filas hasta la n. Así, desde la fila 2 hasta la fila n realizamos en cada paso $n+1$ productos y n sumas, y en total, $(n-1)(n+1)$ productos y $(n-1)n$ sumas.

Tenemos entonces la matriz en la forma

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & a_{1,n+1} \\ 0 & a_{22} & \dots & a_{2n} & a_{2,n+1} \\ \vdots & & & & \\ 0 & a_{n2} & \dots & a_{nn} & a_{n,n+1} \end{array}\right).$$

Ahora, para cada k variando entre las filas 3 y n, tendremos que calcular un multiplicador $m_k = \frac{a_{k2}}{a_{22}}$ (un producto), y aplicar la fórmula $a_{kj} - m_k a_{2j}$, con $j = 3, \ldots, n+1$ ($(n-1)$ productos y $(n-1)$ sumas). En total, para hacer ceros en la segunda columna necesitamos $(n-2)n$ productos y $(n-2)(n-1)$ sumas.

Al procesar las dos últimas filas para hacer un cero en la posición $(n, n-1)$, realizamos $1\cdot$ sumas y $1\cdot 3$ productos. En definitiva, el proceso de eliminación gaussiana precisa $\sum_{k=1}^{n-1} k(k+1)$ sumas, y $\sum_{k=1}^{n-1} k(k+2)$ productos.

Ahora vamos a calcular el número de operaciones necesarias para la sustitución hacia atrás. Comenzamos con $x_n = \frac{a_{n,n+1}}{a_{nn}}$, que es un producto. De la fórmula

$$x_i = \frac{1}{a_{ii}}(a_{i,n+1} - a_{i,i+1}x_{i+1} - a_{i,i+2}x_{i+2} - \cdots - a_{in}x_n), i = n-1, n-2, \ldots, 2, 1,$$

vemos que para cada i se necesitan $n-i$ productos y $n-i$ sumas. En total, en el primer paso 0 sumas y 1 producto, en el segundo 1 suma y 2 productos, hasta el último, con $n-1$ sumas y n productos, es decir, $\sum_{k=1}^{n-1} k$ sumas y $\sum_{k=1}^{n} k$ productos.

Si sumamos la parte de la eliminación en la matriz con la sustitución, nos queda:

Sumas

$$\begin{aligned}\sum_{k=1}^{n-1} k(k+1) + \sum_{k=1}^{n-1} k &= \sum_{k=1}^{n-1} k^2 + 2\sum_{k=1}^{n-1} k \\ &= \frac{1}{6}(n-1)n(2n-1) + (n-1)n = \frac{1}{3}n^3 + \frac{1}{2}n^2 - \frac{5}{6}n.\end{aligned}$$

Productos

$$\begin{aligned}\sum_{k=1}^{n-1} k(k+2) + \sum_{k=1}^{n} k &= \sum_{k=1}^{n} k^2 + 2\sum_{k=1}^{n} k + \sum_{k=1}^{n} k \\ &= \frac{1}{6}(n-1)n(2n-1) + (n-1)n + \frac{1}{2}n(n+1) \\ &= \frac{1}{3}n^3 + n^2 - \frac{1}{3}n.\end{aligned}$$

Número de operaciones en la eliminación gaussiana

La eliminación gaussiana con sustitución hacia atrás en un sistema $n \times n$ requiere

$$\frac{n^3}{3} + n^2 - \frac{n}{3} \text{ multiplicaciones/divisiones}$$

y

$$\frac{n^3}{3} + \frac{n^2}{2} - \frac{5n}{6} \text{ sumas/restas.}$$

Cuando n crece, el término $n^3/3$ domina estas expresiones. Por tanto, lo importante a recordar es que la eliminación gaussiana con sustitución hacia atrás sobre un sistema $n \times n$ precisa alrededor de $n^3/3$ multiplicaciones/divisiones y sobre el mismo número de sumas/restas. O bien, mediante agrupación, decimos que la eliminación gaussiana tiene un coste del orden de $2n^3/3$ flops.

1.4. Método de Gauss-Jordan

En esta sección introducimos una variante de la eliminación gaussiana, conocida como método de Gauss-Jordan. Aunque hay confusión con respecto al nombre, este método fue usado por Wilhelm Jordan (1842-1899), profesor de geodesia alemán, y no por Camille Jordan (1838-1922), matemático francés de quien hablaremos más adelante. Las características que distinguen el método

Figura 1.3: Wilhelm Jordan (1842-1899)

de Gauss-Jordan de la eliminación gaussiana son como sigue:

- En cada paso, el elemento pivote tiene que ser 1.

- En cada paso, todos los términos por *encima* del pivote así como todos los que están por debajo deben ser anulados.

En otras palabras, si

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & b_n \end{array}\right)$$

es la matriz ampliada del sistema, entonces mediante operaciones elementales la reducimos a

$$\left(\begin{array}{cccc|c} 1 & 0 & \dots & 0 & s_1 \\ 0 & 1 & \dots & 0 & s_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & s_n \end{array}\right).$$

La solución aparece en la última columna ($x_i = s_i$), por lo que no es necesaria la sustitución hacia atrás.

Ejemplo 1.4.1. Apliquemos Gauss-Jordan al siguiente sistema:

$$\begin{array}{rcrcrcr} 2x_1 & + & 2x_2 & + & 6x_3 & = & 4, \\ 2x_1 & + & x_2 & + & 7x_3 & = & 6, \\ -2x_1 & - & 6x_2 & - & 7x_3 & = & -1. \end{array}$$

La sucesión de operaciones se indican en cada paso, y se marca el pivote.

$$\left(\begin{array}{ccc|c} \boxed{2} & 2 & 6 & 4 \\ 2 & 1 & 7 & 6 \\ -2 & -6 & -7 & -1 \end{array}\right) \xrightarrow{F_1/2} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 3 & 2 \\ 2 & 1 & 7 & 6 \\ -2 & -6 & -7 & -1 \end{array}\right)$$

$$\xrightarrow[F_3+2F_1]{F_2-2F_1} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 3 & 2 \\ 0 & -1 & 1 & 2 \\ 0 & -4 & -1 & 3 \end{array}\right)$$

$$\xrightarrow{-F_2} \left(\begin{array}{ccc|c} 1 & 1 & 3 & 2 \\ 0 & \boxed{1} & -1 & -2 \\ 0 & -4 & -1 & 3 \end{array}\right)$$

$$\xrightarrow[F_3+4F_2]{F_1-F_2} \left(\begin{array}{ccc|c} 1 & 0 & 4 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & -5 & -5 \end{array}\right)$$

$$\xrightarrow{-F_3/5} \left(\begin{array}{ccc|c} 1 & 0 & 4 & 4 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & \boxed{1} & 1 \end{array}\right)$$

$$\xrightarrow[F_2+F_3]{F_1-4F_3} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & \boxed{1} & 1 \end{array}\right).$$

Por tanto, la solución es $x_1 = 0, x_2 = -1, x_3 = 1$.

En principio, puede parecer que hay poca diferencia entre la eliminación gaussiana con sustitución hacia atrás y el método de Gauss-Jordan, porque la eliminación de términos por encima del pivote en Gauss-Jordan parece equivalente a la sustitución hacia atrás. Pero esto no es correcto. El método de Gauss-Jordan necesita más operaciones aritméticas que la sustitución hacia atrás.

Número de operaciones de Gauss-Jordan

Para un sistema $n \times n$, el método de Gauss-Jordan necesita

$$\frac{n^3}{2} + \frac{n^2}{2} \text{ multiplicaciones/divisiones}$$

y

$$\frac{n^3}{2} - \frac{n}{2} \text{ sumas/restas.}$$

En otras palabras, el método de Gauss-Jordan requiere alrededor de $n^3/2$ multiplicaciones/divisiones y sobre el mismo número de sumas/restas, por lo que decimos que su coste total es del orden de n^3 flops.

En la última sección vimos que la eliminación de Gauss con sustitución hacia atrás usaba solamente unas $n^3/3$ multiplicaciones/divisiones, y alrededor del mismo número de sumas/restas. Si comparamos este número con el factor $n^3/2$ del método de Gauss-Jordan, vemos que este último require un esfuerzo adicional de un 50 %. Para sistemas pequeños, como los que aparecerán en los ejemplos ($n = 3,4$), estas comparaciones no muestran grandes diferencias. Sin embargo, en la práctica, los sistemas que se encuentran son bastante grandes, y la diferencia entre los dos métodos puede ser significativa. Por ejemplo, si $n = 100$, entonces $n^3/3 \approx 333333$, mientras que $n^3/2 \approx 500000$, que supone una diferencia de 166667 multiplicaciones/divisiones, así como de sumas/restas.

Aunque el método de Gauss-Jordan no es recomendable para resolver sistemas de ecuaciones en la práctica, tiene ciertas ventajas teóricas. Además, puede ser una técnica útil para tareas distintas a la resolución de sistemas de ecuaciones. Usaremos el método de Gauss-Jordan cuando tratemos la inversión de matrices.

1.5. La eliminación gaussiana en la práctica

Ahora que ya entendemos la técnica básica de la eliminación gaussiana, es momento de describir un algoritmo práctico para las aplicaciones reales. Para cálculos con lápiz y papel, donde hacemos aritmética exacta, la estrategia es mantener las cosas tan simples como sea posible, para así minimizar esos errores de cálculo que todos cometemos. Pero muy pocos problemas en el mundo real son de la clase que aparecen en los libros de texto, y las aplicaciones prácticas que involucran a sistemas lineales requieren el uso de un ordenador.

1.5.1. Coma flotante

Los ordenadores no se preocupan por lo complicada que pueda resultar una expresión con fracciones y no cometen errores al sumar números racionales o cambian un signo de manera inadvertida. Un ordenador genera un tipo más predecible de error, llamado de *redondeo*, y es importante gastar un poco de tiempo para entender este tipo de error y su efecto en la resolución de sistemas.

El cálculo numérico en los sistemas digitales se realiza aproximando el conjunto infinito de los números reales por un conjunto finito.

Números en coma flotante

Un ***número en coma flotante*** con t-dígitos y base β tiene la forma

$$f = \pm .d_1 d_2 \dots d_t \times \beta^{\epsilon} \text{ con } d_1 \neq 0,$$

donde la base β, el exponente ϵ y los dígitos $0 \leq d_i \leq \beta - 1$ son enteros. En la representación interna de la máquina, $\beta = 2$ (representación binaria), pero en los ejemplos de lápiz y papel es más conveniente usar $\beta = 10$. El valor de t, llamado la ***precisión***, y el exponente ϵ pueden variar con la máquina y el programa usado.

Por ejemplo, MAPLE tiene un valor inicial de 10 dígitos, aunque modificable. MATLAB trabaja internamente con 16 dígitos, y podemos parametrizar la salida.

Los números en coma flotante no son más que la adaptación del concepto familiar de notación científica con $\beta = 10$, que será el valor que usemos en nuestros ejemplos. Para un conjunto de valores fijados para t, β y ϵ, el conjunto correspondiente $\mathscr{F}$ de números en coma flotante es necesariamente finito, por lo que muchos números reales no los podremos encontrar en $\mathscr{F}$. Existe más de una forma de aproximar números reales con números en coma flotante. En todo lo que siga, usaremos el *redondeo* que a continuación describimos. Dado un número real x, la aproximación en coma flotante $\text{fl}(x)$ se define como el elemento de $\mathscr{F}$ más cercano a x, y en caso de empate entre dos opciones, elegimos la más lejana a cero. Esto significa que para una precisión de t dígitos con $\beta = 10$, tenemos que mirar el dígito d_{t+1} en

$$x = .d_1 d_2 \dots d_t d_{t+1} \cdots \times 10^{\epsilon}, \text{ con } d_1 \neq 0,$$

y entonces escribimos

$$\text{fl}(x) = \begin{cases} .d_1 d_2 \dots d_t \times 10^{\epsilon} & \text{si } d_{t+1} < 5, \\ ([.d_1 d_2 \dots d_t] + 10^{-t}) \times 10^{\epsilon} & \text{si } d_{t+1} \geq 5. \end{cases}$$

Por ejemplo, con 2 dígitos de precisión en aritmética decimal,

$$\mathrm{fl}(3/80) = \mathrm{fl}(0{,}0375) = \mathrm{fl}(0{,}375 \times 10^{-1}) = {,}38 \times 10^{-1} = 0{,}038.$$

Consideremos $\eta = 21/2$ y $\xi = 11/2$. Con 2 dígitos de precisión, en aritmética decimal,

$$\begin{aligned} \mathrm{fl}(\eta+\xi) &= \mathrm{fl}(32/2) = 16 = 0{,}16 \times 10^2, \\ \mathrm{fl}(\eta) + \mathrm{fl}(\xi) &= \mathrm{fl}(0{,}105 \times 10^2) + \mathrm{fl}(0{,}55 \times 10^1) = 0{,}11 \times 10^2 + 0{,}55 \times 10 \\ &= 0{,}165 \times 10^2 \xrightarrow{\mathrm{fl}} 0{,}17 \times 10^2. \end{aligned}$$

Por tanto, $\mathrm{fl}(\eta+\xi) \neq \mathrm{fl}(\mathrm{fl}(\eta)+\mathrm{fl}(\xi))$. Igualmente se puede comprobar que $\mathrm{fl}(\eta\xi) \neq \mathrm{fl}(\mathrm{fl}(\eta)\,\mathrm{fl}(\xi))$.

Además, otras propiedades habituales de la aritmética real no se verifican para la aritmética en coma flotante; por ejemplo, la propiedad asociativa es una de las más llamativas. Esto, entre otras razones, convierte el análisis del cálculo en coma flotante en algo difícil. También significa que hay que ser cuidadoso cuando se trabajen los ejercicios que se propongan, porque la mayor parte de calculadoras y ordenadores tienen una precisión interna fija con la que realizan todos los cálculos antes de mostrarlos en pantalla. La precisión interna de la calculadora es mayor que la precisión que usaremos para algunos ejemplos, por lo que cada vez que realice un cálculo con su calculadora y se pidan t dígitos de precisión, deberá efectuar el redondeo a mano, y reintroducir el número en la calculadora antes de seguir con el siguiente cálculo. En otras palabras, no encadene operaciones en su calculadora u ordenador.

Vamos a ver algunos ejemplos interesantes respecto a los problemas que encontraremos con la aritmética de coma flotante.

Ejemplo 1.5.1. 1. Adición de números positivos en orden ascendente. Supongamos que queremos efectuar la suma de los siguientes números, en aritmética de coma flotante de cuatro dígitos:

$$\begin{array}{lllll} 0{,}2897 \times 10^0, & 0{,}4976 \times 10^0, & 0{,}2488 \times 10^1, & 0{,}7259 \times 10^1, & 0{,}1638 \times 10^2, \\ 0{,}6249 \times 10^2, & 0{,}2162 \times 10^3, & 0{,}5233 \times 10^3, & 0{,}1403 \times 10^4, & 0{,}5291 \times 10^4. \end{array}$$

Si sumamos tal como aparecen en la lista (orden ascendente), obtenemos las siguientes sumas parciales:

$$\begin{array}{lllll} 0{,}7873 \times 10^0, & 0{,}3275 \times 10^1, & 0{,}1053 \times 10^2, & 0{,}2691 \times 10^2, & 0{,}8940 \times 10^2, \\ 0{,}3056 \times 10^3, & 0{,}8289 \times 10^3, & 0{,}2232 \times 10^4, & \mathbf{0{,}7523 \times 10^4}. & \end{array}$$

Si hacemos la suma en orden inverso (orden descendente), las sumas parciales son

$$\begin{array}{lllll} 0{,}6694 \times 10^4, & 0{,}7217 \times 10^4, & 0{,}7433 \times 10^4, & 0{,}7495 \times 10^4, & 0{,}7511 \times 10^4, \\ 0{,}7518 \times 10^4, & 0{,}7520 \times 10^4, & 0{,}7520 \times 10^4, & \mathbf{0{,}7520 \times 10^4}. & \end{array}$$

La suma correcta a ocho cifras se puede encontrar conservando todos los dígitos en cada suma, y es $0{,}75229043 \times 10^4$. Entonces el error en la suma ascendente es $-0{,}1 \times 10^0$, mientras que en la suma descendente es $2{,}9 \times 10^0$, unas 30 veces mayor.

2. Adición de números aproximadamente iguales. Consideremos los números
$$x_1 = 0{,}5243 \times 10^0, \quad x_2 = 0{,}5262 \times 10^0, \quad x_3 = 0{,}5226 \times 10^0,$$
$$x_4 = 0{,}5278 \times 10^0.$$
Si sumamos uno a continuación del otro, y con el redondeo en cada adición, obtenemos como valor de la suma $0{,}2102 \times 10^1$. Hay otra estrategia. Sumamos separadamente $x_1 + x_2 \overset{\text{fl}}{\to} 0{,}1051 \times 10^1$, y $x_3 + x_4 \overset{\text{fl}}{\to} 0{,}1050 \times 10^1$, y ahora estas dos cantidades, con resultado $0{,}2101 \times 10^1$. La suma exacta es $0{,}21009 \times 10^1$. En general, si deseamos sumar n^2 números positivos de aproximadamente igual magnitud, el error total por redondeo se reduce si se suman n grupos de n elementos cada uno, y después se suman las n sumas parciales.

3. Sustracción de dos números aproximadamente iguales. En esta diferencia, la cancelación puede hacer que los dígitos significativos desaparezcan, dejando en el resultado dígitos contaminados por los errores de redondeo. Supongamos $b = 3{,}34, a = 1{,}22, c = 2{,}28$, y queremos calcular $b^2 - 4ac$. Su valor exacto es 0,0292, pero
$$b^2 \overset{\text{fl}}{\to} 11{,}2, 4ac \overset{\text{fl}}{\to} 11{,}1, b^2 - 4ac \overset{\text{fl}}{\to} 0{,}1.$$

Una buena referencia sobre los efectos desastrosos que puede tener una inadecuada gestión de las excepciones en coma flotante se puede encontrar en

`http://www.ima.umn.edu/~arnold/disasters/ariane5rep.html`

Un interesante artículo sobre los fundamentos de la aritmética en coma flotante, y diferentes estrategias, se puede ver en [**?**].

Consejos sobre la coma flotante

- Cuando se van a sumar o restar números, empiece siempre con los números más pequeños (en valor absoluto).
- Evite la sustracción de dos números aproximadamente iguales. A menudo dicha expresión se puede reescribir para evitarla.

1.5.2. Aplicación a la eliminación gaussiana

Para entender cómo funciona la eliminación gaussiana con la aritmética de coma flotante, comparemos el uso de aritmética exacta con el cálculo con 3 dígitos de precisión sobre el siguiente sistema:

$$\begin{aligned} 47x + 28y &= 19, \\ 89x + 53y &= 36. \end{aligned}$$

En la eliminación gaussiana con aritmética exacta, multiplicamos la primera ecuación por el factor $m = 89/47$ y restamos el resultado a la segunda ecuación:

$$\left(\begin{array}{cc|c} 47 & 28 & 19 \\ 0 & -1/47 & 1/47 \end{array}\right).$$

Mediante sustitución hacia atrás, la solución *exacta* es $x = 1, y = -1$.

Con aritmética de 3 dígitos, el multiplicador es

$$\mathrm{fl}(m) = \mathrm{fl}(\frac{89}{47}) = 0{,}189 \times 10^1 = 1{,}89.$$

Observemos ahora la secuencia de operaciones:

$$\begin{aligned} \mathrm{fl}(\mathrm{fl}(m)\,\mathrm{fl}(47)) &= \mathrm{fl}(1{,}89 \times 47) = 0{,}888 \times 10^2 = 88{,}8, \\ \mathrm{fl}(\mathrm{fl}(m)\,\mathrm{fl}(28)) &= \mathrm{fl}(1{,}89 \times 28) = 0{,}529 \times 10^2 = 52{,}9, \\ \mathrm{fl}(\mathrm{fl}(m)\,\mathrm{fl}(19)) &= \mathrm{fl}(1{,}89 \times 19) = 0{,}359 \times 10^2 = 35{,}9. \end{aligned}$$

El primer paso de la eliminación gaussiana queda, con 3 dígitos de precisión,

$$\left(\begin{array}{cc|c} 47 & 28 & 19 \\ \mathrm{fl}(89-88{,}8) & \mathrm{fl}(53-52{,}9) & \mathrm{fl}(36-35{,}9) \end{array}\right) = \left(\begin{array}{cc|c} 47 & 28 & 19 \\ \boxed{,2} & ,1 & ,1 \end{array}\right).$$

El objetivo era triangular el sistema, y producir un cero en la posición $(2,1)$, pero esto no se puede hacer con aritmética de 3 dígitos. A menos que el valor $\boxed{,2}$ sea reemplazado por cero, la sustitución hacia atrás no se podrá llevar a cabo. Por tanto, acordaremos introducir 0 en la posición que estamos intentando anular, independientemente del valor que la aritmética de punto flotante haya dado. El valor de la posición que se quiere anular no se suele calcular. Así, no nos preocupamos de calcular

$$\mathrm{fl}(89 - \mathrm{fl}(\mathrm{fl}(m)\,\mathrm{fl}(47))) = \mathrm{fl}(89 - 88{,}8) = ,2$$

en el ejemplo anterior. Por tanto, el resultado de la eliminación gaussiana con 3 dígitos es

$$\left(\begin{array}{cc|c} 47 & 28 & 19 \\ 0 & ,1 & ,1 \end{array}\right).$$

Aplicamos sustitución hacia atrás, con la aritmética de 3 dígitos:

$$y = \mathrm{fl}(\frac{,1}{,1}) = 1,$$
$$x = \mathrm{fl}(\frac{19-28}{47}) = \mathrm{fl}(\frac{-9}{47}) = -,191.$$

La gran discrepancia entre la solución exacta $(1,-1)$ y la calculada con 3 dígitos de precisión $(-,191,1)$ ilustra algunos de los problemas con los que nos vamos a encontrar al resolver sistemas lineales con aritmética en coma flotante. En algunos casos, una precisión mayor puede ayudar, pero esto no es siempre posible porque en todas las máquinas hay un límite natural que convierte a la aritmética con precisión extendida no práctica a partir de un punto. Incluso si es posible incrementar la precisión, puede que no sea ventajoso porque hay casos en los que un incremento en precisión no produce una disminución comparable en los errores de redondeo. Dada una precisión particular t, no es difícil encontrar ejemplos de sistemas lineales para los que la solución calculada con t dígitos es tan mala como la de nuestro ejemplo con 3 dígitos.

Aunque los efectos del redondeo no pueden ser completamente eliminados, hay algunas técnicas sencillas que ayudan a minimizar estos errores inducidos por la máquina.

Pivoteo parcial

En cada paso, buscamos la posición desde la posición pivotal que contenga el coeficiente de mayor módulo. Si es necesario, realizamos el intercambio de filas adecuado para llevar este coeficiente máximo a la posición pivotal.

En principio, no parece claro por qué el pivoteo parcial debería mejorar la resolución del sistema. El siguiente ejemplo no solamente muestra que el pivoteo parcial puede mejorar mucho, sino que también indica qué hace esta estrategia efectiva.

Ejemplo 1.5.2. Es fácil ver que la solución exacta del sistema

$$\begin{aligned} -10^{-4}x + y &= 1, \\ x + y &= 2, \end{aligned}$$

es

$$x = \frac{1}{1{,}0001} \approx 0{,}99990000,\ y = \frac{1{,}0002}{1{,}0001} \approx 1{,}00009999.$$

Si usamos aritmética de 3 dígitos *sin* pivoteo parcial, el resultado es

$$\left(\begin{array}{cc|c} -10^{-4} & 1 & 1 \\ 1 & 1 & 2 \end{array}\right) \xrightarrow{F_2+10^4F_1} \left(\begin{array}{cc|c} -10^{-4} & 1 & 1 \\ 0 & 10^4 & 10^4 \end{array}\right),$$

porque

$$\mathrm{fl}(1+10^4) = \mathrm{fl}(,100001\times 10^5) = ,100\times 10^5 = 10^4 \tag{1.5.1}$$

y

$$\mathrm{fl}(2+10^4) = \mathrm{fl}(,100002\times 10^5) = ,100\times 10^5 = 10^4. \tag{1.5.2}$$

La sustitución hacia atrás nos da $x=0, y=1$.

Aunque la solución calculada para y es próxima a su solución exacta, la de x no es muy próxima al valor exacto. La solución calculada de x no se aproxima con tres dígitos significativos a la solución exacta. Si usamos aritmética de 3 dígitos, pero *con* pivoteo parcial, entonces

$$\left(\begin{array}{cc|c} -10^{-4} & 1 & 1 \\ 1 & 1 & 2 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 2 \\ -10^{-4} & 1 & 1 \end{array}\right) \xrightarrow{F_2+10^4F_1} \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1 \end{array}\right),$$

porque

$$\mathrm{fl}(1+10^{-4}) = \mathrm{fl}(,10001\times 10^1) = ,100\times 10^1 = 1 \tag{1.5.3}$$

y

$$\mathrm{fl}(1+2\times 10^{-4}) = \mathrm{fl}(,10002\times 10^1) = ,100\times 10^1 = 1. \tag{1.5.4}$$

Esta vez, la sustitución hacia atrás nos da la solución $x=1, y=1$, que es mucho más próxima a la solución exacta en la medida que uno puede razonablemente aceptar: la solución calculada coincide con la exacta en tres dígitos significativos.

¿Por qué el pivoteo parcial produce esta diferencia? La respuesta está en comparar las igualdades 1.5.1 y 1.5.2 con 1.5.3 y 1.5.4.

Sin pivoteo parcial, el multiplicador es 10^4, y tiene un valor tan grande que desborda la aritmética que implica a los números relativamente pequeños 1 y 2, y evita que sean tomados en cuenta. Esto es, los números más pequeños 1 y 2 son *borrados* como si no estuvieran presentes, de tal forma que nuestro ordenador de 3 dígitos calcula la solución exacta de otro sistema, a saber

$$\left(\begin{array}{cc|c} -10^{-4} & 1 & 1 \\ 1 & 0 & 0 \end{array}\right),$$

que es bastante diferente del sistema original. *Con* pivoteo parcial, el multiplicador es 10^{-4}, y es lo bastante pequeño para no solapar los números 1 y 2. En

este caso, el ordenador con 3 dígitos de precisión calcula la solución exacta del sistema

$$\left(\begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 1 & 2 \end{array}\right),$$

que es más próximo al original.

La respuesta a la pregunta "¿Qué sistema hemos resuelto, y cuán próximo es este sistema al original?" se denomina *análisis de error hacia atrás*, en oposición al análisis hacia delante, que intenta responder a la pregunta "¿Cómo de próxima será una solución calculada a la solución exacta?" El análisis hacia atrás ha demostrado ser un camino efectivo para analizar la estabilidad numérica de algoritmos.

El villano en el ejemplo anterior es el multiplicador tan grande que previene el tomar en cuenta a números más pequeños, por lo que tenemos la solución exacta de otro sistema que es muy diferente del original. Mediante el aumento de la magnitud del pivote en cada paso, minimizamos la magnitud del multiplicador asociado, lo que nos permite controlar el crecimiento de los números que surgen durante el proceso de eliminación. Esto ayuda a burlar algunos de los efectos del error de redondeo. El problema del crecimiento en el proceso de eliminación lo veremos más adelante, cuando estudiemos normas matriciales.

Cuando se usa pivoteo parcial, ningún multiplicador excede a 1 en magnitud. Para ver que esto es así, consideremos los siguientes pasos típicos en un proceso de eliminación:

$$\left(\begin{array}{ccccc|c} * & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & 0 & \boxed{p} & * & * & * \\ 0 & 0 & q & * & * & * \\ 0 & 0 & r & * & * & * \end{array}\right) \xrightarrow[F_5-(r/p)F_3]{F_4-(q/p)F_3} \left(\begin{array}{ccccc|c} * & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & 0 & \boxed{p} & * & * & * \\ 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * & * \end{array}\right).$$

El pivote es p, mientras que q/p y r/p son los multiplicadores. Si estamos usando pivoteo parcial, entonces $|p| \geq |q|$ y $|p| \geq |r|$, por lo que

$$\left|\frac{q}{p}\right| \leq 1 \text{ y } \left|\frac{r}{p}\right| \leq 1.$$

Con la garantía de que ningún multiplicador excede a 1 en magnitud, la posibilidad de crear números relativamente grandes que puedan eclipsar otros valores más pequeños es muy reducida. Para ver qué más se puede hacer, consideremos el siguiente ejemplo.

Ejemplo 1.5.3. La solución exacta del sistema

$$\begin{array}{rcrcr} -10x & + & 10^5 y & = & 10^5, \\ x & + & y & = & 2, \end{array}$$

es

$$x = \frac{1}{1{,}00001}, y = \frac{1{,}0002}{1{,}0001}.$$

Supongamos que se emplea aritmética de 3 dígitos con pivoteo parcial. Como $|-10| > 1$, no se necesita intercambio y obtenemos

$$\left(\begin{array}{cc|c} -10 & 10^5 & 10^5 \\ 1 & 1 & 2 \end{array}\right) \xrightarrow{F_2 + 10^{-1}F_1} \left(\begin{array}{cc|c} -10 & 10^5 & 10^5 \\ 0 & 10^4 & 10^4 \end{array}\right),$$

porque

$$\mathrm{fl}(1 + 10^4) = \mathrm{fl}(,10001 \times 10^5) = ,100 \times 10^5 = 10^4$$

y

$$\mathrm{fl}(2 + 10^4) = \mathrm{fl}(,10002 \times 10^5) = ,100 \times 10^5 = 10^4.$$

La sustitución hacia atrás nos da $x = 0, y = 1$, que debe ser considerada muy mala: la solución de y no es demasiado mala, pero la de x es terrible.

¿Qué ha pasado en el ejemplo anterior? En esta ocasión, no podemos echarle la culpa al multiplicador. El problema procede de que la primera ecuación contiene coeficientes que son mucho mayores que los coeficientes de la segunda. Esto es, hay un problema de *escala* debido a que los coeficientes son de diferentes órdenes de magnitud. Por ello, deberíamos reescalar el sistema antes de resolverlo.

Esto apunta a que el éxito del pivoteo parcial puede depender del mantenimiento de la escala adecuada entre los coeficientes. Así, el segundo refinamiento necesario para hacer la eliminación gaussiana práctica es una estrategia razonable de escalado. Por desgracia, no hay una tal estrategia que produzca óptimos resultados en todos los sistemas, por lo que debemos decidir por una estrategia que funcione la mayoría de las veces. La estrategia es combinar **escalado por filas** con **escalado por columnas**. La primera implica la comparación de términos en cada fila. El escalado por filas no afecta a la solución exacta, pero el escalado por columnas sí. Este escalado es equivalente a cambiar las unidades de medida de una incógnita. Por ejemplo, si las unidades en que medimos la incógnita x_k en $[A|\boldsymbol{b}]$ son milímetros, y la k-ésima columna de A se multiplica por ,001, entonces la k-ésima incógnita en el sistema escalado $[\hat{A}|\boldsymbol{b}]$ es $\hat{x}_k = 1000x_k$, y las unidades de $\hat{x}_k$ son ahora metros.

Escalado por columnas

Escoja unidades que sean naturales al problema y no distorsionen las relaciones entre los tamaños de las cosas. Estas unidades naturales son habitualmente claras, y un posterior escalado de columnas después de este punto no se suele hacer.

La experiencia muestra que el combinado de escalado por filas con escalado por columnas funciona habitualmente bastante bien. El pivoteo parcial junto a una estrategia de escalado hace a la eliminación gaussiana con sustitución hacia atrás una herramienta muy efectiva. A lo largo del tiempo, esta técnica ha demostrado ser fiable para resolver la mayoría de sistemas lineales encontrados en la práctica.

Vamos a describir dos técnicas de escalado por filas.

Escalado previo de filas y pivoteo parcial

1. Escale las filas del sistema $[A|\boldsymbol{b}]$ para que el coeficiente de mayor magnitud en cada fila de A sea igual a 1. Esto es, divida cada ecuación por el coeficiente de mayor valor absoluto de la matriz de coeficientes.
2. Aplique entonces pivoteo parcial al sistema resultante.

Ejemplo 1.5.4. Aplicamos escalado previo al sistema

$$\begin{cases} -10x & + & 10^5 y & = & 10^5, \\ x & + & y & = & 2, \end{cases}$$

con aritmética de 3 dígitos en coma flotante. La primera ecuación la dividimos por 10^5, y la segunda por 1. Tenemos entonces el sistema

$$\begin{cases} -10^{-4}x + y & = & 1, \\ x + y & = & 2, \end{cases}$$

Ahora aplicamos pivoteo parcial, que ya lo hemos detallado en el ejemplo 1.5.2.

En [**?**, secc. 3.5.2] se hacen diversos comentarios sobre el escalado de filas y columnas. Por ejemplo, sobre el escalado de filas apunta al método que hemos

usado de dividir cada fila por el elemento de mayor módulo. De esta forma se reduce la probabilidad de sumar un número muy pequeño a uno muy grande durante el proceso de eliminación. Igualmente, enfatiza que el simple escalado de filas y columnas no resuelve el problema, y que se debe proceder en una aproximación problema a problema. Deben considerarse tanto las unidades de medida como el error de los datos originales.

Otra aproximación diferente es la que se conoce como ***pivoteo escalado*** ([**?**, secc. 1.3.2.7], [**?**, secc. 2.4]). Se utiliza el escalado como el método de decisión del pivote, pero sin efectuar la división de cada fila por el término de mayor módulo, con el objetivo de no realizar muchas operaciones. Se implementa como sigue. Antes de aplicar la eliminación, se crea un vector con los cocientes de los elementos de la primera columna por el elemento de mayor módulo de su fila de coeficientes correspondiente. El pivote se escoge según el término de mayor módulo de este vector, y se realiza el intercambio de filas. A continuación, se efectúa la eliminación en la primera columna. De nuevo, antes de aplicar eliminación a la segunda columna, se construye un nuevo vector con los cocientes de los elementos de la segunda columna por el término de mayor módulo de su fila de coeficientes, entre 2 y n, y se escoge el pivote. Hemos de notar que las ecuaciones originales no se alteran.

Pivoteo escalado

1. En cada paso de la eliminación gaussiana, cree un vector con los cocientes de los elementos de la columna correspondiente por el elemento de mayor módulo de su filas de coeficientes correspondiente.
2. Coloque en la posición pivote la fila correspondiente al término de mayor módulo.
3. Aplique la eliminación de los elementos de la columna.

Ejemplo 1.5.5. Consideremos el sistema

$$\begin{cases} 3x_1 & +2x_2 & +105x_3 & = & 104, \\ 2x_1 & -3x_2 & +103x_3 & = & 98, \\ x_1 & +x_2 & +3x_3 & = & 3. \end{cases}$$

La solución exacta es $x_1 = -1{,}0, x_2 = 1{,}0$ y $x_3 = 1{,}0$. Aplicamos pivoteo parcial,

con aritmética de tres dígitos significativos en los cálculos.

$$\left(\begin{array}{ccc|c} 3 & 2 & 105 & 104 \\ 2 & -3 & 103 & 98 \\ 1 & 1 & 3 & 3 \end{array}\right) \xrightarrow{F_2-0{,}667F_1, F_3-0{,}333F_1} \left(\begin{array}{ccc|c} 3 & 2 & 105 & 104 \\ 0 & -4{,}33 & 33{,}0 & 28{,}6 \\ 0 & 0{,}334 & -32{,}0 & -31{,}6 \end{array}\right)$$

$$\xrightarrow{F_3+0{,}0771F_2} \left(\begin{array}{ccc|c} 3 & 2 & 105 & 104 \\ 0 & -4{,}33 & 33{,}0 & 28{,}6 \\ 0 & 0 & -29{,}5 & -29{,}4 \end{array}\right).$$

Mediante sustitución hacia atrás, nos queda $x_3 = 0{,}997$, $x_2 = 0{,}924$, $x_1 = -0{,}844$, que no casan muy bien con las soluciones exactas. Aplicamos ahora el pivoteo escalado. En primer lugar calculamos

$$\boldsymbol{a}_1 = \begin{pmatrix} \frac{3}{105} \\ \frac{2}{103} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 0{,}0286 \\ 0{,}0194 \\ 0{,}3333 \end{pmatrix}.$$

El tercer elemento de $\boldsymbol{a}_1$ es el de mayor módulo, lo que indica que las filas 1 y 3 se deben intercambiar, y luego procedemos a la eliminación.

$$\left(\begin{array}{ccc|c} 3 & 2 & 105 & 104 \\ 2 & -3 & 103 & 98 \\ 1 & 1 & 3 & 3 \end{array}\right) \xrightarrow{F_{13}} \left(\begin{array}{ccc|c} 1 & 1 & 3 & 1 \\ 2 & -3 & 103 & 98 \\ 3 & 2 & 105 & 104 \end{array}\right)$$

$$\xrightarrow{F_2-2F_1, F_3-3F_1} \left(\begin{array}{ccc|c} 1 & 1 & 3 & 3 \\ 0 & -5 & 97 & 92 \\ 0 & -1 & 96 & 95 \end{array}\right).$$

Ahora revisamos los elementos de la segunda columna, desde las posiciones 2 a 3. Calculamos

$$\boldsymbol{a}_2 = \begin{pmatrix} \frac{5}{97} \\ \frac{1}{96} \end{pmatrix} = \begin{pmatrix} 0{,}0516 \\ 0{,}0104 \end{pmatrix},$$

lo que indica que no es necesario el intercambio de filas. Realizamos entonces la eliminación

$$\left(\begin{array}{ccc|c} 1 & 1 & 3 & 3 \\ 0 & -5 & 97 & 92 \\ 0 & -1 & 96 & 95 \end{array}\right) \xrightarrow{F_3-0{,}2F_2} \left(\begin{array}{ccc|c} 1 & 1 & 3 & 3 \\ 0 & -5 & 97 & 92 \\ 0 & 0 & 76{,}6 & 76{,}6 \end{array}\right).$$

La sustitución hacia atrás nos da $x_3 = 1{,}0$, $x_2 = 1{,}0$, $x_1 = -1{,}0$, que es la solución.

Por completar, comparemos este resultado con la técnica de escalado previo y pivoteo parcial (tres dígitos significativos). En primer lugar, el escalado:

$$\left[\begin{array}{cccc} 3 & 2 & 105 & 104 \\ 2 & -3 & 103 & 98 \\ 1 & 1 & 3 & 3 \end{array}\right] \xrightarrow{\frac{1}{105}F_1, \frac{1}{103}F_2, \frac{1}{3}F_3} \left(\begin{array}{cccc} 0{,}0286 & 0{,}0190 & 1{,}0 & 0{,}990 \\ 0{,}0194 & -0{,}0291 & 1{,}0 & 0{,}951 \\ 0{,}333 & 0{,}333 & 1{,}0 & 1{,}0 \end{array}\right).$$

En la primera columna, el elemento de mayor módulo se encuentra en la tercera fila, y efectuamos el intercambio. Los multiplicadores para la eliminación son $m_1 = \text{fl}(0{,}0194/0{,}333) = 0{,}0583$, $m_2 = \text{fl}(0{,}0286/0{,}333) = 0{,}0859$.

$$\begin{pmatrix} 0{,}333 & 0{,}333 & 1{,}0 & 1{,}0 \\ 0{,}0194 & -0{,}0291 & 1{,}0 & 0{,}951 \\ 0{,}0286 & 0{,}0190 & 1{,}0 & 0{,}990 \end{pmatrix} \xrightarrow{F_2 - m_1 F_1, F_3 - m_2 F_1} \begin{pmatrix} 0{,}333 & 0{,}333 & 1{,}0 & 1{,}0 \\ 0{,}0 & -0{,}0485 & 0{,}942 & 0{,}893 \\ 0{,}0 & -0{,}00960 & 0{,}914 & 0{,}904 \end{pmatrix}.$$

El nuevo multiplicador es $m_3 = \text{fl}(-0{,}00960/-0{,}0485) = 0{,}198$, y la reducción queda

$$\begin{pmatrix} 0{,}333 & 0{,}333 & 1{,}0 & 1{,}0 \\ 0{,}0 & -0{,}0485 & 0{,}942 & 0{,}893 \\ 0{,}0 & -0{,}00960 & 0{,}914 & 0{,}904 \end{pmatrix} \xrightarrow{F_3 - m_3 F_2} \begin{pmatrix} 0{,}333 & 0{,}333 & 1{,}0 & 1{,}0 \\ 0{,}0 & -0{,}0485 & 0{,}942 & 0{,}893 \\ 0{,}0 & 0{,}0 & 0{,}727 & 0{,}727 \end{pmatrix}$$

Entonces

$$\begin{aligned} x_3 &= \text{fl}(0{,}727/0{,}727) = 1{,}0, \\ x_2 &= \frac{1}{-0{,}0485}(0{,}893 - 0{,}942 * x_3) \overset{\text{fl}}{=} 1{,}01, \\ x_1 &= \frac{1}{0{,}333}(1 - x_3 - 0{,}333 * x_2) = 1{,}01. \end{aligned}$$

Es un resultado algo peor, y además consume más operaciones.

1.6. * Pivoteo completo

Aunque no es ampliamente usada, existe una extensión del pivoteo parcial conocido como *pivoteo completo o total*, que, en algunos casos, puede ser más efectivo que el pivoteo parcial para controlar los efectos del error de redondeo.

Pivoteo completo

Si $[A|\boldsymbol{b}]$ es la matriz ampliada en el k-ésimo paso de la eliminación gaussiana, entonces hay que buscar el elemento de mayor módulo en las posiciones por debajo o a la derecha de la posición pivotal. Si es necesario, se realizan los apropiados cambios de filas y columnas para llevar dicho coeficiente a la posición pivotal.

Como una situación típica, consideremos el tercer paso en la siguiente matriz

$$\left(\begin{array}{ccccc|c} * & * & * & * & * & * \\ 0 & * & * & * & * & * \\ 0 & 0 & \boxed{S} & S & S & * \\ 0 & 0 & S & S & S & * \\ 0 & 0 & S & S & S & * \end{array}\right).$$

Buscamos el coeficiente de mayor módulo entre las posiciones marcadas con "S". Si es necesario, intercambiamos filas y columnas para llevar este elemento máximo a la posición pivotal marcada. Esto tiene el efecto de renombrar las incógnitas asociadas.

El pivoteo completo es tan efectivo como el parcial. Es incluso posible construir ejemplos donde el pivoteo completo es superior al parcial. Sin embargo, en la práctica se encuentran pocos ejemplos.

Ejemplo 1.6.1. Con aritmética de 3 dígitos y pivoteo completo, resolvemos el sistema

$$\begin{array}{rcrcr} x & - & y & = & -2, \\ -9x & + & 10y & = & 12. \end{array}$$

Como 10 es el coeficiente de mayor módulo, intercambiamos la primera y segunda filas, y también la primera y segunda columnas.

$$\left(\begin{array}{cc|c} 1 & -1 & -2 \\ -9 & 10 & 12 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} -9 & 10 & 12 \\ 1 & -1 & -2 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{cc|c} 10 & -9 & 12 \\ -1 & 1 & -2 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 10 & -9 & 12 \\ 0 & ,1 & -,8 \end{array}\right)$$

El efecto de intercambiar las columnas es renombrar las incógnitas a $\hat{x}$ y $\hat{y}$, donde $\hat{x} = y$ y $\hat{y} = x$. La sustitución hacia atrás nos da $\hat{y} = -8$ y $\hat{x} = -6$, por lo que

$$x = \hat{y} = -8 \text{ y } y = \hat{x} = -6.$$

En este caso, la solución con 3 dígitos y la solución exacta coinciden. Si solamente se hubiera usado pivoteo parcial, la solución con 3 dígitos no habría sido tan precisa. Sin embargo, la combinación de escalado y pivoteo parcial da el mismo resultado que pivoteo completo.

Si el coste del pivoteo completo fuera similar al coste del pivoteo parcial, siempre usaríamos pivoteo completo. Sin embargo, el pivoteo completo necesita, en el paso k-ésimo, calcular el máximo de un conjunto de elementos de k^2 elementos. Cada comparación es una resta, por lo que la suma de todas estas operaciones añade del orden de $\frac{1}{3}n^3$ flops. El pivoteo parcial añade del orden

de $\frac{1}{2}n^2$ comparaciones, lo que no afecta en gran medida al coste de la eliminación gaussiana. Si juntamos estos hechos con lo raro que resulta en la práctica encontrar sistemas donde pivoteo parcial y escalado no es adecuado y pivoteo completo sí, es fácil entender por qué el pivoteo completo es raramente usado en la práctica. La eliminación gaussiana con escalado y pivoteo parcial es el método preferido para resolver sistemas densos de tamaño moderado.

1.7. Sistemas mal condicionados

La eliminación gaussiana con pivoteo parcial sobre un sistema escalado adecuadamente es quizás el algoritmo más fundamental en el uso práctico del álgebra lineal. Sin embargo, no es un algoritmo universal ni puede usarse a ciegas. En esta sección apuntamos a que en la resolución de un sistema lineal debemos usar algo de buen juicio porque hay sistemas que son tan sensibles a pequeñas perturbaciones que *no hay* ninguna técnica numérica que pueda usarse con confianza.

Ejemplo 1.7.1. Consideremos el siguiente sistema

$$\begin{array}{rcrcr} ,835x & + & ,667y & = & ,168 \\ ,333x & + & ,266y & = & ,067 \end{array}',$$

que tiene como solución exacta $x = 1, y = -1$. Si $b_2 = ,067$ se modifica ligeramente a $\hat{b}_2 = ,066$, entonces la solución exacta cambia a $\hat{x} = -666, \hat{y} = 834$.

Este es un ejemplo de un sistema cuya solución es muy sensible a pequeñas variaciones. Esta sensibilidad es intrínseca al sistema, y no depende del método numérico para resolverlo. Por tanto, no podemos esperar algún “truco numérico"que elimine esta sensibilidad. Si la solución exacta es sensible a pequeñas perturbaciones, entonces cualquier solución calculada con aritmética en coma flotante sufre del mismo problema, y este comportamiento es independiente del algoritmo usado.

Sistemas lineales mal condicionados

Un sistema de ecuaciones se dice ***mal condicionado*** cuando pequeñas perturbaciones en los coeficientes del sistema producen grandes cambios en la solución exacta. En otro caso, decimos que el sistema está ***bien condicionado***.

Es fácil visualizar lo que causa que un sistema 2×2 sea mal condicionado. Desde el punto de vista geométrico, dos ecuaciones con dos incógnitas representan dos líneas rectas, y el punto de intersección es la solución del sistema. Un sistema mal condicionado son dos líneas rectas casi paralelas.

Dado que los errores de redondeo se pueden ver como perturbaciones de los coeficientes del sistema original, incluso el empleo de una buena técnica numérica (por ejemplo, aritmética exacta) sobre un sistema mal condicionado, lleva el riesgo de producir resultados sin sentido.

Un científico, cuando trata un sistema mal condicionado, se enfrenta con un problema más básico, y más preocupante, que el de resolver el sistema. Incluso si pudiera realizar un milagro y calcular la solución exacta, el científico podría obtener una solución sin sentido que le llevase a conclusiones totalmente falsas. El problema procede de que los coeficientes se obtienen con frecuencia de manera empírica y son conocidos dentro de unas tolerancias. Para un sistema mal condicionado, una pequeña incertidumbre en alguno de los coeficientes puede significar una incertidumbre enorme en la solución. Esta gran incertidumbre puede llevar a considerar la solución exacta inútil.

Ejemplo 1.7.2. Supongamos que en el sistema

$$\begin{array}{rcrcl} ,835x & + & ,667y & = & b_1, \\ ,333x & + & ,266y & = & b_2, \end{array}$$

los números b_1 y b_2 se obtienen como resultados de un experimento y se leen de la pantalla de un instrumento de medición. Supongamos que el sensor está calibrado con una tolerancia de $\pm$,001, y que los valores leídos de b_1 y b_2 son ,168 y ,067, respectivamente. Esto nos da el sistema mal condicionado del ejemplo anterior, y vimos que la solución exacta era

$$(x, y) = (1, -1). \tag{1.7.1}$$

Sin embargo, debido a la incertidumbre de la lectura, tenemos que

$$,167 \leq b_1 \leq ,169 \text{ y } ,066 \leq b_2 \leq ,068. \tag{1.7.2}$$

Por ejemplo, esto significa que la solución asociada con la lectura $(b_1, b_2) = (,168, ,067)$ es tan válida como la solución asociada con la lectura $(b_1, b_2) = (,167, ,068)$, o la lectura $(b_1, b_2) = (,169, ,066)$, o cualquier otra lectura que caiga en el rango dado por (1.7.2). Para la lectura $(b_1, b_2) = (,167, ,068)$, la solución exacta es

$$(x, y) = (934, -1169), \tag{1.7.3}$$

mientras que para $(b_1, b_2) = (,169, ,066)$, la solución exacta es

$$(x, y) = (-932, 1167). \tag{1.7.4}$$

¿Se atrevería a ser el primero en volar en un avión o atravesar un puente cuyo diseño incorporara una solución a este problema?

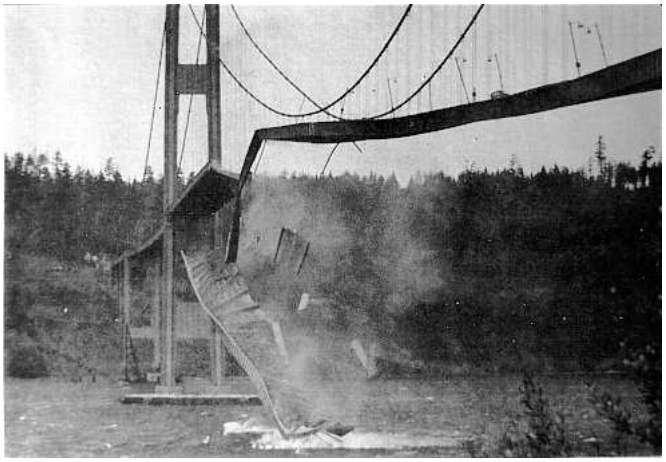

Figura 1.4: Puente de Tacoma

Como ninguna de las soluciones 1.7.1, 1.7.3 o 1.7.4 puede considerarse mejor que las otras, es imaginable que diseños totalmente diferentes se realicen dependiendo de cómo el técnico lea el último dígito significativo de la pantalla. Debido a la naturaleza mal condicionada de un sistema lineal, el buen diseño de un avión o un puente puede depender de la suerte ciega más que de principios científicos.

Antes que extraer información de soluciones con mucha precisión de sistemas mal condicionados, es mejor invertir tiempo y recursos en diseñar de otra forma los experimentos asociados o los métodos de captura de datos para evitar sistemas mal condicionados.

Hay otro aspecto desconcertante en los sistemas mal condicionados. Se refiere a lo que los estudiantes llaman “verificación de la respuesta" mediante la sustitución de una solución calculada en el lado izquierdo de la ecuación y ver cuán próxima es al lado derecho. Más formalmente, si

$$x_c = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{pmatrix}$$

es una solución calculada del sistema

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \dots & + & a_{bm}x_n & = & b_n, \end{array}$$

entonces los números

$$r_i = a_{i1}\xi_1 + a_{i2}\xi_2 + \cdots + a_{in}\xi_n - b_i \text{ para } i = 1, 2, \ldots, n$$

se denominan ***residuos***. Supongamos que calculamos una solución x_c, y que los residuos son relativamente pequeños. ¿Nos garantiza esto que x_c es cercana a la solución exacta? Sorprendentemente, la respuesta es un sonoro **NO** cuando el sistema está mal condicionado.

Ejemplo 1.7.3. Para el sistema mal condicionado del ejemplo 1.7.1, supongamos que de alguna forma hemos calculado que la solución es

$$\xi_1 = -666, \xi_2 = 834.$$

Si intentamos “verificar" el error de esta solución mediante la sustitución en el sistema, encontramos, con aritmética exacta, que los residuos son

$$\begin{aligned} r_1 &= ,835\xi_1 + ,667\xi_2 - ,168 = 0, \\ r_2 &= ,333\xi_1 + ,266\xi_2 - ,067 = -,001. \end{aligned}$$

Así, la solución calculada $(-666, 834)$ satisface exactamente la primera ecuación y está muy cerca de verificar la segunda. En principio, esto parece sugerir que la solución calculada debería ser muy próxima a la solución exacta. Una persona ingenua podría ser inducida a creer que la solución calculada está en un rango de $\pm,001$ de la solución exacta. Evidentemente, no está cerca de la solución exacta, que es

$$x = 1, y = -1.$$

Siempre es un choque mental ver esto por primera vez porque va en contra de la intuición del aprendiz. Por desgracia, muchos estudiantes salen de los cursos creyendo que siempre se puede verificar la exactitud de sus cálculos mediante la simple sustitución en las ecuaciones originales; es bueno saber que tú, querido lector, no estás entre ellos.

Lo anterior nos lleva a la pregunta de cómo podemos comprobar si una solución calculada es más o menos correcta. Afortunadamente, si el sistema está bien condicionado, los residuos proporcionan un buen método para medir la precisión. Una mirada con mayor detalle la veremos más adelante. Pero esto significa que debemos ser capaces de responder a más preguntas. Por ejemplo, ¿cómo podemos decir a priori si un sistema está mal condicionado? ¿Cómo podemos medir el grado de mal condicionamiento de un sistema?

Es posible realizar experimentos con los coeficientes, y estudiar cómo afectan a la solución, pero esto es caro y nada satisfactorio. Pero antes de que podamos decir algo, necesitamos herramientas más sofisticadas.

Capítulo 2

Sistemas rectangulares y formas escalonadas

2.1. Forma escalonada por filas y rango

Ya estamos preparados para analizar sistemas rectangulares con m ecuaciones y n incógnitas

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

donde m puede ser diferente de n; decimos que el sistema es ***rectangular***. El caso $m = n$ también queda comprendido en lo que digamos.

La primera tarea es extender la eliminación gaussiana de sistemas cuadrados a sistemas rectangulares. Recordemos que para un sistema cuadrado con solución única, las posiciones pivote siempre se localizan a lo largo de la ***diagonal principal*** de la matriz de coeficientes A, por lo que la eliminación gaussiana resulta en una reducción de A a una matriz ***triangular***, similar, para $n = 4$, a

$$T = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Recordemos que un pivote debe ser siempre un valor no nulo. Para sistemas cuadrados con una única solución, probaremos que siempre podremos obtener un valor no nulo en cada posición pivotal a lo largo de la diagonal principal

(hablamos para aritmética exacta). Sin embargo, en el caso de un sistema rectangular general, no siempre es posible tener las posiciones pivote en la diagonal principal de la matriz de coeficientes. Esto significa que el resultado final de la eliminación gaussiana *no* será una forma triangular. Por ejemplo, consideremos el siguiente sistema:

$$\begin{array}{rcrcrcrcrcl} x_1 & + & 2x_2 & + & x_3 & + & 3x_4 & + & 3x_5 & = & 5, \\ 2x_1 & + & 4x_2 & & & + & 4x_4 & + & 4x_5 & = & 6, \\ x_1 & + & 2x_2 & + & 3x_3 & + & 5x_4 & + & 5x_5 & = & 9, \\ 2x_1 & + & 4x_2 & & & + & 4x_4 & + & 7x_5 & = & 9. \end{array}$$

Consideremos la matriz ampliada $\left(\begin{array}{c|c} A & \boldsymbol{b} \end{array}\right)$, donde

$$A = \begin{pmatrix} 1 & 2 & 1 & 3 & 3 \\ 2 & 4 & 0 & 4 & 4 \\ 1 & 2 & 3 & 5 & 5 \\ 2 & 4 & 0 & 4 & 7 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 5 \\ 6 \\ 9 \\ 9 \end{pmatrix}. \tag{2.1.1}$$

Aplicamos eliminación gaussiana a $\left(\begin{array}{c|c} A & \boldsymbol{b} \end{array}\right)$ y obtenemos el siguiente resultado:

$$\left(\begin{array}{ccccc|c} \boxed{1} & 2 & 1 & 3 & 3 & 5 \\ 2 & 4 & 0 & 4 & 4 & 6 \\ 1 & 2 & 3 & 5 & 5 & 9 \\ 2 & 4 & 0 & 4 & 7 & 9 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 3 & 5 \\ \boxed{0} & 0 & -2 & -2 & -2 & -4 \\ 0 & 0 & 2 & 2 & 2 & 4 \\ 0 & 0 & -2 & -2 & 1 & -1 \end{array}\right).$$

En el proceso de eliminación básico, nos movemos abajo y a la derecha, a la siguiente posición pivotal. Si encontramos un cero en esta posición, se efectúa un intercambio con una fila inferior para llevar un número no nulo a la posición pivotal. Sin embargo, en este ejemplo, es imposible llevar un elemento no nulo a la posición $(2,2)$ mediante el intercambio de la segunda fila con una fila inferior.

Para manejar esta situación, debemos modificar el procedimiento.

Eliminación gaussiana modificada

Supongamos que U es la matriz ampliada asociada a un sistema tras haber completado $i-1$ pasos de eliminación. Para ejecutar el i-ésimo paso, procedemos como sigue:

- De izquierda a derecha en U, localizamos la primera columna que contiene un valor no nulo en o por debajo de la i-ésima posición. Digamos que es U_{*j}.
- La posición pivotal para el i-ésimo paso es la posición (i, j).
- Si es necesario, intercambia la i-ésima fila con una fila inferior para llevar un número no nulo a la posición (i, j), y entonces anula todas las entradas por debajo de este pivote.
- Si la fila U_{i*} así como todas las filas de U por debajo de U_{i*} consisten en filas nulas, entonces el proceso de eliminación está completo.

Ilustremos lo anterior aplicando la versión modificada de la eliminación gaussiana a la matriz dada en 2.1.1

Ejemplo 2.1.1. Aplicamos la eliminación gaussiana modificada a la matriz

$$\left(\begin{array}{c|c} A & \boldsymbol{b} \end{array}\right) = \left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 3 & 5 \\ 2 & 4 & 0 & 4 & 4 & 6 \\ 1 & 2 & 3 & 5 & 5 & 9 \\ 2 & 4 & 0 & 4 & 7 & 9 \end{array}\right),$$

y marcamos las posiciones pivote.

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

$$\to \left(\begin{array}{ccccc|c} \boxed{1} & 2 & 1 & 3 & 3 & 5 \\ 0 & 0 & \boxed{-2} & -2 & -2 & 4 \\ 0 & 0 & 0 & 0 & \boxed{0} & 0 \\ 0 & 0 & 0 & 0 & 3 & 3 \end{array}\right) \to \left(\begin{array}{ccccc|c} \boxed{1} & 2 & 1 & 3 & 3 & 5 \\ 0 & 0 & \boxed{-2} & -2 & -2 & -4 \\ 0 & 0 & 0 & 0 & \boxed{3} & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Observemos que el resultado final de aplicar la eliminación gaussiana en el ejemplo anterior no es una forma triangular, sino un tipo escalonado de forma triangular. De aquí en adelante, una matriz que muestre esta estructura la llamaremos ***forma escalonada por filas***.

Forma escalonada por filas

Una matriz E de orden $m \times n$ con filas E_{i*} y columnas E_{*j} se dice que está en ***forma escalonada por filas*** si se verifica lo siguiente.

- Si E_{i*} es una fila de ceros, entonces todas las filas por debajo de E_{i*} son también nulas.
- Si la primera entrada no nula de E_{i*} está en la j-ésima posición, entonces todas las entradas por debajo de la i-ésima posición en las columnas $E_{*1}, E_{*2}, \ldots, E_{*j}$ son nulas.

Una estructura típica de una forma escalonada por filas, con los pivotes marcados, es

$$\begin{pmatrix} \boxed{*} & * & * & * & * & * & * & * \\ 0 & 0 & \boxed{*} & * & * & * & * & * \\ 0 & 0 & 0 & \boxed{*} & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{*} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Los pivotes son las primeras entradas no nulas en cada fila. Podemos tener también columnas de ceros a la izquierda de la matriz.

Como hay flexibilidad para elegir las operaciones por filas que reducen una matriz A a una forma escalonada E, las entradas de E no están unívocamente determinadas por A. No obstante, se puede probar que la forma de E es única en el sentido de que las *posiciones de los pivotes* en E están completamente determinadas por las entradas de A. Esto lo veremos tras la siguiente sección, donde damos un paso más.

2.2. Forma escalonada reducida por filas

En cada paso del método de Gauss-Jordan, forzábamos a que el pivote fuera 1, y entonces todas las entradas por encima y por debajo del pivot se anulaban. Si A es la matriz de coeficientes de un sistema cuadrado con solución única,

entonces el resultado final de aplicar el método de Gauss-Jordan a A es una matriz con 1 en la diagonal y 0 en el resto. Esto es,

$$A \xrightarrow{\text{Gauss-Jordan}} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Pero si la técnica de Gauss-Jordan se aplica a matrices rectangulares $m \times n$, entonces el resultado final no es necesariamente como el descrito antes. El siguiente ejemplo ilustra qué ocurre en el caso rectangular.

Ejemplo 2.2.1. Aplicamos la eliminación de Gauss-Jordan a la matriz

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 3 \\ 2 & 4 & 0 & 4 & 4 \\ 1 & 2 & 3 & 5 & 5 \\ 2 & 4 & 0 & 4 & 7 \end{pmatrix}$$

y marcamos las posiciones pivote.

$$\begin{pmatrix} \boxed{1} & 2 & 1 & 3 & 3 \\ 2 & 4 & 0 & 4 & 4 \\ 1 & 2 & 3 & 5 & 5 \\ 2 & 4 & 0 & 4 & 7 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 2 & 1 & 3 & 3 \\ 0 & 0 & \boxed{-2} & -2 & -2 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & -2 & -2 & 1 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} \boxed{1} & 2 & 1 & 3 & 3 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 \\ 0 & 0 & -2 & -2 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 2 & 0 & 2 & 2 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{0} \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} \boxed{1} & 2 & 0 & 2 & 2 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} \boxed{1} & 2 & 0 & 2 & 2 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} \boxed{1} & 2 & 0 & 2 & 0 \\ 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Si obtenemos una forma escalonada de la matriz A, llegamos a una expresión

$$\left(\begin{array}{ccccc|c} \boxed{1} & 2 & 1 & 3 & 3 & 5 \\ 0 & 0 & \boxed{-2} & -2 & -2 & 4 \\ 0 & 0 & 0 & 0 & \boxed{0} & 0 \\ 0 & 0 & 0 & 0 & 3 & 3 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} \boxed{1} & 2 & 1 & 3 & 3 & 5 \\ 0 & 0 & \boxed{-2} & -2 & -2 & -4 \\ 0 & 0 & 0 & 0 & \boxed{3} & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Vemos que la forma de la matriz final es la misma en ambos casos, que tiene que ver con la unicidad que hemos comentado anteriormente. La única diferencia es el valor numérico de algunas entradas. Por la naturaleza de la eliminación de Gauss-Jordan, cada pivote es 1 y todas las entradas por encima y por debajo son nulas. Por tanto, la forma escalonada por filas que produce el método de Gauss-Jordan contiene un número reducido de entradas no nulas, por lo que parece natural llamarla ***forma escalonada reducida por filas***.

Forma escalonada reducida por filas

Una matriz $E_{m\times n}$ está en ***forma escalonada reducida por filas*** si se verifican las siguientes condiciones:

- E está en forma escalonada por filas.
- La primera entrada no nula de cada fila (el pivote) es 1.
- Todas las entradas por encima del pivote son cero.

Una estructura típica de una matriz en forma escalonada reducida por filas es

$$\begin{pmatrix} \boxed{1} & * & 0 & 0 & * & * & 0 & * \\ 0 & 0 & \boxed{1} & 0 & * & * & 0 & * \\ 0 & 0 & 0 & \boxed{1} & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Como comentamos antes, si una matriz A se transforma en una forma escalonada por filas mediante operaciones por fila, entonces la forma está unívocamente determinada por A, pero las entradas individuales de la forma no son únicas. Sin embargo, si A se transforma mediante operaciones por fila a una forma *reducida* por filas E_A, se puede probar que tanto la forma como las entradas individuales en E_A están unívocamente determinadas por A. En otras palabras, la forma escalonada reducida por filas E_A generada por A es independiente del camino de eliminación que usemos. Es claro que producir una forma escalonada no reducida es más eficiente desde el punto de vista de computación, pero la unicidad de E_A la hace útil para ciertas cuestiones teóricas.

Unicidad de E_A

La forma escalonada reducida por filas de una matriz $A_{m\times n}$ es única.

PRUEBA: Comenzaremos probando un resultado que luego generalizaremos. Si la matriz B se obtiene de A a partir de una operación elemental, y una columna de A satisface una relación del tipo

$$A_{*k} = \sum_{j=1}^{n} \alpha_j A_{*j},$$

entonces la columna correspondiente de B satisface una relación análoga, esto es,

$$B_{*k} = \sum_{j=1}^{n} \alpha_j B_{*j}. \tag{2.2.1}$$

Es una comprobación inmediata para cada una de las tres operaciones elementales. Supongamos entonces que hemos llegado, mediante las operaciones por filas, a dos matrices U y V en forma escalonada reducida por filas. La entrada no nula situada más a la izquierda en una fila de U es un 1, que ocupa una posición pivote, y su columna la llamamos columna pivote. Las columnas pivote de las matrices U y V son precisamente las columnas no nulas que no dependen linealmente de las columnas a su izquierda. Como U y V se pueden transformar una en otra mediante transformaciones elementales, sus columnas tiene las mismas relaciones de dependencia lineal. Por tanto, las columnas pivote de U y V aparecen en la misma posición. Si hay r de tales columnas, como U y V están en forma escalonada reducida por filas, sus columnas pivote son las r primeras columnas de la matriz identidad de orden $m\times m$. Por tanto, las columnas pivote correspondientes de U y V son iguales.

Consideremos ahora cualquier columna no pivote de U, por ejemplo, la j-ésima. Esta columna es cero o es una combinación lineal de las columnas pivote de su izquierda. En cualquier caso, la correspondiente columna j-ésima de V verifica la misma relación, por lo que es igual a la de U. En definitiva, U y V son iguales. □

Como las posiciones pivote son únicas, se sigue que el número de pivotes, que es el mismo que el número de filas no nulas de E, también está unívocamente determinado por A. Este número se denomina ***rango*** de A, y es uno de los conceptos fundamentales del curso.

Rango de una matriz

Supongamos que una matriz $A_{m\times n}$ se reduce mediante operaciones por filas a una forma escalonada E. El ***rango*** de A es el número

$$\begin{aligned}\text{rango}(A) &= \text{número de pivotes}\\ &= \text{número de filas no nulas de } E\\ &= \text{número de columnas básicas de } A,\end{aligned}$$

donde las columnas básicas de A son aquellas columnas de A que contienen las posiciones pivote.

Ejemplo 2.2.2. Determinemos el rango y columnas básicas de la matriz

$$A=\begin{pmatrix}1&2&1&1\\2&4&2&2\\3&6&3&4\end{pmatrix}.$$

Reducimos A a forma escalonada por filas.

$$\begin{pmatrix}\boxed{1}&2&1&1\\2&4&2&2\\3&6&3&4\end{pmatrix}\to\begin{pmatrix}\boxed{1}&2&1&1\\0&0&0&\boxed{0}\\0&0&0&1\end{pmatrix}\to\begin{pmatrix}\boxed{1}&2&1&1\\0&0&0&\boxed{1}\\0&0&0&0\end{pmatrix}=E$$

Por tanto, rango$(A)=2$. Las posiciones pivote están en la primera y cuarta columna, por lo que las columnas básicas de A son A_{*1} y A_{*4}. Esto es,

$$\text{Columnas básicas } =\left\{\begin{pmatrix}1\\2\\3\end{pmatrix},\begin{pmatrix}1\\2\\4\end{pmatrix}\right\}.$$

Es importante resaltar que las columnas básicas se extraen de A y no de la forma escalonada E.

Notación

Para una matriz A, el símbolo E_A denotará la única forma escalonada reducida por filas derivada de A mediante operaciones por fila. También escribiremos

$$A\overset{\text{rref}}{\longrightarrow}E_A.$$

Ejemplo 2.2.3. Determinemos E_A, calculemos rango(A) e identifiquemos las columnas básicas de

$$A = \begin{pmatrix} 1 & 2 & 2 & 3 & 1 \\ 2 & 4 & 4 & 6 & 2 \\ 3 & 6 & 6 & 9 & 6 \\ 1 & 2 & 4 & 5 & 3 \end{pmatrix}.$$

$$\begin{pmatrix} \boxed{1} & 2 & 2 & 3 & 1 \\ 2 & 4 & 4 & 6 & 2 \\ 3 & 6 & 6 & 9 & 6 \\ 1 & 2 & 4 & 5 & 3 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 2 & 2 & 3 & 1 \\ 0 & 0 & \boxed{0} & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 2 & 2 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 2 & 2 & 3 & 1 \\ 0 & 0 & \boxed{2} & 2 & 2 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} \boxed{1} & 2 & 2 & 3 & 1 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 2 & 0 & 1 & -1 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} \boxed{1} & 2 & 0 & 1 & -1 \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 2 & 0 & 1 & 0 \\ 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Por tanto, rango$(A) = 3$, y $\{A_{*1}, A_{*3}, A_{*5}\}$ son las tres columnas básicas.

El ejemplo anterior ilustra otra importante característica de E_A, y explica por qué las columnas básicas reciben ese nombre. Cada columna no básica es expresable como combinación lineal de las columnas básicas. En el ejemplo,

$$A_{*2} = 2A_{*1}, A_{*4} = A_{*1} + A_{*3}. \tag{2.2.2}$$

Observemos que las mismas relaciones se tienen en E_A, esto es,

$$E_{*2} = 2E_{*1}, E_{*4} = E_{*1} + E_{*3}. \tag{2.2.3}$$

La razón la encontramos en la prueba de 2.2.1. La matriz E_A se obtiene mediante transformaciones elementales de A, por lo que las relaciones entre las columnas de A son las mismas que las de E_A (y al revés). Las relaciones entre las columnas básicas y no básicas en una matriz general A no se ven a simple vista, pero las relaciones entre las columnas de E_A son completamente transparentes. Por ejemplo, los coeficientes usados en las relaciones (2.2.2) y (2.2.3) aparecen explícitamente en las dos columnas no básicas de E_A. Son precisamente las entradas no nulas en estas columnas no básicas. Esto es importante, porque usaremos E_A como un mapa o clave para revelar las relaciones ocultas entre las columnas de A.

Finalmente, observemos del ejemplo que únicamente las columnas básicas *a la izquierda* de una columna no básica dada se necesitan para expresar la columna no básica como combinación lineal de las columnas básicas. Así, la expresión de A_{*2} requiere únicamente de A_{*1}, y no de A_{*3} o A_{*5}, mientras que la expresión de A_{*4} precisa únicamente de A_{*1} y A_{*3}. Esto es lo que hemos probado en el teorema de unicidad de la forma escalonada reducida por filas.

Relaciones de las columnas en A y E_A

- Cada columna no básica E_{*k} de E_A es una combinación lineal de las columnas básicas de E_A a la izquierda de E_{*k}. Esto es,

$$E_{*k} = \mu_1 E_{*b_1} + \mu_2 E_{*b_2} + \cdots + \mu_j E_{*b_j},$$

  donde las E_{*b_i} son las columnas básicas a la izquierda de E_{*k}, y los coeficientes μ_j son las primeras j entradas de E_{*k}.

- Las relaciones que existen entre las columnas de A son exactamente las mismas relaciones que existen entre las columnas de E_A. En particular, si A_{*k} es una columna no básica de A, entonces

$$A_{*k} = \mu_1 A_{*b_1} + \mu_2 A_{*b_2} + \cdots + \mu_j A_{*b_j},$$

  donde las A_{*b_i} son las columnas básicas de A situadas a la izquierda de A_{*k} y los coeficientes μ_j son los descritos antes.

Lo que tenemos es una expresión de la forma

$$\begin{aligned} E_{*k} &= \mu_1 E_{*b_1} + \mu_2 E_{*b_2} + \cdots + \mu_j E_{*b_j} \\ &= \mu_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \mu_j \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_j \\ \vdots \\ 0 \end{pmatrix}. \end{aligned}$$

Ejemplo 2.2.4. Escribamos las columnas no básicas de la matriz

$$A = \begin{pmatrix} 2 & -4 & -8 & 6 & 3 \\ 0 & 1 & 3 & 2 & 3 \\ 3 & -2 & 0 & 0 & 8 \end{pmatrix}$$

como combinación lineal de las básicas. Para ello, calculamos la forma escalonada reducida por filas E_A.

$$\begin{pmatrix} \boxed{2} & -4 & -8 & 6 & 3 \\ 0 & 1 & 3 & 2 & 3 \\ 3 & -2 & 0 & 0 & 8 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & -2 & -4 & 3 & \frac{3}{2} \\ 0 & 1 & 3 & 2 & 3 \\ 3 & -2 & 0 & 0 & 8 \end{pmatrix} \to \begin{pmatrix} \boxed{1} & -2 & -4 & 3 & \frac{3}{2} \\ 0 & \boxed{1} & 3 & 2 & 3 \\ 0 & 4 & 12 & -9 & \frac{7}{2} \end{pmatrix} \to$$
$$\begin{pmatrix} \boxed{1} & 0 & 2 & 7 & \frac{15}{2} \\ 0 & \boxed{1} & 3 & 2 & 3 \\ 0 & 0 & 0 & -17 & -\frac{17}{2} \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 0 & 2 & 7 & \frac{15}{2} \\ 0 & \boxed{1} & 3 & 2 & 3 \\ 0 & 0 & 0 & \boxed{1} & \frac{1}{2} \end{pmatrix} \to \begin{pmatrix} \boxed{1} & 0 & 2 & 0 & 4 \\ 0 & \boxed{1} & 3 & 0 & 2 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \end{pmatrix}.$$

Las columnas tercera y quinta son no básicas. Revisando las columnas de E_A, tenemos que

$$E_{*3} = 2E_{*1} + 3E_{*2} \text{ y } E_{*5} = 4E_{*1} + 2E_{*2} + \frac{1}{2}E_{*4}.$$

Las relaciones que existen entre las columnas de A son exactamente las mismas que las de E_A, esto es,

$$A_{*3} = 2A_{*1} + 3A_{*2} \text{ y } A_{*5} = 4A_{*1} + 2A_{*2} + \frac{1}{2}A_{*4}.$$

En resumen, la utilidad de E_A reside en su habilidad para revelar las dependencias entre los datos almacenados en la matriz A. Las columnas no básicas de A representan información redundante en el sentido de que esta información se puede expresar en términos de los datos contenidos en las columnas básicas.

Aunque la compresión de datos no es la razón primaria para introducir a E_A, la aplicación a estos problemas es clara. Para una gran matriz de datos, es más eficiente almacenar únicamente las columnas básicas de A con los coeficientes μ_j obtenidos de las columnas no básicas de E_A. Entonces los datos redundantes contenidos en las columnas no básicas de A siempre se pueden reconstruir cuando los necesitemos. Algo parecido ocurrirá cuando tratemos el problema de la colinealidad de datos.

2.3. Compatibilidad de los sistemas lineales

Un sistema de m ecuaciones y n incógnitas se dice ***compatible*** si posee el menos una solución. Si no tiene soluciones, decimos que el sistema es ***incompatible***. El propósito de esta sección es determinar las condiciones bajo las que un sistema es compatible.

Establecer dichas condiciones para un sistema de dos o tres incógnitas es fácil. Una ecuación lineal con dos incógnitas representa una recta en el plano,

y una ecuación lineal con tres incógnitas es un plano en el espacio de tres dimensiones. Por tanto, un sistema lineal de m ecuaciones con dos incógnitas es compatible si y solamente si las m rectas definidas por las m ecuaciones tienen un punto común de intersección. Lo mismo ocurre para m planos en el espacio. Sin embargo, para m grande, estas condiciones geométricas pueden ser difíciles de verificar visualmente, y cuando $n > 3$ no es posible esta representación con los ojos.

Mejor que depender de la geometría para establecer la compatibilidad, usaremos la eliminación gaussiana. Si la matriz ampliada asociada $[A|\boldsymbol{b}]$ se reduce mediante operaciones por filas a una forma escalonada por filas $[E|\boldsymbol{c}]$, entonces la compatibilidad o no del sistema es evidente. Supongamos que en un momento del proceso de reducción de $[A|\boldsymbol{b}]$ a $[E|\boldsymbol{c}]$ llegamos a una situación en la que la única entrada no nula de una fila aparece en el lado derecho, como mostramos a continuación:

$$\text{Fila } i \rightarrow \left(\begin{array}{cccccc|c} * & * & * & * & * & * & * \\ 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right) \leftarrow \alpha \neq 0.$$

Si esto ocurre en la i-ésima fila, entonces la i-ésima ecuación del sistema asociado es

$$0 \cdot x_1 + 0 \cdot x_2 + \cdots + 0 \cdot x_n = \alpha.$$

Para $\alpha \neq 0$, esta ecuación no tiene solución, y el sistema original es incompatible (recordemos que las operaciones por filas no alteran el conjunto de soluciones). El recíproco también se verifica. Esto es, si el sistema es incompatible, entonces en algún momento del proceso de eliminación llegamos a una fila de la forma

$$\left(\begin{array}{ccccc|c} 0 & 0 & \ldots & 0 & | & \alpha \end{array}\right), \alpha \neq 0. \tag{2.3.1}$$

En otro caso, la sustitución hacia atrás se podría realizar y obtener una solución. No hay incompatibilidad si se llega a una fila de la forma

$$\left(\begin{array}{ccccc|c} 0 & 0 & \ldots & 0 & | & 0 \end{array}\right).$$

Esta ecuación dice simplemente $0 = 0$, y aunque no ayuda a determinar el valor de ninguna incógnita, es verdadera.

Existen otras formas de caracterizar la compatibilidad (o incompatibilidad) de un sistema. Una es observando que si la última columna $\boldsymbol{b}$ de la matriz ampliada $[A|\boldsymbol{b}]$ es una columna no básica, entonces no puede haber un pivote en la última columna y, por tanto, el sistema es compatible, porque la situación

2.3.1 no puede ocurrir. Recíprocamente, si el sistema es compatible, entonces la situación 2.3.1 no puede ocurrir, y en consecuencia la última columna no puede ser básica. En otras palabras, $[A|\boldsymbol{b}]$ es compatible si y solamente si $\boldsymbol{b}$ no es columna básica.

Decir que $\boldsymbol{b}$ no es columna básica en $[A|\boldsymbol{b}]$ es equivalente a decir que todas las columnas básicas de $[A|\boldsymbol{b}]$ están en la matriz de coeficientes A. Como el número de columnas básicas es el rango, la compatibilidad puede ser caracterizada diciendo que un sistema es compatible si y sólo si $\text{rango}([A|\boldsymbol{b}]) = \text{rango}(A)$.

Recordemos que una columna no básica se puede expresar como combinación lineal de las columnas básicas. Como un sistema compatible se caracteriza porque el lado derecho $\boldsymbol{b}$ es una columna no básica, se sigue que un sistema es compatible si $\boldsymbol{b}$ es una combinación lineal de las columnas de la matriz de coeficientes A. Resumimos todas estas condiciones.

Compatibilidad

Cada uno de las siguientes enunciados es equivalente a que el sistema con matriz ampliada $[A|\boldsymbol{b}]$ es compatible.

- En la reducción por filas de $[A|\boldsymbol{b}]$, nunca aparece una fila de la forma
$$\begin{pmatrix} 0 & 0 & \dots & 0 & | & \alpha \end{pmatrix}, \alpha \neq 0.$$
- $\boldsymbol{b}$ es una columna no básica de $[A|\boldsymbol{b}]$.
- $\text{rango}([A|\boldsymbol{b}]) = \text{rango}(A)$.
- $\boldsymbol{b}$ es combinación lineal de las columnas de A.

Ejemplo 2.3.1. Determinemos si el sistema

$$\begin{cases} x_1 + x_2 + 2x_3 + 2x_4 + x_5 = 1, \\ 2x_1 + 2x_2 + 4x_3 + 4x_4 + 3x_5 = 1, \\ 2x_1 + 2x_2 + 4x_3 + 4x_4 + 2x_5 = 2, \\ 3x_1 + 5x_2 + 8x_3 + 6x_4 + 5x_5 = 3, \end{cases}$$

es compatible. Aplicamos eliminación gaussiana a la matriz ampliada $[A|\boldsymbol{b}]$.

$$\left(\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 2 & 1 & 1 \\ 2 & 2 & 4 & 4 & 3 & 1 \\ 2 & 2 & 4 & 4 & 2 & 2 \\ 3 & 5 & 8 & 6 & 5 & 3 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 2 & 1 & 1 \\ 0 & \boxed{0} & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 2 & 0 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccc|c} \boxed{1} & 1 & 2 & 2 & 1 & 1 \\ 0 & \boxed{2} & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Como no hay ninguna fila de la forma $\begin{pmatrix} 0 & 0 & \dots & 0 & | & \alpha \end{pmatrix}$, con $\alpha \neq 0$, el sistema es compatible. También observamos que $\boldsymbol{b}$ no es una columna básica en $[A|\boldsymbol{b}]$, por lo que rango$([A|\boldsymbol{b}])$ = rango(A). Los pivotes nos indican también que $\boldsymbol{b}$ es combinación lineal de A_{*1}, A_{*2} y A_{*5}. En concreto, como la forma escalonada reducida por filas es

$$\left(\begin{array}{ccccc|c} 1 & 0 & 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

vemos que $\boldsymbol{b} = A_{*1} + A_{*2} - A_{*5}$.

Ejemplo 2.3.2. Estudiemos ahora si el sistema

$$\left\{\begin{array}{rcrcrcr} 6x_1 & + & 4x_2 & + & 7x_3 & = & -1, \\ 3x_1 & + & 2x_2 & - & 5x_3 & = & 4, \\ 3x_1 & + & 2x_2 & - & 2x_3 & = & 5 \end{array}\right.$$

es compatible. Aplicamos la eliminación gaussiana a la matriz ampliada:

$$\left[\begin{array}{ccc|c} 6 & 4 & 7 & -1 \\ 3 & 2 & -5 & 4 \\ 3 & 2 & -2 & 5 \end{array}\right] \rightarrow \left(\begin{array}{ccc|c} 6 & 4 & 7 & -1 \\ 0 & 0 & -17/2 & 9/2 \\ 0 & 0 & -11/2 & 11/2 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccc|c} 6 & 4 & 7 & -1 \\ 0 & 0 & -17/2 & 9/2 \\ 0 & 0 & 0 & \frac{44}{17} \end{array}\right).$$

Aparece un pivote en la columna correspondiente al término independiente, que representa una ecuación de la forma $0 \cdot x_3 = 44/17$. Por tanto, el sistema es incompatible.

2.4. Sistemas homogéneos

Un sistema de m ecuaciones y n incógnitas

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \dots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \dots & + & a_{2n}x_n & = & 0 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \dots & + & a_{mn}x_n & = & 0, \end{array}$$

en el que el lado derecho contiene únicamente ceros se denomina ***homogéneo***. Si al menos unos de los coeficientes de la derecha es no nulo, decimos que es ***no homogéneo***. En esta sección vamos a examinar algunas de las propiedades más elementales de los sistemas homogéneos.

La compatibilidad nunca es un problema con un sistema homogéneo, pues $x_1 = x_2 = \cdots = x_n = 0$ siempre es una solución del sistema, independientemente de los coeficientes. Esta solución se denomina ***solución trivial***. La pregunta es si hay otras soluciones además de la trivial, y cómo podemos describirlas. Como antes, la eliminación gaussiana nos dará la respuesta.

Mientras reducimos la matriz ampliada $[A|\mathbf{0}]$ de un sistema homogéneo a una forma escalonada mediante la reducción gaussiana, la columna de ceros de la derecha no se ve alterada por ninguna de las operaciones elementales. Así, cualquier forma escalonada derivada de $[A|\mathbf{0}]$ tendrá la forma $[E|\mathbf{0}]$. Esto significa que la columna de ceros puede ser eliminada a la hora de efectuar los cálculos. Simplemente reducimos la matriz A a una forma escalonada E, y recordamos que el lado derecho es cero cuando procedamos a la sustitución hacia atrás. El proceso se comprende mejor con un ejemplo.

Ejemplo 2.4.1. Vamos a examinar las soluciones del sistema homogéneo

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & 2x_3 & + & 3x_4 & = & 0, \\ 2x_1 & + & 4x_2 & + & x_3 & + & 3x_4 & = & 0, \\ 3x_1 & + & 6x_2 & + & x_3 & + & 4x_4 & = & 0. \end{array} \tag{2.4.1}$$

Reducimos la matriz de coeficientes a una forma escalonada por filas:

$$A = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 2 & 3 \\ 0 & 0 & -3 & -3 \\ 0 & 0 & -5 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 2 & 3 \\ 0 & 0 & -3 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix} = E. \tag{2.4.2}$$

Entonces, el sistema homogéneo inicial es equivalente al sistema homogéneo

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & 2x_3 & + & 3x_4 & = & 0, \\ & & & & -3x_3 & - & 3x_4 & = & 0. \end{array}$$

Como hay cuatro incógnitas y solamente dos ecuaciones, es imposible extraer una solución única para cada incógnita. Lo mejor que podemos hacer es elegir dos incógnitas básicas, que llamaremos ***variables básicas***, y resolver el sistema en función de las otras dos, que llamaremos ***variables libres***. Aunque hay distintas posibilidades para escoger las variables básicas, el convenio es siempre resolver las incógnitas que se encuentran en las posiciones pivote.

En este ejemplo, los pivotes, así como las columnas básicas, están en la primera y tercera posición, por lo que la estrategia es aplicar sustitución hacia atrás en la resolución del sistema, y expresar las variables básicas x_1 y x_3 en función de las variables libres x_2 y x_4.

La segunda ecuación nos da

$$x_3 = -x_4$$

y la sustitución hacia atrás produce

$$\begin{aligned} x_1 &= -2x_2 - 2x_3 - 3x_4 \\ &= -2x_2 - 2(-x_4) - 3x_4 \\ &= \qquad\qquad\qquad\qquad -2x_2 - x_4. \end{aligned}$$

Las soluciones del sistema homogéneo original pueden ser descritas como

$$\begin{aligned} x_1 &= -2x_2 - x_4, \\ x_2 &= \text{ libre }, \\ x_3 &= -x_4, \\ x_4 &= \text{ libre.} \end{aligned}$$

Las expresiones anteriores describen todas las soluciones.

Mejor que describir las soluciones de esta forma, es más conveniente expresarlas como

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2x_2 - x_4 \\ x_2 \\ -x_4 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix},$$

entendiendo que x_2 y x_4 son variables libres que pueden tomar cualquier valor. Esta representación se denominará ***solución general*** del sistema homogéneo. Esta expresión de la solución general enfatiza que cada solución es combinación lineal de las dos soluciones

$$\boldsymbol{h}_1 = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{h}_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}.$$

Consideremos ahora un sistema homogéneo general $[A|\mathbf{0}]$ de m ecuaciones y n incógnitas. Si la matriz de coeficientes A es de rango r, entonces, por lo que hemos visto antes, habrá r variables básicas, correspondientes a las posiciones de las columnas básicas de A, y $n-r$ variables libres, que se corresponden con las columnas no básicas de A. Mediante la reducción de A a una forma escalonada por filas por eliminación gaussiana y sustitución hacia atrás, expresamos las variables básicas en función de las variables libres y obtenemos la ***solución general***, de la forma

$$\boldsymbol{x} = x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r},$$

donde $x_{f_1}, x_{f_2}, \ldots, x_{f_{n-r}}$ son las variables libres, y $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_{n-r}$ son vectores columna que representan soluciones particulares.

Observemos que el vector $\boldsymbol{h}_1$ tiene un 1 en la posición f_1, y los restantes vectores $\boldsymbol{h}_j$ tienen un cero en esa posición. Lo mismo se aplica a todos los vectores $\boldsymbol{h}_i$: tienen un valor 1 en la posición f_i, y los restantes vectores $\boldsymbol{h}_j$ tienen un cero en esa posición.

Si calculamos la forma escalonada reducida por filas del ejemplo, nos queda

$$A = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = E_A,$$

y el sistema a resolver es

$$\begin{aligned} x_2 + 2x_2 + x_4 &= 0, \\ x_3 + x_4 &= 0. \end{aligned}$$

Si resolvemos x_1 y x_3 en función de x_2 y x_4 obtenemos el mismo resultado que antes. Por ello, y para evitar la sustitución hacia atrás, puede resultar más conveniente usar Gauss-Jordan para calcular la forma escalonada reducida por filas E_A y construir directamente la solución general a partir de las entradas de E_A.

Una última pregunta que nos planteamos es cuándo la solución trivial de un sistema homogéneo es la única solución. Lo anterior nos muestra la respuesta. Si hay al menos una variable libre, entonces el sistema tendrá infinitas soluciones. Por tanto, la solución trivial será la única solución si y solamente si no hay variables libres, esto es, $n-r=0$. Podemos reformular esto diciendo que un sistema homogéneo tiene únicamente la solución trivial si y solamente si $\text{rango}(A) = n$.

Ejemplo 2.4.2. El sistema homogéneo

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 0, \\ 2x_1 + 5x_2 + 7x_3 &= 0, \\ 3x_1 + 6x_2 + 8x_3 &= 0, \end{aligned}$$

tiene solamente la solución trivial porque

$$A=\begin{pmatrix}1&2&2\\2&5&7\\3&6&8\end{pmatrix}\to\begin{pmatrix}1&2&2\\0&1&3\\0&0&2\end{pmatrix}=E$$

prueba que rango$(A)=3=n$. Se ve fácilmente que la aplicación de la sustitución hacia atrás desde $[E|\mathbf{0}]$ únicamente devuelve la solución trivial.

Ejemplo 2.4.3. Calculemos la solución general del sistema

$$\begin{array}{rcrcrcl}x_1&+&2x_2&+&2x_3&=&0,\\2x_1&+&5x_2&+&7x_3&=&0,\\3x_1&+&6x_2&+&6x_3&=&0.\end{array}$$

Se tiene que

$$A=\begin{pmatrix}1&2&2\\2&5&7\\3&6&6\end{pmatrix}\to\begin{pmatrix}1&2&2\\0&1&3\\0&0&0\end{pmatrix}=E,$$

de donde rango$(A)=2<n=3$. Como las columnas básicas están en las posiciones uno y dos, x_1 y x_2 son las variables básicas, y x_3 es libre. Mediante sustitución hacia atrás en $[E|\mathbf{0}]$, nos queda $x_2=-3x_3$ y $x_1=-2x_2-2x_3=4x_3$, y la solución general es

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=x_3\begin{pmatrix}4\\-3\\1\end{pmatrix},\text{ donde } x_3 \text{ es libre.}$$

2.5. Sistemas no homogéneos

Recordemos que un sistema de m ecuaciones y n incógnitas

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

es ***no homogéneo*** cuando $b_i\neq 0$ para algún i. A diferencia de los sistemas homogéneos, los no homogéneos pueden ser incompatibles y las técnicas que conocemos las aplicaremos para saber si una solución existe. A menos que se diga lo contrario, suponemos que los sistemas de esta sección son compatibles.

Para describir el conjunto de todas las posibles soluciones de un sistema no homogéneo compatible, vamos a construir una solución general de la misma forma que hicimos para los homogéneos.

- Usamos eliminación gaussiana para reducir la matriz ampliada $[A|\boldsymbol{b}]$ a una forma escalonada por filas $[E|\boldsymbol{c}]$.
- Identificamos las variables básicas y las libres.
- Aplicamos sustitución hacia atrás a $[E|\boldsymbol{c}]$ y resolvemos las variables básicas en función de las libres.
- Escribimos el resultado en la forma

$$\boldsymbol{x} = \boldsymbol{p} + x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r},$$

donde $x_{f_1}, x_{f_2}, \ldots, x_{f_{n-r}}$ son las variables libres, y $\boldsymbol{p}, \boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_{n-r}$ son vectores columna de orden n. Esta es la ***solución general*** del sistema no homogéneo.

Como las variables libres x_{f_i} recorren todos los posibles valores, la solución general contiene todas las posibles soluciones del sistema $[A|\boldsymbol{b}]$. Como en el caso homogéneo, podemos reducir completamente $[A|\boldsymbol{b}]$ a $E_{[A|\boldsymbol{b}]}$ mediante Gauss-Jordan y evitamos la sustitución hacia atrás.

La diferencia entre la solución general de un sistema no homogéneo y la de uno homogéneo es la columna $\boldsymbol{p}$ que aparece. Para entender de dónde viene, consideremos el sistema no homogéneo

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 4, \\ 2x_1 & + & 4x_2 & + & x_3 & + & 3x_4 & = & 5, \\ 3x_1 & + & 6x_2 & + & x_3 & + & 4x_4 & = & 7, \end{array}$$

en el que la matriz de coeficientes es la misma que la matriz de coeficientes de 2.4.1. Si $[A|\boldsymbol{b}]$ se reduce por Gauss-Jordan a $E_{[A|\boldsymbol{b}]}$, tenemos

$$[A|\boldsymbol{b}] \to \begin{pmatrix} 1 & 2 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 & 5 \\ 3 & 6 & 1 & 4 & 7 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = E_{[A|\boldsymbol{b}]}.$$

Nos queda el sistema equivalente

$$\begin{array}{rcrcrcl} x_1 & + & 2x_2 & + & x_4 & = & 2, \\ & & x_3 & + & x_4 & = & 1. \end{array}$$

Resolvemos las variables básicas x_1 y x_3 en función de las libres x_2 y x_4, y obtenemos

$$\begin{array}{rcl} x_1 & = & 2 - 2x_2 - x_4, \\ x_2 & & \text{es libre}, \\ x_3 & = & 1 - x_4, \\ x_4 & & \text{es libre}. \end{array}$$

La solución general se sigue escribiendo estas ecuaciones en la forma

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2-x_2-x_4 \\ x_2 \\ 1-x_4 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}. \tag{2.5.1}$$

La columna

$$\begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

que aparece en la expresión (2.5.1) es una *solución particular* del sistema no homogéneo; se tiene cuando $x_2 = 0, x_4 = 0$.

Además, la solución general del sistema homogéneo

$$\begin{array}{rcrcrcrcl} x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 0, \\ 2x_1 & + & 4x_2 & + & x_3 & + & 3x_4 & = & 0, \\ 3x_1 & + & 6x_2 & + & x_3 & + & 4x_4 & = & 0, \end{array} \tag{2.5.2}$$

es

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}.$$

Así, la solución general del sistema homogéneo (2.5.2) es una parte de la solución general del sistema no homogéneo original (2.5.1).

Estas dos observaciones se pueden combinar diciendo que *la solución general del sistema no homogéneo viene dado por una solución particular más la solución general del sistema homogéneo asociado.*

Veamos que esta observación es siempre cierta. Supongamos que $[A|\boldsymbol{b}]$ representa un sistema $m \times n$ compatible, donde $\text{rango}(A) = r$. La compatibilidad garantiza que $\boldsymbol{b}$ no es una columna básica de $[A|\boldsymbol{b}]$, por lo que las columnas básicas de $[A|\boldsymbol{b}]$ están en la misma posición que las columnas básicas de $[A|\boldsymbol{0}]$. Esto significa que el sistema no homogéneo y el sistema homogéneo asociado tienen exactamente el mismo conjunto de variables básicas así como de libres. Además, no es difícil ver que

$$E_{[A|\boldsymbol{0}]} = [E_A|\boldsymbol{0}] \text{ y } E_{[A|\boldsymbol{b}]} = [E_A|\boldsymbol{c}],$$

donde $\boldsymbol{c}$ es una columna de la forma

$$\boldsymbol{c} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_r \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Si resolvemos la i-ésima ecuación en el sistema homogéneo reducido para la i-ésima variable básica x_{b_i} en función de las variables libres $x_{f_1}, x_{f_2}, \ldots, x_{f_{n-r}}$ para conseguir

$$x_{b_i} = \alpha_i x_{f_i} + \alpha_{i+1} x_{f_{i+1}} + \cdots + \alpha_{n-r} x_{f_{n-r}},$$

entonces la solución de la i-ésima variable básica en el sistema **no homogéneo** reducido debe tener la forma

$$x_{b_i} = \xi_i + \alpha_i x_{f_i} + \alpha_{i+1} x_{f_{i+1}} + \cdots + \alpha_{n-r} x_{f_{n-r}}.$$

Esto es, las dos soluciones se diferencian únicamente en la presencia de la constante ξ_i en la última. Si organizamos como columnas las expresiones anteriores, podemos decir que si la solución general del sistema homogéneo es de la forma

$$\boldsymbol{x} = x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r},$$

entonces la solución general del sistema no homogéneo tiene la forma similar

$$\boldsymbol{x} = \boldsymbol{p} + x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r},$$

donde la columna $\boldsymbol{p}$ contiene las constantes ξ_i junto con ceros.

Ejemplo 2.5.1. Calculemos la solución general del sistema

$$\begin{array}{rcrcrcrcrcl} x_1 & + & x_2 & + & 2x_3 & + & 2x_4 & + & x_5 & = & 1, \\ 2x_1 & + & 2x_2 & + & 4x_3 & + & 4x_4 & + & 3x_5 & = & 1, \\ 2x_1 & + & 2x_2 & + & 4x_3 & + & 4x_4 & + & 2x_5 & = & 2, \\ 3x_1 & + & 5x_2 & + & 8x_3 & + & 6x_4 & + & 5x_5 & = & 3, \end{array}$$

y la comparamos con la solución general del sistema homogéneo asociado.

En primer lugar, calculamos la forma escalonada reducida por filas de la matriz ampliada $[A|\boldsymbol{b}]$.

$$
\begin{aligned}
[A|\boldsymbol{b}] &= \left(\begin{array}{ccccc|c} 1&1&2&2&1&1\\ 2&2&4&4&3&1\\ 2&2&4&4&2&2\\ 3&5&8&6&5&3 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 1&1&2&2&1&1\\ 0&0&0&0&1&-1\\ 0&0&0&0&0&0\\ 0&2&2&0&2&0 \end{array}\right)\\
&\rightarrow \left(\begin{array}{ccccc|c} 1&1&2&2&1&1\\ 0&2&2&0&2&0\\ 0&0&0&0&1&-1\\ 0&0&0&0&0&0 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 1&1&2&2&1&1\\ 0&1&1&0&1&0\\ 0&0&0&0&1&-1\\ 0&0&0&0&0&0 \end{array}\right)\\
&\rightarrow \left(\begin{array}{ccccc|c} 1&0&1&2&0&1\\ 0&1&1&0&1&0\\ 0&0&0&0&1&-1\\ 0&0&0&0&0&0 \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 1&0&1&2&0&1\\ 0&1&1&0&0&1\\ 0&0&0&0&1&-1\\ 0&0&0&0&0&0 \end{array}\right) = E_{[A|\boldsymbol{b}]}.
\end{aligned}
$$

El sistema es compatible, pues la última columna es no básica. Resolvemos el sistema reducido para las variables básicas x_1, x_2, x_5 en función de las variables libres x_3, x_4 para obtener

$$
\begin{array}{lcl} x_1 &=& 1-x_3-2x_4,\\ x_2 &=& 1-x_3,\\ x_3 && \text{es libre},\\ x_4 && \text{es libre},\\ x_5 &=& -1. \end{array} \quad \Rightarrow \quad \begin{array}{lcllll} x_1 &=& 1 & -x_3 & -2x_4,\\ x_2 &=& 1 & -x_3, & \\ x_3 &=& & x_3, & \\ x_4 &=& & & x_4,\\ x_5 &=& -1. & & \end{array}
$$

La solución general del sistema no homogéneo es

$$
\boldsymbol{x} = \begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5 \end{pmatrix} = \begin{pmatrix} 1-x_3-2x_4\\ 1-x_3\\ x_3\\ x_4\\ -1 \end{pmatrix} = \begin{pmatrix} 1\\ 1\\ 0\\ 0\\ -1 \end{pmatrix} + x_3 \begin{pmatrix} -1\\ -1\\ 1\\ 0\\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -2\\ 0\\ 0\\ 1\\ 0 \end{pmatrix}.
$$

La solución general del sistema homogéneo asociado es

$$
\boldsymbol{x} = \begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5 \end{pmatrix} = \begin{pmatrix} -x_3-2x_4\\ -x_3\\ x_3\\ x_4\\ 0 \end{pmatrix} = x_3 \begin{pmatrix} -1\\ -1\\ 1\\ 0\\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -2\\ 0\\ 0\\ 1\\ 0 \end{pmatrix}.
$$

Ahora volvemos a la pregunta: ¿cuándo un sistema compatible tiene solución única? Sabemos que la solución general de un sistema no homogéneo compatible de orden $m \times n$, con rango r, es de la forma

$$\boldsymbol{x} = \boldsymbol{p} + x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r},$$

donde

$$x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r}$$

es la solución general del sistema homogéneo asociado. Por tanto, es evidente que el sistema $[A|\boldsymbol{b}]$ tendrá una única solución si y solamente si no hay variables libres, esto es, si y solamente si $r = n$. Esto es lo mismo que decir que el sistema homogéneo asociado $[A|\mathbf{0}]$ tiene solamente la solución trivial.

Ejemplo 2.5.2. Consideremos el siguiente sistema no homogéneo:

$$\begin{array}{rcrcrcr} 2x_1 & + & 4x_2 & + & 6x_3 & = & 2, \\ x_1 & + & 2x_2 & + & 3x_3 & = & 1, \\ x_1 & & & + & x_3 & = & -3, \\ 2x_1 & + & 4x_2 & & & = & 8. \end{array}$$

La forma escalonada reducida por filas de $[A|\boldsymbol{b}]$ es

$$[A|\boldsymbol{b}] = \left(\begin{array}{ccc|c} 2 & 4 & 6 & 2 \\ 1 & 2 & 3 & 1 \\ 1 & 0 & 1 & -3 \\ 2 & 4 & 0 & 8 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right) = E_{[A|\boldsymbol{b}]}.$$

El sistema es compatible porque la última columna no es básica, o bien porque rango$(A) = 3 =$ número de incógnitas (no hay variables libres). El sistema homogéneo asociado tiene únicamente la solución trivial, y la solución del sistema es

$$\boldsymbol{p} = \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix}.$$

Resumen

Sea $[A|\boldsymbol{b}]$ la matriz ampliada de un sistema no homogéneo compatible, de orden $m \times n$, con rango$(A) = r$.

- Mediante la reducción de $[A|\boldsymbol{b}]$ a una forma escalonada usando la eliminación gaussiana, resolvemos las variables básicas en función de las libres y llegamos a que la ***solución general*** del sistema es de la forma

$$\boldsymbol{x} = \boldsymbol{p} + x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r}.$$

- La columna $\boldsymbol{p}$ es una solución particular del sistema no homogéneo.
- La expresión

$$x_{f_1}\boldsymbol{h}_1 + x_{f_2}\boldsymbol{h}_2 + \cdots + x_{f_{n-r}}\boldsymbol{h}_{n-r}$$

  es la solución general del sistema homogéneo asociado.
- El sistema tiene una solución única si y solamente si se verifica alguna de las siguientes condiciones:
  - rango$(A) = n$ = número de incógnitas.
  - No hay variables libres.
  - El sistema homogéneo asociado solamente tiene la solución trivial.

Capítulo 3

Álgebra matricial

3.1. Adición y trasposición

El conjunto de los números reales se notará por $\mathbb{R}$, y el de los números complejos por $\mathbb{C}$. Al principio, no hay mucho inconveniente en pensar únicamente en números reales, pero después se hará inevitable el uso de números complejos.

El conjunto de n-uplas de números reales se notará por $\mathbb{R}^n$, y el conjunto de n-uplas de números complejos por $\mathbb{C}^n$. Análogamente, $\mathbb{R}^{m\times n}$ y $\mathbb{C}^{m\times n}$ denotarán las matrices de orden $m \times n$ que contienen números reales y complejos, respectivamente.

Dos matrices $A = (a_{ij})$ y $B = (b_{ij})$ son ***iguales*** cuando A y B tienen la misma forma y las entradas correspondientes son iguales.

Esta definición se aplica a matrices como

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ y } \boldsymbol{v} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}.$$

Aunque podamos pensar que $\boldsymbol{u}$ y $\boldsymbol{v}$ describen el mismo punto en $\mathbb{R}^3$, no podemos decir que sean iguales como matrices, pues sus formas son diferentes.

Una matriz formada por una sola columna se denomina ***vector columna***, y si tiene una sola fila se llama ***vector fila***.

Suma de matrices

Si A y B son matrices de orden $m \times n$, la suma de A y B se define como la matriz de orden $m \times n$ notada por $A+B$, cuyas entradas verifican

$$[A+B]_{ij} = [A]_{ij} + [B]_{ij} \text{ para cada } i, j.$$

La matriz $-A$, llamada ***opuesta*** de A, se define como

$$[-A]_{ij} = -[A]_{ij}.$$

La ***diferencia*** de A y B es

$$A - B = A + (-B).$$

Ejemplo 3.1.1. Sean

$$A = \begin{pmatrix} 2 & -3 & 4 & 0 \\ 1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} -4 & -2 & 0 & 3 \\ -3 & 1 & -1 & -2 \\ -1 & 4 & 2 & -1 \end{pmatrix}.$$

Entonces

$$A+B = \begin{pmatrix} -2 & -5 & 4 & 3 \\ -2 & -1 & 0 & -1 \\ -1 & 4 & 2 & -1 \end{pmatrix}, -A = \begin{bmatrix} -2 & 3 & -4 & 0 \\ -1 & 2 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, A-B = \begin{bmatrix} 6 & -1 & 4 & -3 \\ 4 & -3 & 2 & 3 \\ 1 & -4 & -2 & 1 \end{bmatrix}.$$

Si

$$C = \begin{bmatrix} -1 & -3 & -2 & -2 & 1 \\ -3 & -3 & 0 & 1 & -3 \\ 1 & -2 & 3 & 2 & 1 \end{bmatrix},$$

no podemos calcular $A+C$, pues A es de orden 3×4 y C es de orden 3×5.

Propiedades de la suma de matrices

Sean A, B y C matrices de orden $m \times n$. Se verifican las siguientes propiedades:

- $A+B$ es una matriz de orden $m \times n$.
- $(A+B)+C = A+(B+C)$.
- $A+B = B+A$.
- La matriz $0_{m\times n}$ que tiene todas sus entradas nulas verifica $A+0 = A$.
- La matriz $-A$ es de orden $m \times n$ y verifica $A+(-A) = 0_{m\times n}$.

Multiplicación por un escalar

El producto de un escalar α por una matriz A de orden $m \times n$, notada por αA, se define como la matriz de orden $m \times n$ que verifica

$$[\alpha A]_{ij} = \alpha [A]_{ij}.$$

Ejemplo 3.1.2. Si

$$A = \begin{bmatrix} 2 & -3 & 4 & 0 \\ 1 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

entonces

$$(-3)\cdot A = \begin{bmatrix} -6 & 9 & -12 & 0 \\ -3 & 6 & -3 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}, 0\cdot A \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Propiedades de la multiplicación por un escalar

Sean A, B matrices de orden $m \times n$, y α, β escalares.

- αA es una matriz $m \times n$.
- $(\alpha\beta)A = \alpha(\beta A)$.
- $\alpha(A+B) = \alpha A + \alpha B$.
- $(\alpha+\beta)A = \alpha A + \beta A$.
- $1 \cdot A = A$.

Se tienen propiedades análogas para $A\alpha = \alpha A$.

Trasposición

La traspuesta de una matriz $A_{m\times n}$ es la matriz notada por A^t de orden $n \times m$ definida como

$$[A^t]_{ij} = [A]_{ji}.$$

La matriz ***conjugada*** de una matriz $A_{m\times n}$ es la matriz de orden $m \times n$ notada por $\bar{A}$ definida como

$$[\overline{A}]_{ij} = \overline{[A]_{ij}}.$$

La matriz ***conjugada traspuesta*** de una matriz $A_{m\times n}$ es la matriz de orden $n \times m$ notada por A^* y definida como

$$[A^*]_{ij} = \overline{[A]_{ji}}.$$

Ejemplo 3.1.3. Sea

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

Entonces

$$A^t = \begin{bmatrix} 2 & 1 & 0 \\ -3 & -2 & 0 \\ 4 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, A^* = \begin{bmatrix} 2 & 1 & 0 \\ -3 & -2 & 0 \\ 4 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Para que haya alguna diferencia entre A^t y A^* debemos emplear matrices con entradas complejas. Por ejemplo, si

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 1-i \\ i \\ 2 \end{pmatrix}, \text{ entonces } \boldsymbol{u}^* = \begin{pmatrix} 1 & 1+i & -i & 2 \end{pmatrix}.$$

Es evidente que $(A^t)^t = A, (A^*)^* = A$. En el caso de matrices **reales**, $\overline{A} = A$ y $A^* = A^t$.

Propiedades de la matriz traspuesta

Sean A y B matrices de la misma forma y α un escalar. Entonces

- $(A+B)^t = A^t + B^t$ y $(A+B)^* = A^* + B^*$.
- $(\alpha A)^t = \alpha A^t$ y $(\alpha A)^* = \overline{\alpha} A^*$.

Simetrías

Sea $A = (a_{ij})$ una matriz cuadrada.

- Decimos que A es simétrica si $A = A^t$, esto es, $a_{ij} = a_{ji}$.
- Decimos que A es anti-simétrica si $A = -A^t$, esto es, $a_{ij} = -a_{ji}$.
- Decimos que A es hermitiana si $A = A^*$, esto es, $a_{ij} = \overline{a_{ji}}$.
- Decimos que A es anti-hermitiana si $A = -A^*$, esto es, $a_{ij} = -\overline{a_{ji}}$.

Ejemplo 3.1.4. Sean

$$A = \begin{bmatrix} 1 & 0 & 5 \\ 0 & -3 & 2 \\ 5 & 2 & -3 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & 2 \\ 5 & 2 & -3 \end{bmatrix}, C = \begin{pmatrix} 1 & 1+i \\ 1-i & 3 \end{pmatrix}.$$

Entonces A es simétrica, B no es simétrica y C es hermitiana.

Un número asociado a una matriz cuadrada es la ***traza***. Si $A = (a_{ij})$ es una matriz cuadrada de orden n, entonces la traza de A es el número $\text{traza}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$, es decir, la suma de sus elementos diagonales. Por ejemplo, si

$$A = \begin{bmatrix} 1 & 1 & 5 \\ 1 & -3 & 2 \\ 5 & 2 & -3 \end{bmatrix}, \text{ entonces } \text{traza}(A) = 1 + (-3) + (-3) = -5.$$

3.2. Multiplicación matricial

- Dos matrices A y B se dicen ***ajustadas*** para multiplicación en el orden AB cuando el número de columnas de A es igual al número de filas de B, esto es, si A es de orden $m \times p$ y B es de orden $p \times n$.

- Para matrices ajustadas $A_{m\times p} = (a_{ij})$ y $B_{p\times n} = (b_{ij})$, la ***matriz producto*** AB se define como

$$[AB]_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

Puede ocurrir que exista AB, pero que no tenga sentido BA. Aun cuando tengan sentido los dos productos, la multiplicación matricial no es conmutativa. Considere lo que ocurre al tomar

$$A = \begin{pmatrix} 1 & 2 \end{pmatrix}, B = \begin{pmatrix} 3 \\ 4 \end{pmatrix},$$

y calcular AB y BA.

Filas y columnas de un producto

Supongamos que $A_{m\times p} = (a_{ij})$ y $B_{p\times n} = (b_{ij})$.

- $[AB]_{i*} = A_{i*}B$; esto es, la i-ésima fila de AB es la i-esima fila de A multiplicada por B.
- $[AB]_{*j} = AB_{*j}$; esto es, la j-ésima columna de AB es A multiplicada por la j-ésima columna de B.
- $[AB]_{i*} = a_{i1}B_{1*} + a_{i2}B_{2*} + \cdots + a_{ip}B_{p*} = \sum_{k=1}^{p} a_{ik}B_{k*}$.
- $[AB]_{*j} = A_{*1}b_{1j} + A_{*2}b_{2j} + \cdots + A_{*p}b_{pj} = \sum_{k=1}^{p} A_{*k}b_{kj}$.

PRUEBA: Las dos primeras propiedades son inmediatas a partir de la definición. Para la tercera, se tiene

$$\begin{aligned}(AB)_{i*} &= \begin{pmatrix} c_{i1} & c_{i2} & \dots & c_{in} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=1}^{p} a_{ik}b_{k1} & \sum_{k=1}^{p} a_{ik}b_{k2} & \dots & \sum_{k=1}^{p} a_{ik}b_{kn} \end{pmatrix} \\ &= \sum_{k=1}^{p} a_{ik}\begin{pmatrix} b_{k1} & b_{k2} & \dots & b_{kn} \end{pmatrix} = \sum_{k=1}^{p} a_{ik}B_{k*}.\end{aligned}$$

La cuarta propiedad es análoga. □

Las dos últimas ecuaciones indican que las filas de AB son combinación lineal de las filas de B, y que las columnas de AB son combinación lineal de las columnas de A.

Sistemas lineales

Todo sistema de m ecuaciones y n incógnitas

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

se puede escribir en forma matricial como $A\boldsymbol{x} = \boldsymbol{b}$, donde

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Recíprocamente, toda ecuación matricial $A_{m\times n}\boldsymbol{x}_{n\times 1} = \boldsymbol{b}_{m\times 1}$ representa un sistema lineal de m ecuaciones y n incógnitas.

Ejemplo 3.2.1. El sistema de ecuaciones

$$\left\{ \begin{array}{rrrrcr} & & x_3 & +x_4 & = & 1, \\ -2x_1 & -4x_2 & +x_3 & & = & -1, \\ 3x_1 & +6x_2 & -x_3 & +x_4 & = & 2. \end{array} \right.$$

se escribe como $A\boldsymbol{x} = \boldsymbol{b}$, donde

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ -2 & -4 & 1 & 0 \\ 3 & 6 & -1 & 1 \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

3.3. Propiedades de la multiplicación matricial

Propiedades distributiva y asociativa

Para matrices ajustadas se verifica

- $A(B+C) = AB + AC$.
- $(D+E)F = DF + EF$.
- $A(BC) = (AB)C$.

PRUEBA:

- Supongamos que $A_{m\times p}, B_{p\times n}, C_{p\times n}$. Sea $G = B + C$ y llamamos $g_{ij} = b_{ij} + c_{ij}$; identificamos de forma análoga los elementos de las matrices $H = AM, Y = AB, Z = AC$. Entonces

$$\begin{aligned} h_{ij} &= \sum_{k=1}^{p} a_{ik} g_{kj} = \sum_{k=1}^{p} a_{ik}(b_{kj} + c_{kj}) \\ &= \sum_{k=1}^{p} a_{ik} b_{kj} + \sum_{k=1}^{p} a_{ik} c_{kj} = y_{ij} + z_{ij}, \end{aligned}$$

  o bien que $H = Y + Z$.

- Se prueba de manera similar a la anterior.

- Supongamos que $A_{m\times p}, B_{p\times q}, C_{q\times n}$ y llamemos $D = BC, E = AD, F = AB, G = FC$. Entonces

$$\begin{aligned} e_{ij} = \sum_{k=1}^{p} a_{ik} d_{kj} = \sum_{k=1}^{p} a_{ik} \sum_{l=1}^{q} b_{kq} c_{qj} = \sum_{l=1}^{q} \left(\sum_{k=1}^{p} a_{ik} b_{kq} \right) c_{qj} \\ = \sum_{l=1}^{q} f_{iq} c_{qj} = g_{ij}, \end{aligned}$$

  lo que es equivalente a decir que $E = G$.

□

Matriz identidad

La matriz de orden $n \times n$ con unos en la diagonal y ceros en el resto

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

se denomina ***matriz identidad*** de orden n.

Para toda matriz A de orden $m \times n$ se verifica

$$AI_n = A \text{ y } I_m A = A.$$

El subíndice de I_n se elimina cuando el tamaño es obvio por el contexto. Las columnas de I_n se representan por los vectores $e_1, e_2, \dots, e_n$, es decir, el vector e_i es el vector de n componentes cuya componente i-ésima es igual a 1 y las restantes iguales a cero. Observemos que la notación tiene la ambigüedad respecto al tamaño del vector, y se deduce por los tamaños de las matrices que intervengan en la expresión.

Trasposición y producto

Para matrices ajustadas A y B se verifica que

$$(AB)^t = B^t A^t, \text{ y } (AB)^* = B^* A^*.$$

Ejemplo 3.3.1. Para cada matriz $A_{m\times n}$

- las matrices AA^t y $A^t A$ son simétricas, y
- las matrices AA^* y $A^* A$ son hermitianas.

Ejemplo 3.3.2. Para matrices $A_{m\times n}$ y $B_{n\times m}$ se verifica

$$\text{traza}(AB) = \text{traza}(BA).$$

De lo anterior se deduce que $\text{traza}(ABC) = \text{traza}(BCA) = \text{traza}(CAB)$, pero, en general, $\text{traza}(ABC) \neq \text{traza}(BAC)$.

Multiplicación por bloques

Supongamos que A y B se particionan en submatrices, también llamados bloques, como sigue:

$$A=\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1r} \\ A_{21} & A_{22} & \dots & A_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ A_{s1} & A_{s2} & \dots & A_{sr} \end{pmatrix}, B=\begin{pmatrix} B_{11} & B_{12} & \dots & B_{1t} \\ B_{21} & B_{22} & \dots & B_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ B_{r1} & B_{r2} & \dots & B_{rt} \end{pmatrix}.$$

Si los pares (A_{ik}, B_{kj}) son ajustados para el producto, entonces decimos que A y B tienen una ***partición ajustada***. Para tales matrices, el producto AB se forma combinando los bloques exactamente de la misma forma como se hace con los escalares en la multiplicación ordinaria. Esto es, el bloque (i, j) en AB es

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ir}B_{rj}.$$

Ejemplo 3.3.3. Consideremos las matrices particionadas

$$A=\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right)=\begin{pmatrix} C & I \\ I & \mathbf{0} \end{pmatrix}, B=\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 2 & 1 & 2 \\ 3 & 4 & 3 & 4 \end{array}\right)=\begin{pmatrix} I & 0 \\ C & C \end{pmatrix},$$

donde

$$I=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ y } C=\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Mediante la multiplicación por bloques, el producto AB es fácil de obtener:

$$AB=\begin{pmatrix} C & I \\ I & \mathbf{0} \end{pmatrix}\begin{pmatrix} I & 0 \\ C & C \end{pmatrix}=\begin{pmatrix} 2C & C \\ I & \mathbf{0} \end{pmatrix}=\left(\begin{array}{cc|cc} 2 & 4 & 1 & 2 \\ 6 & 8 & 3 & 4 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right)$$

Ejemplo 3.3.4. Reducibilidad. Supongamos que $T_{n\times n}\boldsymbol{x} = \boldsymbol{b}$ representa un sistema de ecuaciones en el que la matriz de coeficientes es ***triangular por bloques***. Esto es, T se puede particionar como

$$T=\begin{pmatrix} A & B \\ \mathbf{0} & C \end{pmatrix}, \text{ donde } A \text{ es } r\times r \text{ y } C \text{ es } (n-r)\times(n-r).$$

Si $\boldsymbol{x}$ y $\boldsymbol{b}$ se particionan de igual forma como

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix},$$

entonces la multiplicación por bloques muestra que $T\boldsymbol{x} = \boldsymbol{b}$ se reduce a dos sistemas más pequeños:

$$\begin{aligned} A\boldsymbol{x}_1 + B\boldsymbol{x}_2 &= \boldsymbol{b}_1, \\ C\boldsymbol{x}_2 &= \boldsymbol{b}_2, \end{aligned}$$

3.4. Inversa de una matriz

Inversa de una matriz

Para una matriz cuadrada $A_{n\times n}$, la matriz $B_{n\times n}$ que verifica las condiciones

$$AB = I_n \text{ y } BA = I_n$$

se denomina ***inversa*** de A, y la notaremos por $B = A^{-1}$. No todas la matrices cuadradas tienen inversa. Una matriz con inversa se denomina ***no singular*** y una matriz cuadrada sin inversa se llama ***singular***.

Aunque no todas las matrices tienen inversa, cuando existe, es única. Supongamos que X_1 y X_2 son inversas de una matriz no singular A. Entonces

$$X_1 = X_1 I = X_1(AX_2) = (X_1 A)X_2 = IX_2 = X_2.$$

Ecuaciones matriciales

- Si A es una matriz no singular, entonces existe una única solución para X en la ecuación matricial $A_{n\times n}X_{n\times p} = B_{n\times p}$, que es $X = A^{-1}B$.
- Un sistema de n ecuaciones y n incógnitas se puede escribir como una ecuación matricial $A_{n\times n}\boldsymbol{x}_{n\times 1} = \boldsymbol{b}_{n\times 1}$. Por lo anterior, si A es no singular, el sistema tiene solución única igual a $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

Sin embargo, debemos hacer hincapié en que la representación de la solución como $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ es conveniente desde el punto de vista teórico o de notación. En la práctica, un sistema no singular $A\boldsymbol{x} = \boldsymbol{b}$ ***nunca*** se resuelve calculando A^{-1} y entonces el producto $A^{-1}\boldsymbol{b}$. La razón aparecerá cuando estudiemos el coste del cálculo de A^{-1}.

Como no todas las matrices cuadradas tienen inversa, se necesitan métodos para distinguir entre matrices singulares y no singulares. Los más importantes son los que siguen.

Existencia de inversa

Sea A una matriz cuadrada de orden n. Son equivalentes:

1. A^{-1} existe (A es no singular).
2. rango$(A) = n$.
3. $A \xrightarrow{\text{Gauss-Jordan}} I_n$.
4. $A\boldsymbol{x} = \boldsymbol{0}$ implica que $\boldsymbol{x} = \boldsymbol{0}$.

PRUEBA: El hecho de 2) $\Leftrightarrow$ 3) es una consecuencia directa de la definición de rango. La equivalencia 3) $\Leftrightarrow$ 4) la hemos visto en el tema anterior. Solamente falta por establecer 1) $\Leftrightarrow$ 2) para completar la prueba.

1) $\Rightarrow$ 2). Consideremos la matriz $X = \begin{pmatrix} X_{*1} & X_{*2} & \dots & X_{*n} \end{pmatrix}$. Esta matriz X verifica la ecuación $AX = I$ si y solamente si X_{*j} es solución del sistema $A\boldsymbol{x} = I_{*j}$. Si A es no singular, entonces sabemos que existe una solución única de $AX = I$, y por tanto cada sistema $A\boldsymbol{x} = I_{*j}$ tiene solución única. Pero sabemos que un sistema tiene solución única si y solamente si el rango de la matriz de coeficientes es igual al número de incógnitas, esto es, rango$(A) = n$.

2) $\Rightarrow$ 1). Si rango$(A) = n$, entonces cada sistema $A\boldsymbol{x} = I_{*j}$ es compatible, porque rango$([A|I_{*j}]) = n =$ rango(A). Además, la solución es única, por lo que la ecuación matricial $AX = I$ tiene una única solución. Nos gustaría decir ya que $X = A^{-1}$, pero nos hace falta primero probar que $XA = I$. Supongamos que no es cierto, esto es, $XA - I \neq \mathbf{0}$. Como

$$A(XA - I) = AXA - A = IA - A = \mathbf{0},$$

se sigue que cada columna no nula de $XA - I$ es una solución no trivial del sistema homogéneo $A\boldsymbol{x} = \boldsymbol{0}$. Pero esto es una contradicción. Por tanto, $XA - I = \mathbf{0}$, y $XA = AX = I$. □

Como un subproducto de la prueba anterior, hemos probado que si $A_{n\times n}$ es una matriz para la que existe $X_{n\times n}$ con $AX = I_n$, entonces $X = A^{-1}$. En efecto, en el contexto de la prueba necesitamos ver que si existe X tal que $AX = I_n$, entonces rango$(A) = n$. Todos los sistemas $A\boldsymbol{x} = \boldsymbol{e}_i, i = 1,2,\ldots,n$ tienen solución, por lo que cada vector $\boldsymbol{e}_i, i = 1,\ldots,n$ se expresa como combinación lineal de las columnas de A. Esto implica que la matriz $B = \begin{pmatrix} A & I_n \end{pmatrix}$ tiene rango igual al de la matriz A, y es claro que rango$(B) = n$.

Aunque evitaremos el cálculo de la inversa de una matriz, hay veces que debemos hacerlo. Para construir un algoritmo que nos devuelva A^{-1} cuando $A_{n\times n}$ es no singular, recordemos que determinar A^{-1} es equivalente a resolver la ecuación matricial $AX = I$, que es lo mismo que resolver los n sistemas de ecuaciones definidos por

$$A\boldsymbol{x} = I_{*j}, j = 1,2,\ldots,n.$$

En otras palabras, si $X_{*1}, X_{*2}, \ldots, X_{*n}$ son las respectivas soluciones, entonces

$$X = \begin{pmatrix} X_{*1} & X_{*2} & \ldots & X_{*n} \end{pmatrix}$$

resuelve la ecuación $AX = I$ y de aquí $X = A^{-1}$.

Si A es no singular, el método de Gauss-Jordan reduce la matriz ampliada $[A|I_{*j}]$ a $[I|X_{*j}]$, y sabemos que X_{*j} es la única solución de $A\boldsymbol{x} = I_{*j}$. En otras palabras,

$$[A|I_{*j}] \xrightarrow{\text{Gauss-Jordan}} [I|[A^{-1}]_{*j}].$$

Pero mejor que resolver cada sistema $A\boldsymbol{x} = I_{*j}$ de forma independiente, podemos resolverlos simultáneamente aprovechando que todos tienen la misma matriz de coeficientes. En otras palabras, si aplicamos Gauss-Jordan a la matriz ampliada $[A|I_{*1}|I_{*2}|\ldots|I_{*n}]$ obtenemos

$$[A|I_{*1}|I_{*2}|\ldots|I_{*n}] \xrightarrow{\text{Gauss-Jordan}} [I|[A^{-1}]_{*1}|[A^{-1}]_{*2}|\ldots|[A^{-1}]_{*n}],$$

o de manera más compacta

$$[A|I] \xrightarrow{\text{Gauss-Jordan}} [I|A^{-1}].$$

¿Qué ocurre si intentamos invertir una matriz singular con este procedimiento? El resultado anterior nos indica que una matriz singular A no puede ser reducida mediante Gauss-Jordan a la matriz I porque una fila de ceros aparecerá en algún momento en la zona correspondiente a la matriz A. Por ello, no tenemos que saber a priori si la matriz que tenemos es o no singular, pues resultará evidente en el proceso de cálculo.

Cálculo de la inversa

La eliminación de Gauss-Jordan se puede usar para el cálculo de la inversa de una matriz A mediante la reducción

$$[A|I] \xrightarrow{Gauss-Jordan} [I|A^{-1}].$$

La única posibilidad de que este método falle es porque aparezca una fila de ceros en el lado izquierdo de la matriz ampliada, y esto ocurre si y solamente si la matriz A es singular.

Aunque no están incluidos en los ejemplos de esta sección, recordemos que el pivoteo y el escalado son necesarios, y que los efectos del mal condicionamiento se deben considerar cuando calculamos una inversa con datos en coma flotante.

Ejemplo 3.4.1. Calculemos, si existe, la inversa de la matriz

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

Aplicamos el método de Gauss-Jordan para obtener

$$[A|I] = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 2 & -1 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right).$$

Por tanto , la matriz es no singular y

$$A^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

Número de operaciones para calcular la inversa

El cálculo de $A^{-1}_{n\times n}$ mediante Gauss-Jordan aplicado a $[A|I]$ requiere

- n^3 multiplicaciones/divisiones,
- $n^3 - 2n^2 + n$ sumas/restas.

En total, del orden de $2n^3$ flops. Observemos que es tres veces el coste de la eliminación gaussiana, por lo que para resolver un sistema $A\boldsymbol{x} = \boldsymbol{b}$ no es recomendable usar la fórmula $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ (vea también [**?**, p.103]).

A primera vista, podría parecer que la inversión de una matriz es mucho más compleja que la multiplicación. Sin embargo, la multiplicación estándar entre matrices necesita n^3 multiplicaciones y $n^3 - n^2$ sumas, lo que convierte a la inversión y al producto de matrices en operaciones del mismo orden de coste. Como nota final, decir que hay un algoritmo de multiplicación de matrices que baja el coste al orden de $n^{2,8}$.

Propiedades de la inversión de matrices

Para matrices no singulares A y B, se verifica que

- $(A^{-1})^{-1} = A$.
- El producto AB es no singular.
- $(AB)^{-1} = B^{-1}A^{-1}$.
- $(A^{-1})^t = (A^t)^{-1}$ y $(A^{-1})^* = (A^*)^{-1}$.

PRUEBA: La primera es inmediata. La segunda y la tercera se prueban simultáneamente. Sea $X = B^{-1}A^{-1}$. Entonces $(AB)X = I$, y como son matrices cuadradas, tenemos que $X = (AB)^{-1}$. La última propiedad tiene un tratamiento similar. Sea $X = (A^{-1})^t$, que sabemos que existe (observemos que todavía no podemos garantizar el carácter no singular de A^t). Entonces

$$A^t X = A^t (A^{-1})^t = (A^{-1}A)^t = I^t = I,$$

de donde A^t es no singular y $(A^t)^{-1} = (A^{-1})^t$. La prueba de la segunda parte es similar. □

Fórmula de Sherman-Morrison

Si $A_{n\times n}$ es una matriz no singular, y $\boldsymbol{c}, \boldsymbol{d}$ son vectores columna $n\times 1$ tales que $1 + \boldsymbol{d}^t A^{-1}\boldsymbol{c} \neq 0$, entonces la suma $A + \boldsymbol{c}\boldsymbol{d}^t$ es no singular, y

$$(A + \boldsymbol{c}\boldsymbol{d}^t)^{-1} = A^{-1} - \frac{A^{-1}\boldsymbol{c}\boldsymbol{d}^t A^{-1}}{1 + \boldsymbol{d}^t A^{-1}\boldsymbol{c}}.$$

La utilidad de la fórmula de Sherman-Morrison se aprecia cuando, habiendo calculado A^{-1}, necesitamos obtener la inversa de la matriz resultado de cambiar un elemento de la matriz A. No es necesario empezar desde el principio para calcular la nueva inversa. Supongamos que cambiamos a_{ij} por $a_{ij}+\alpha$. Sean $\boldsymbol{c}=\boldsymbol{e}_i$ y $\boldsymbol{d}=\alpha\boldsymbol{e}_j$, donde $\boldsymbol{e}_i$ y $\boldsymbol{e}_j$ son los vectores correspondientes a las columnas i-ésima y j-ésima de la matriz identidad, respectivamente. La matriz $\boldsymbol{c}\boldsymbol{d}^t$ tiene α en la posición (i,j) y cero en el resto, por lo que

$$B = A + \boldsymbol{c}\boldsymbol{d}^t = A + \alpha \boldsymbol{e}_i \boldsymbol{e}_j^t$$

es la matriz actualizada. Según la fórmula de Sherman-Morrison,

$$\begin{aligned} B^{-1} = (A+\alpha\boldsymbol{e}_i\boldsymbol{e}_j^t)^{-1} &= A^{-1} - \alpha\frac{A^{-1}\boldsymbol{e}_i\boldsymbol{e}_j^t A^{-1}}{1+\alpha\boldsymbol{e}_j^t A^{-1}\boldsymbol{e}_i} \\ &= A^{-1} - \alpha\frac{[A^{-1}]_{*i}[A^{-1}]_{j*}}{1+\alpha[A^{-1}]_{ji}}. \end{aligned}$$

Esto muestra cómo cambia A^{-1} cuando a_{ij} es modificado, y da un algoritmo útil para actualizar A^{-1}.

Matrices idempotentes

Una matriz cuadrada A es **idempotente** si $A^2 = A$.

Ejemplo 3.4.2. La matriz identidad es una matriz idempotente. La matriz

$$A_1 = \begin{bmatrix} 2/3 & -1/3 \\ -2/3 & 1/3 \end{bmatrix}$$

es idempotente. Si A es idempotente y P es una matriz no singular, entonces la matriz $B = P^{-1}AP$ es idempotente.

Existen matrices idempotentes que se utilizan con mucha frecuencia en estadística para representar operaciones muy usuales. Sea $\mathbf{1}$ el vector de n componentes con todas sus entradas iguales a 1 y definimos la matriz

$$C = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t.$$

La matriz C es idempotente:

$$\begin{aligned} C^2 &= I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{1}{n^2}\mathbf{1}\underbrace{\mathbf{1}^t\mathbf{1}}_{\text{escalar}}\mathbf{1}^t \\ &= I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t - \frac{1}{n}\mathbf{1}\mathbf{1}^t + \frac{n}{n^2}\mathbf{1}\mathbf{1}^t \\ &= C. \end{aligned}$$

Se verifican las siguientes propiedades:

1. Sea $\boldsymbol{a}$ un vector y $\bar{a}$ la media de sus componentes. Entonces

$$C\boldsymbol{a} = \begin{pmatrix} a_1 - \bar{a} \\ a_2 - \bar{a} \\ \vdots \\ a_n - \bar{a} \end{pmatrix},$$

   es decir, se obtiene el vector centrado alrededor de la media de sus componentes.

2. Si A es una matriz de orden $n \times p$, entonces

$$CA = \begin{pmatrix} a_{11} - \bar{a}_1 & \dots & a_{1p} - \bar{a}_p \\ a_{21} - \bar{a}_1 & \dots & a_{2p} - \bar{a}_p \\ \vdots & & \vdots \\ a_{n1} - \bar{a}_1 & \dots & a_{np} - \bar{a}_p \end{pmatrix},$$

   donde $\bar{a}_1, \dots, \bar{a}_p$ son las medias de las columnas respectivas de A.

3. $C\mathbf{1} = \mathbf{0}$.
4. $\mathbf{1}^t C = \mathbf{0}^t$.
5. $\mathbf{1}\mathbf{1}^t C = C\mathbf{1}\mathbf{1}^t = 0_{n\times n}$.
6. $\sum_{i=1}^n (x_i - \bar{x})^2 = \boldsymbol{x}^t C \boldsymbol{x}$.

Propiedades elementales de las matrices idempotentes

Sean A y B matrices idempotentes de orden n. Entonces

1. AB es una matriz idempotente si, además, $AB = BA$.
2. $I - A$ es idempotente.
3. $A(I - A) = (I - A)A = 0_{n\times n}$.

3.5. Matrices elementales y equivalencia

Vamos a ver que las operaciones elementales que usamos para la eliminación gaussiana pueden interpretarse como productos por ciertas matrices de estructura muy sencilla.

Matrices elementales

Las ***matrices elementales*** son las matrices de la forma $I - \boldsymbol{u}\boldsymbol{v}^t$, donde $\boldsymbol{u}$ y $\boldsymbol{v}$ son vectores columna $n \times 1$ tales que $\boldsymbol{v}^t\boldsymbol{u} \neq 1$.

Estas matrices tienen inversa; en concreto,

$$(I - \boldsymbol{u}\boldsymbol{v}^t)^{-1} = I + \frac{\boldsymbol{u}\boldsymbol{v}^t}{1 - \boldsymbol{v}^t\boldsymbol{u}},$$

que a su vez son matrices elementales.

Una matriz elemental de tipo I es de la forma $E_1 = I - \boldsymbol{u}\boldsymbol{u}^t$, con $\boldsymbol{u} = \boldsymbol{e}_i - \boldsymbol{e}_j$. Esta matriz se obtiene a partir de la matriz identidad intercambiando las filas i y j. Por ejemplo,

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I - (\boldsymbol{e}_1 - \boldsymbol{e}_2)(\boldsymbol{e}_1 - \boldsymbol{e}_2)^t$$

es una matriz elemental de tipo I, resultado de intercambiar las filas 1 y 2 de I_3. Se las llama *matrices de permutación*, y se representan por P_{ij}, con i y j las filas implicadas.

Una matriz elemental de tipo II es de la forma $E_2 = I - (1-\alpha)\boldsymbol{e}_i\boldsymbol{e}_i^t$. Esta matriz se obtiene a partir de la matriz identidad multiplicando la i-ésima fila por α. Por ejemplo,

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = I - (1-\alpha)\boldsymbol{e}_2\boldsymbol{e}_2^t$$

es una matriz elemental de tipo II, resultado de multiplicar la segunda fila de la matriz I_3 por α. La notaremos por $T_i(\alpha)$.

Una matriz elemental de tipo III es de la forma $E_3 = I + \alpha\boldsymbol{e}_j\boldsymbol{e}_j^i$, $i \neq j$. Esta matriz se obtiene a partir de la matriz identidad y poniendo en la posición (j,i) el valor α. Por ejemplo,

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix} = I + \alpha\boldsymbol{e}_3\boldsymbol{e}_1^t.$$

La notaremos por $T_{ij}(\alpha)$.

Propiedades de las matrices elementales

- Cuando una matriz elemental de tipo I, II o III multiplica *a la izquierda* a una matriz, produce la correspondiente transformación elemental *por filas*.
- Cuando una matriz elemental de tipo I, II o III multiplica *a la derecha* a una matriz, produce la correspondiente transformación elemental *por columnas*.

PRUEBA: Las matrices elementales de tipos I y II se comprueban fácilmente. Veamos las de tipo III. Sea $E_3 = I + \alpha \boldsymbol{e}_j \boldsymbol{e}_i^t$. Entonces

$$(I + \alpha \boldsymbol{e}_j \boldsymbol{e}_i^t)A = A + \alpha \boldsymbol{e}_j A_{i*} = A + \alpha \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \leftarrow \text{fila } j \ .$$

Esta es exactamente la matriz producida por una operación de tipo III sobre las filas i y j: la fila j-ésima de A se cambia por ella más la fila i-ésima multiplicada por α.

Cuando multiplicamos a la derecha, nos queda

$$A(I + \alpha \boldsymbol{e}_j \boldsymbol{e}_i^t) = A + \alpha A_{*j} \boldsymbol{e}_i^t = A + \alpha \begin{pmatrix} & & \text{col.}\,i & & \\ & & \downarrow & & \\ 0 & \dots & a_{1j} & \dots & 0 \\ 0 & \dots & a_{2j} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & a_{mj} & \dots & 0 \end{pmatrix}.$$

Se ha cambiado la columna i-ésima de A por ella más la columna j-ésima multiplicada por α. □

Aunque no hemos hablado de dimensiones, lo anterior es válido para matrices generales de orden $m \times n$.

Ejemplo 3.5.1. Consideremos la sucesión de operaciones para reducir

$$A=\begin{pmatrix}1&2&4\\2&4&8\\3&6&13\end{pmatrix}$$

a su forma escalonada reducida por filas E_A.

$$A=\begin{pmatrix}1&2&4\\2&4&8\\3&6&13\end{pmatrix}\xrightarrow[F_3-3F_1]{F_2-2F_1}\begin{pmatrix}1&2&4\\0&0&0\\0&0&1\end{pmatrix}$$

$$\xrightarrow{\text{Cambia } F_2 \text{ y } F_3}\begin{pmatrix}1&2&4\\0&0&1\\0&0&0\end{pmatrix}\xrightarrow{F_1-4F_2}\begin{pmatrix}1&2&0\\0&0&1\\0&0&0\end{pmatrix}=E_A$$

La reducción se puede ver como una sucesión de multiplicaciones a izquierda por la matrices elementales correspondientes.

$$\begin{pmatrix}1&-4&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\-3&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\-2&1&0\\0&0&1\end{pmatrix}A=E_A.$$

Inversas de matrices elementales

- La inversa de una matriz elemental de tipo I es ella misma: $P_{ij}^{-1}=P_{ij}$.
- La inversa de una matriz elemental de tipo II es una matriz elemental de tipo II: $T_i(\alpha)^{-1}=T_i(\alpha^{-1})$.
- La inversa de una matriz elemental de tipo II es una matriz elemental de tipo III: $T_{ij}(\alpha)^{-1}=T_{ij}(-\alpha)$.

PRUEBA: Es algo inmediato a partir del efecto que tienen estas transformaciones elementales sobre una matriz. Sin embargo, vamos a hacer la prueba basándonos en la definición original. Recordemos que

$$(I-\boldsymbol{u}\boldsymbol{v}^t)^{-1}=I+\frac{\boldsymbol{u}\boldsymbol{v}^t}{1-\boldsymbol{v}^t\boldsymbol{u}},\ \text{si } 1-\boldsymbol{v}^t\boldsymbol{u}\neq 0.$$

- Por definición, $P_{ij} = I - (e_i - e_j)(e_i - e_j)^t$. Entonces

$$\begin{aligned} P_{ij}^{-1} &= I + \frac{(e_i - e_j)(e_i - e_j)^t}{1 - (e_i - e_j)^t(e_i - e_j)} \\ &= I - (e_i - e_j)(e_i - e_j)^t, \text{ pues } 1 - (e_i - e_j)^t(e_i - e_j) = 1 - 2 = -1, \\ &= P_{ij}. \end{aligned}$$

- Ahora tenemos que $T_i(\alpha) = I - (1-\alpha)e_i e_i^t$, para $\alpha \neq 0$. Entonces

$$\begin{aligned} T_i(\alpha)^{-1} &= I + \frac{(1-\alpha)e_i e_i^t}{1 - e_i^t(1-\alpha)e_i} \\ &= I + (1-\alpha)e_i e_i^t \frac{1}{1-1+\alpha} \\ &= I - (1 - \frac{1}{\alpha})e_i e_i^t = T_i(\alpha^{-1}). \end{aligned}$$

- Para las de tipo III sabemos que $T_{ij}(\alpha) = I + \alpha e_j e_i^t$, con $i \neq j$. Entonces

$$\begin{aligned} T_{ij}(\alpha)^{-1} &= I - \frac{\alpha e_j e_i^t}{1 + e_i^t \alpha e_j} \\ &= I - \alpha e_j e_i^t, \text{ porque } e_i^t e_j = 0 \text{ para } i \neq j, \\ &= T_{ij}(-\alpha). \end{aligned}$$

□

Producto de matrices elementales

Una matriz A es no singular si y solamente si A es el producto de matrices elementales de tipos I, II, o III.

PRUEBA: Si A es no singular, el método de Gauss-Jordan reduce A a la matriz I mediante operaciones por fila. Si $G_1, G_2, \ldots, G_k$ son las correspondientes matrices elementales, entonces

$$G_k \ldots G_2 G_1 A = I, \text{ o bien } A = G_1^{-1} G_2^{-1} \ldots G_k^{-1}.$$

Como la inversa de una matriz elemental es una matriz elemental, esto prueba que A se puede expresar como producto de matrices elementales.

Recíprocamente, si $A = E_1 E_2 \ldots E_k$ es un producto de matrices elementales, entonces A es no singular, pues es el producto de matrices no singulares. □

Ejemplo 3.5.2. Expresemos

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

como producto de matrices elementales. Mediante la reducción a su forma escalonada reducida por filas, comprobaremos que A es no singular, y la expresaremos como dicho producto. En efecto,

$$\begin{aligned} A &\overset{-\frac{1}{2}F_1}{\longrightarrow} \begin{bmatrix} 1 & -3/2 \\ 1 & 0 \end{bmatrix} \\ &\overset{F_2-F_1}{\longrightarrow} \begin{bmatrix} 1 & -3/2 \\ 0 & 3/2 \end{bmatrix} \\ &\overset{\frac{2}{3}F_2}{\longrightarrow} \begin{bmatrix} 1 & -3/2 \\ 0 & 1 \end{bmatrix} \\ &\overset{F_1+\frac{3}{2}F_2}{\longrightarrow} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Entonces

$$\begin{pmatrix} 1 & \frac{3}{2} \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \frac{2}{3} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} A = I_2,$$

de donde

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

Equivalencia de matrices

Cuando una matriz B se obtiene de una matriz A mediante operaciones elementales de filas y columnas, escribiremos $A \sim B$ y diremos que A y B son ***matrices equivalentes***. Otra forma de expresarlo es que

$A \sim B \Leftrightarrow B = PAQ$ para matrices no singulares P y Q.

Ejemplo 3.5.3. Las matrices

$$A = \begin{pmatrix} -1 & 4 & 0 & 4 \\ -3 & 2 & -4 & 2 \\ -2 & -2 & -2 & -4 \end{pmatrix} \text{ y } B = \begin{pmatrix} 86 & -15 & -16 & -33 \\ -6 & 51 & -8 & -43 \\ 28 & 33 & -12 & -43 \end{pmatrix}$$

son equivalentes porque $PAQ = B$ para las matrices no singulares

$$P = \begin{pmatrix} 2 & 3 & 2 \\ 2 & -3 & -2 \\ 2 & -1 & -1 \end{pmatrix}, Q = \begin{pmatrix} 0 & 3 & 2 & 3 \\ 2 & 3 & 1 & -3 \\ -3 & 0 & -1 & -3 \\ 3 & 0 & -2 & -1 \end{pmatrix}$$

Equivalencia por filas y columnas

- Decimos que dos matrices A, B de la misma dimensión $m \times n$ son ***equivalentes por filas*** si existe una matriz P de orden m no singular tal que $PA = B$. Lo notamos como $A \overset{\text{f}}{\sim} B$.
- Decimos que dos matrices A, B de la misma dimensión $m \times n$ son ***equivalentes por columnas*** si existe una matriz Q de orden n no singular tal que $AQ = B$. Lo notamos como $A \overset{\text{c}}{\sim} B$.

Estas relaciones son de equivalencia.

Ejemplo 3.5.4. Toda matriz A es equivalente por filas a su forma escalonada reducida por filas E_A. La matriz

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ es equivalente por columnas a la matriz}$$

$$B \cdot \begin{pmatrix} 0 & 3 & 2 & 3 \\ 2 & 3 & 1 & -3 \\ -3 & 0 & -1 & -3 \\ 3 & 0 & -2 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 3 & 2 & 3 \\ 2 & 3 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Relaciones entre filas y columnas

- Si $A \overset{\text{f}}{\sim} B$, entonces las relaciones que existen entre las columnas de A también se tienen entre las columnas de B. Esto es,

$$B_{*k} = \sum_{j=1}^{n} \alpha_j B_{*j} \Leftrightarrow A_{*k} = \sum_{j=1}^{n} \alpha_j A_{*j}.$$

- Si $A \overset{\text{c}}{\sim} B$, entonces las relaciones que existen entre las filas de A también se tienen entre las filas de B.

En particular, las relaciones entre columnas en A y E_A deben ser las mismas, por lo que las columnas no básicas de A son combinación lineal de las básicas, tal como describimos en su momento.

PRUEBA: Si $A \overset{\text{f}}{\sim} B$, entonces $PA = B$, para una matriz P no singular. Tal como vimos en el producto de matrices,

$$B_{*j} = (PA)_{*j} = PA_{*j}.$$

Por tanto, si $A_{*k} = \sum_{j=1}^{n} \alpha_j A_{*j}$, la multiplicación a la izquierda por P produce $B_{*k} = \sum_{j=1}^{n} \alpha_j B_{*j}$. El recíproco se obtiene con P^{-1}.

El resultado para las columnas se deduce inmediatamente a partir de lo anterior aplicado a A^t y B^t. □

La forma escalonada reducida por filas E_A es lo más lejos que podemos llegar mediante transformaciones por filas. Sin embargo, si permitimos además el uso de transformaciones por columnas, la reducción es mucho mayor.

Forma normal de rango

Si A es una matriz de orden $m \times n$ y $\text{rango}(A) = r$, entonces

$$A \sim N_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

N_r se denomina ***forma normal de rango*** de A.

PRUEBA: Como $A \overset{\text{f}}{\sim} E_A$, existe una matriz no singular P tal que $PA = E_A$. Si rango$(A) = r$, entonces las columnas básicas de E_A son las r columnas unitarias. Mediante intercambio de columnas aplicados a E_A, podemos poner estas r columnas en la parte superior izquierda. Si Q_1 es el producto de las matrices elementales que hacen estos intercambios, entonces

$$PAQ_1 = E_A Q_1 = \begin{pmatrix} I_r & J \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Ahora multiplicamos ambos lados de esta ecuación por la matriz no singular

$$Q_2 = \begin{pmatrix} I_r & -J \\ \mathbf{0} & I \end{pmatrix},$$

y nos queda

$$PAQ_1Q_2 = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Entonces $A \sim N_r$. □

Ejemplo 3.5.5. Veamos que

$$\text{rango}\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} = \text{rango}(A) + \text{rango}(B).$$

Si rango$(A) = r$ y rango$(B) = s$, entonces $A \sim N_r$ y $B \sim N_s$, y

$$\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} \sim \begin{pmatrix} N_r & \mathbf{0} \\ \mathbf{0} & N_s \end{pmatrix},$$

de donde

$$\text{rango}\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} = r + s.$$

Dadas matrices A y B, ¿cómo decidimos si $A \sim B$, $A \overset{\text{f}}{\sim} B$ o $A \overset{\text{c}}{\sim} B$?

Test de equivalencia

Sean A y B matrices de orden $m \times n$. Entonces

- $A \sim B$ si y solamente si rango(A) = rango(B).
- $A \overset{\text{f}}{\sim} B$ si y solamente si $E_A = E_B$.
- $A \overset{\text{c}}{\sim} B$ si y solamente si $E_{A^t} = E_{B^t}$.

En consecuencia, el producto por matrices no singulares no altera el rango.

PRUEBA: Si rango(A) = rango(B), entonces $A \sim N_r$ y $B \sim N_r$, de donde $A \sim N_r \sim B$. Recíprocamente, si $A \sim B$, y rango$(A) = r$, rango$(B) = s$, tenemos que $A \sim N_r$ y $B \sim N_s$, por lo que $N_r \sim N_s$. Existen P y Q no singulares tales que $PN_rQ^{-1} = N_s$, o bien que $PN_r = N_sQ$. La forma escalonada reducida por filas de PN_r es N_r, por lo que tiene rango r y entonces rango$(N_sQ) = r$. Por otro lado, como $N_s \overset{\text{c}}{\sim} N_sQ$, las relaciones entre las filas de N_s y N_sQ son las mismas, lo que implica que N_sQ tiene $m-s$ filas nulas al final y s filas no nulas al principio. Por la definición de rango, esto quiere decir que rango$(N_sQ) \leq s$ y tenemos que $r \leq s$. De forma análoga, a partir de $N_rQ^{-1} = P^{-1}N_s$, llegamos a $r \geq s$ y tenemos el resultado.

Supongamos ahora que $A \overset{\text{f}}{\sim} B$. Como $B \overset{\text{f}}{\sim} E_B$, entonces $A \overset{\text{f}}{\sim} E_B$, y dado que la forma escalonada reducida por filas es única, se sigue que $E_B = E_A$. Recíprocamente, si $E_A = E_B$, entonces

$$A \overset{\text{f}}{\sim} E_A = E_B \overset{\text{f}}{\sim} B.$$

Para las columnas, basta considerar que

$$\begin{aligned} A \overset{\text{c}}{\sim} B &\Leftrightarrow AQ = B \Leftrightarrow (AQ)^t = B^t \\ &\Leftrightarrow Q^tA^t = B^t \Leftrightarrow A^t \overset{\text{f}}{\sim} B^t. \end{aligned}$$

□

Rango y trasposición

$$\text{rango}(A) = \text{rango}(A^t) \text{ y } \text{rango}(A) = \text{rango}(A^*).$$

PRUEBA: Sea $\text{rango}(A) = r$, y sean P y Q matrices no singulares tales que

$$PAQ = N_r = \begin{pmatrix} I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}.$$

Entonces $N_r^t = Q^t A^t P^t$. Como Q^t y P^t son no singulares, se sigue que $A^t \sim N_r^t$, y entonces

$$\text{rango}(A^t) = \text{rango}(N_r^t) = \text{rango}\begin{pmatrix} I_r & \mathbf{0}_{r\times(m-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(m-r)} \end{pmatrix} = r = \text{rango}(A).$$

De forma análoga, $N_r^* = Q^* A^* P^*$, donde Q^*, P^* son matrices no singulares. Como

$$N_r^* = \text{rango}\begin{pmatrix} I_r & \mathbf{0}_{r\times(m-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(m-r)} \end{pmatrix},$$

se tiene que $\text{rango}(N^*) = r$, y como $\text{rango}(A^*) = \text{rango}(N_r^*)$ por equivalencia de matrices, tenemos que $\text{rango}(A^*) = r = \text{rango}(A)$. □

3.6. Aplicaciones del álgebra matricial

Los sistemas dinámicos discretos son una herramienta extremadamente útil en una amplia variedad de campos.

Sistema dinámico lineal discreto

Un ***sistema lineal dinámico discreto*** es una sucesión de vectores $\boldsymbol{x}^{(k)}, k = 0, 1, \ldots$, llamados estados, que se definen por un vector inicial $\boldsymbol{x}^{(0)}$ y una regla

$$\boldsymbol{x}^{(k+1)} = A\boldsymbol{x}^{(k)}, k = 0, 1, \ldots,$$

donde A es una matriz cuadrada fija, llamada matriz de transición del sistema.

Ejemplo 3.6.1. Supongamos que dos compañías de pasta de dientes compiten por los clientes de un mercado fijo, en el que cada consumidor usa la marca A o la marca B. Supongamos que un análisis de mercado muestra que los hábitos de mercado siguen la siguiente tendencia: cada 3 meses, el 30 % de los usuarios de A se cambian a B, mientras que el resto permanece en A. Además, el 40 % de los usuarios de B cambiarán a A, y el resto de usuarios de B serán fieles a la marca. Si suponemos que este patrón no cambia de trimestre en trimestre, tenemos un ejemplo de una cadena de Markov. Vamos a expresar este sistema en el lenguaje matricial.

Sean a_k y b_k las fracciones de clientes que usan las marcas A y B en el trimestre k-ésimo. Las condiciones del enunciado nos dicen que

$$\begin{aligned} a_{k+1} &= 0{,}7a_k + 0{,}4b_k, \\ b_{k+1} &= 0{,}3a_k + 0{,}6b_k. \end{aligned}$$

En forma matricial queda

$$\boldsymbol{x}^{(k+1)} = A\boldsymbol{x}^{(k)}, \text{ donde } \boldsymbol{x}^{(k)} = \begin{pmatrix} a_k \\ b_k \end{pmatrix}, A = \begin{pmatrix} 0{,}7 & 0{,}4 \\ 0{,}3 & 0{,}6 \end{pmatrix}.$$

Los vectores de estado $\boldsymbol{x}^{(k)}$ tienen componentes no negativas, y suman 1. Además, la matriz A verifica que tiene entradas no negativas, y la suma de cada una de sus columnas es 1, es decir, sus columnas son vectores de probabilidad.

Cadena de Markov

Una ***cadena de Markov*** es un sistema dinámico discreto cuyo vector inicial $\boldsymbol{x}^{(0)}$ es un vector de probabilidad y su matriz de transición es estocástica, esto es, cada columna de A es un vector de probabilidad.

Volvamos al ejemplo anterior. Tenemos que

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= A\boldsymbol{x}^{(k)} \\ &= A(A\boldsymbol{x}^{(k-1)}) \\ &\vdots \\ &= A^{k+1}\boldsymbol{x}^{(0)}. \end{aligned}$$

En realidad, esto es válido para cualquier sistema dinámico. Vamos a analizar una situación especial para la cadena de Markov de nuestro ejemplo. Supongamos que, inicialmente, la marca A tiene todos los clientes, y la marca B

está entrando en el mercado. Veamos que ocurre a largo plazo. Con estas condiciones, $\boldsymbol{x}^{(0)} = (1,0)^t$. Entonces

$$\boldsymbol{x}^{(2)} = A^2\boldsymbol{x}^{(0)} = \begin{pmatrix} 0{,}61 & 0{,}52 \\ 0{,}39 & 0{,}48 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0{,}61 \\ 0{,}39 \end{pmatrix}.$$

Si ampliamos el periodo de cálculo,

$$\boldsymbol{x}^{(20)} = A^{20}\boldsymbol{x}^{(0)} = \begin{pmatrix} 0{,}57 \\ 0{,}43 \end{pmatrix}.$$

Por tanto, tras 20 trimestres, la marca A tendrá el 57 % del mercado, y la marca B el 43 %.

Veamos que ocurre si el escenario de partida es completamente diferente. Por ejemplo, que la marca A no tiene clientes y la marca B los tiene todos. En este caso, $\boldsymbol{x}^{(0)} = (0,1)^t$, y

$$\boldsymbol{x}^{(20)} = A^{20}\boldsymbol{x}^{(0)} = \begin{pmatrix} 0{,}57 \\ 0{,}43 \end{pmatrix}.$$

Hemos obtenido la misma respuesta. No es una coincidencia, y veremos más adelante qué significa esto.

Capítulo 4

Espacios vectoriales

4.1. Espacios y subespacios

Después de que la teoría de matrices fuera establecida hacia el final del siglo XIX, se observó que muchas entidades matemáticas que eran consideradas diferentes a las matrices eran, en realidad, bastante similares. Por ejemplo, objetos como las direcciones en el plano de $\mathbb{R}^2$ o en el espacio $\mathbb{R}^3$, polinomios, funciones continuas, y funciones diferenciables, satisfacen las mismas propiedades aditivas y de multiplicación por un escalar que se tienen para las matrices. La idea de abstracción que permitiera un tratamiento unificado llevó finalmente a la definición axiomática de espacio vectorial por Peano (*Calcolo Geometrico*, 1888).

Un espacio vectorial agrupa a cuatro objetos: dos conjuntos V y $\mathbb{K}$, y dos operaciones algebraicas llamadas adición de vectores y producto por un escalar.

- V es un conjunto no vacío de objetos que llamaremos ***vectores***. Aunque V puede ser bastante general, habitualmente consideraremos V como un conjunto de n-uplas o un conjunto de matrices.
- $\mathbb{K}$ es un cuerpo de escalares. Para nosotros será el conjunto de números reales $\mathbb{R}$ o el de números complejos $\mathbb{C}$.
- La adición de vectores, notada por $\boldsymbol{x}+\boldsymbol{y}$, es una operación entre elementos de V.
- La multiplicación por un escalar, notada por $\alpha\boldsymbol{x}$, es una operación entre elementos de $\mathbb{K}$ y V.

La definición formal de espacio vectorial establece cómo estos cuatro objetos se relacionan entre sí. En esencia, los requisitos son que la suma de vectores

y el producto por escalares tengan las mismas propiedades que vimos para matrices.

Definición de espacio vectorial

El conjunto V se denomina ***espacio vectorial sobre*** $\mathbb{K}$ si la adición de vectores y la multiplicación por escalares satisfacen las siguientes propiedades:

1. $\boldsymbol{x}+\boldsymbol{y} \in V$ para todo $\boldsymbol{x},\boldsymbol{y} \in V$.
2. $(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z} = \boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$ para todo $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in V$.
3. $\boldsymbol{x}+\boldsymbol{y} = \boldsymbol{y}+\boldsymbol{x}$ para todo $\boldsymbol{x},\boldsymbol{y} \in V$.
4. Existe un elemento $\mathbf{0} \in V$ tal que $\boldsymbol{x}+\mathbf{0} = \boldsymbol{x}$ para todo $\boldsymbol{x} \in V$.
5. Para cada $\boldsymbol{x} \in V$ existe un elemento $-\boldsymbol{x} \in V$ tal que $\boldsymbol{x}+(-\boldsymbol{x}) = \mathbf{0}$.
6. $\alpha\boldsymbol{x} \in V$ para todo $\alpha \in \mathbb{K}$ y $\boldsymbol{x} \in V$.
7. $(\alpha\beta)\boldsymbol{x} = \alpha(\beta\boldsymbol{x})$ para todo $\alpha,\beta \in \mathbb{K}$ y $\boldsymbol{x} \in V$.
8. $\alpha(\boldsymbol{x}+\boldsymbol{y}) = \alpha\boldsymbol{x}+\alpha\boldsymbol{y}$ para todo $\alpha \in \mathbb{K}$ y $\boldsymbol{x},\boldsymbol{y} \in V$.
9. $(\alpha+\beta)\boldsymbol{x} = \alpha\boldsymbol{x}+\beta\boldsymbol{x}$ para todo $\alpha,\beta \in \mathbb{K}$ y $\boldsymbol{x} \in V$.
10. $1\boldsymbol{x} = \boldsymbol{x}$ para todo $\boldsymbol{x} \in V$.

De las propiedades anteriores se deduce fácilmente que el elemento $\mathbf{0}$ de un espacio vectorial es único, y se le denomina elemento neutro. Si $\mathbf{0}$ y $\mathbf{0}'$ verifican la condición, entonces

$$\mathbf{0}+\mathbf{0}' = \mathbf{0}, \text{ por ser } \mathbf{0}' \text{ elemento neutro,}$$
$$\mathbf{0}'+\mathbf{0} = \mathbf{0}', \text{ por ser } \mathbf{0} \text{ elemento neutro,}$$

de donde $\mathbf{0} = \mathbf{0}'$.

Otra propiedad es la de cancelación: si $\boldsymbol{x} = \boldsymbol{y}+\boldsymbol{z}$, entonces $\boldsymbol{x}+(-\boldsymbol{z}) = \boldsymbol{y}$, pues basta sumar el opuesto del vector $\boldsymbol{z}$ a ambos lados de la igualdad.

Ejemplo 4.1.1. Como las propiedades anteriores no son más que las mismas que teníamos para matrices, es inmediato que $\mathbb{R}^{m\times n}$ es un espacio vectorial sobre $\mathbb{R}$, y $\mathbb{C}^{m\times n}$ es un espacio vectorial sobre $\mathbb{C}$.

Ejemplo 4.1.2. El ***espacio real coordenado por filas***

$$\mathbb{R}^{1\times n} = \{\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}, x_i \in \mathbb{R}\},$$

y el ***espacio real coordenado por columnas***

$$\mathbb{R}^{n\times 1} = \{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, x_i \in \mathbb{R}\}$$

son casos particulares del ejemplo anterior, pero centrarán nuestra atención. En el contexto de espacios vectoriales, es indiferente si un vector coordenado se pone como una fila o una columna. Cuando la distinción entre fila o columna sea irrelevante, usaremos el símbolo común $\mathbb{R}^n$ para designar el espacio coordenado. Como elección de partida, sin embargo, pensaremos en los vectores de $\mathbb{R}^n$ como ***vectores columna***. Se tiene lo análogo para los espacios coordenados complejos.

Ejemplo 4.1.3. Enumeramos algunos ejemplos clásicos sobre espacios vectoriales. Es habitual el referirse a un espacio vectorial V sobre un cuerpo $\mathbb{K}$ como un $\mathbb{K}$-espacio vectorial.

1. $\mathbb{C}$ es un $\mathbb{R}$-espacio vectorial.

2. $\mathbb{C}$ es un $\mathbb{C}$-espacio vectorial.

3. $\mathbb{R}$ es un $\mathbb{Q}$-espacio vectorial.

4. $\mathbb{Q}$ **no** es un $\mathbb{R}$-espacio vectorial.

5. El conjunto $\mathbb{R}[X]$ de polinomios con coeficientes reales es un $\mathbb{R}$-espacio vectorial.

6. El conjunto $\mathbb{R}_m[X]$ de polinomios con coeficientes reales y de grado menor que m es un $\mathbb{R}$-espacio vectorial.

7. El conjunto $C^k([a,b])$ de funciones $f : [a,b] \to \mathbb{R}$ que son k veces diferenciables es un $\mathbb{R}$-espacio vectorial.

8. El conjunto $\mathbb{Q}[\sqrt{3}] = \{a + b\sqrt{3} \mid a, b \in \mathbb{Q}\}$ es un $\mathbb{Q}$-espacio vectorial.

Nota 4.1.4. Hay unas relaciones sencillas que se deducen de la definición de espacio vectorial. Por ejemplo, el producto $0 \cdot \boldsymbol{x}$ del elemento neutro de la suma del cuerpo por un vector es igual a $\mathbf{0}$, el elemento neutro de la suma de V. En efecto,

$$\boldsymbol{x} = 1 \cdot \boldsymbol{x} = (1+0) \cdot \boldsymbol{x} = 1 \cdot \boldsymbol{x} + 0 \cdot \boldsymbol{x}, \text{ de donde } 0 \cdot \boldsymbol{x} = \mathbf{0},$$

al eliminar $\boldsymbol{x}$ en ambos lados de la igualdad. Otra propiedad relaciona el elemento (-1) con el opuesto de un vector: $(-1) \cdot \boldsymbol{x} = -\boldsymbol{x}$. Por lo anterior,

$$\mathbf{0} = 0 \cdot \boldsymbol{x} = (1 + (-1)) \cdot \boldsymbol{x} = 1 \cdot \boldsymbol{x} + (-1) \cdot \boldsymbol{x}.$$

Entonces $(-1) \cdot \boldsymbol{x}$ es el opuesto de $\boldsymbol{x}$ con respecto a la suma en V.

Subespacios

Sea W un subconjunto no vacío de un espacio vectorial V sobre $\mathbb{K}$. Si W es un espacio vectorial sobre $\mathbb{K}$ con las mismas operaciones de suma vectorial y producto por un escalar, decimos que W es un ***subespacio vectorial*** o ***variedad lineal*** de V.

No es necesario verificar todas las condiciones para determinar si un subconjunto W de V es subespacio. Basta con comprobar que las operaciones son internas, esto es,

1. $\boldsymbol{x}, \boldsymbol{y} \in W \Rightarrow \boldsymbol{x} + \boldsymbol{y} \in W$
2. $\boldsymbol{x} \in W \Rightarrow \alpha \boldsymbol{x} \in W$ para todo $\alpha \in \mathbb{K}$.

PRUEBA: Si W es un subconjunto de V, entonces W hereda todas las propiedades de V, excepto la existencia de elemento neutro y elemento opuesto en W. Sin embargo, $(-\boldsymbol{x}) = (-1)\boldsymbol{x} \in W$ para todo elemento $\boldsymbol{x} \in W$. Además, $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0} \in W$, y tenemos que W cumple todas las propiedades. □

Una condición equivalente a la anterior es que si $\boldsymbol{x}, \boldsymbol{y} \in W$, y $\alpha, \beta \in \mathbb{K}$, se tiene que verificar que $\alpha \boldsymbol{x} + \beta \boldsymbol{y} \in W$.

Ejemplo 4.1.5. Dado un espacio vectorial V, el conjunto $Z = \{\mathbf{0}\}$ es un subespacio vectorial, denominado ***subespacio trivial***.

Ejemplo 4.1.6. Para un conjunto de vectores $\mathscr{L} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$, una ***combinación lineal*** de estos vectores es una expresión de la forma

$$\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_r \boldsymbol{v}_r, \text{ donde } \alpha_1, \ldots, \alpha_r \in \mathbb{K}.$$

Toda combinación lineal de vectores es un elemento del espacio vectorial. El conjunto de todas las posibles combinaciones lineales lo notaremos por

$$\langle \mathscr{L} \rangle = \langle \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r \rangle = \{\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_r \boldsymbol{v}_r \mid \alpha_i \in \mathbb{K}\}.$$

Observemos que $\langle \mathscr{L} \rangle$ es un subespacio de V, que llamaremos el ***subespacio generado*** por $\mathscr{L}$. En efecto, sean $\boldsymbol{v}, \boldsymbol{w} \in \langle \mathscr{L} \rangle$ y escalares $\alpha, \beta \in \mathbb{K}$. Debemos probar que $\alpha \boldsymbol{v} + \beta \boldsymbol{w} \in \langle \mathscr{L} \rangle$. Por hipótesis, podemos expresar los vectores $\boldsymbol{v}$ y $\boldsymbol{w}$ como

$$\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_r \boldsymbol{v}_r, \boldsymbol{w} = \beta_1 \boldsymbol{v}_1 + \beta_2 \boldsymbol{v}_2 + \cdots + \beta_r \boldsymbol{v}_r,$$

para ciertos escalares $\alpha_i, \beta_i \in \mathbb{K}$, $i = 1, \ldots, r$. Entonces

$$\begin{aligned} \alpha \boldsymbol{v} + \beta \boldsymbol{w} &= \alpha \sum_{i=1}^{r} \alpha_i \boldsymbol{v}_1 + \beta \sum_{i=1}^{r} \beta_i \boldsymbol{v}_i \\ &= \sum_{i=1}^{r} (\alpha \alpha_i + \beta \beta_i) \boldsymbol{v}_i = \sum_{i=1}^{r} \gamma_i \boldsymbol{v}_i, \end{aligned}$$

que tiene la forma de los elementos de $\langle \mathscr{L} \rangle$. Si V es un espacio vectorial tal que $V = \langle \mathscr{L} \rangle$, decimos que $\mathscr{L}$ es un ***conjunto generador*** de V. En otras palabras, $\mathscr{L}$ genera V cuando todo vector de V se puede expresar como combinación lineal de vectores de $\mathscr{L}$.

Ejemplo 4.1.7. Sea $A_{m \times n}$ una matriz. Entonces el conjunto de soluciones del sistema lineal homogéneo $A\boldsymbol{x} = \boldsymbol{0}$ es un subespacio vectorial de $\mathbb{K}^n$. Para probarlo, consideremos dos soluciones $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{K}^n$ y escalares $\alpha_1, \alpha_2 \in \mathbb{K}$. Entonces $\alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2$ es una solución, pues

$$A \cdot (\alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2) = \alpha_1 A \cdot \boldsymbol{u}_1 + \alpha_2 A \cdot \boldsymbol{u}_2 = \alpha_1 \boldsymbol{0} + \alpha_2 \boldsymbol{0} = \boldsymbol{0}.$$

Ejemplo 4.1.8. Consideremos un conjunto de vectores columna

$$\mathscr{L} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$$

de $\mathbb{K}^m$, y formemos la matriz A con columnas $\boldsymbol{a}_i$. Entonces $\mathscr{L}$ genera $\mathbb{K}^m$ si y solamente si para cada $\boldsymbol{b} \in \mathbb{K}^m$ existe una columna $\boldsymbol{x}$ tal que $A\boldsymbol{x} = \boldsymbol{b}$, esto es, el sistema $A\boldsymbol{x} = \boldsymbol{b}$ es compatible para cada $\boldsymbol{b} \in \mathbb{K}^m$. Para verlo, tenemos que $\mathscr{L}$ genera $\mathbb{K}^m$ si y solamente si para cada vector $\boldsymbol{b} \in \mathbb{K}^m$ existen escalares α_i tales que

$$\boldsymbol{b} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_n \boldsymbol{a}_n = \begin{pmatrix} \boldsymbol{a}_1 & | & \boldsymbol{a}_2 & | & \ldots & | & \boldsymbol{a}_n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = A\boldsymbol{\alpha}.$$

Nota 4.1.9. Esta simple observación es muy útil. Por ejemplo, para verificar si

$$\mathscr{L} = \left\{ \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 0 \end{pmatrix} \right\}$$

genera todo $\mathbb{R}^3$, colocamos los vectores como columnas de una matriz A, y nos planteamos si el sistema

$$\begin{pmatrix} 2 & 1 & 5 \\ 1 & 2 & 4 \\ -1 & 2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

es compatible para todo vector $\boldsymbol{b} \in \mathbb{R}^3$. Recordemos que el sistema es compatible si y solamente si rango$([A|\boldsymbol{b}])$ = rango(A). En este caso, rango$(A) = 2$, pero rango$([A|\boldsymbol{b}]) = 3$ para algunos $\boldsymbol{b}$, como por ejemplo $b_1 = 0, b_2 = 0, b_3 = 1$. Por tanto, $\mathscr{L}$ no genera $\mathbb{R}^3$. Por otro lado,

$$\mathscr{L}' = \left\{ \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix} \right\}$$

sí es un conjunto generador de $\mathbb{R}^3$, porque

$$A' = \begin{pmatrix} 2 & 1 & 5 \\ 1 & 2 & 4 \\ -1 & 2 & 1 \end{pmatrix}$$

es una matriz no singular, de donde $A'\boldsymbol{x} = \boldsymbol{b}$ es compatible para todo $\boldsymbol{b}$.

También es posible 'sumar' subespacios para generar otro.

Suma de subespacios

Sean W_1 y W_2 subespacios vectoriales de V. Se define la ***suma*** de W_1 y W_2 como el conjunto de todas las posibles sumas de vectores de W_1 y W_2. Esto es,

$$W_1 + W_2 = \{\boldsymbol{w}_1 + \boldsymbol{w}_2 \mid \boldsymbol{w}_1 \in W_1, \boldsymbol{w}_2 \in W_2\}.$$

Ejemplo 4.1.10. Sea $V = \mathbb{R}^3$ y consideremos los subespacios vectoriales $W_1 = \langle e_1 \rangle, W_2 = \langle e_2 \rangle$. Entonces

$$W_1 + W_2 = \{w_1 + w_2 \mid w_1 = \alpha_1 e_1, w_2 = \alpha_2 e_2\},$$

es decir, $W_1 + W_2$ es el conjunto de vectores de la forma

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ 0 \end{pmatrix}, \text{ donde } \alpha_1, \alpha_2 \in \mathbb{R}.$$

Propiedades de la suma de subespacios

- La suma $W_1 + W_2$ es un subespacio vectorial de V.
- Si $W_i = \langle \mathscr{L}_i \rangle, i = 1,2$ entonces $W_1 + W_2 = \langle \mathscr{L}_1 \cup \mathscr{L}_2 \rangle$.

PRUEBA: Para probar la primera parte, debemos comprobar que las operaciones de suma y producto por un escalar son *internas* al conjunto. Por ejemplo, sean $u, v \in W_1 + W_2$. Entonces existen $u_1, v_1 \in W_1, u_2, v_2 \in W_2$ tales que

$$u = u_1 + u_2, v = v_1 + v_2.$$

Por tanto,

$$u + v = (u_1 + v_1) + (u_2 + v_2) \in W_1 + W_2,$$

y tenemos el resultado. Con respecto al producto por un escalar, sabemos que $\alpha u_1 \in W_1, \alpha u_2 \in W_2$ para cualquier escalar α. Entonces $\alpha u = \alpha u_1 + \alpha u_2 \in W_1 + W_2$.

Veamos que la unión de los conjuntos generadores proporciona un conjunto generador de la suma. Sean

$$\mathscr{L}_1 = \{u_1, \ldots, u_r\}, \mathscr{L}_2 = \{v_1, \ldots, v_s\}.$$

Entonces

$$\begin{aligned} w \in \langle \mathscr{L}_1 \cup \mathscr{L}_2 \rangle &\Leftrightarrow w = \sum_{i=1}^{r} \alpha_i u_i + \sum_{i=1}^{s} \beta_i v_i = u + v \text{ con } u \in W_1, v \in W_2 \\ &\Leftrightarrow w \in W_1 + W_2. \end{aligned}$$

□

4.2. Subespacios asociados a una aplicación lineal

Aplicación lineal

Sea $f : V \to V'$ una aplicación entre dos espacios vectoriales sobre un mismo cuerpo $\mathbb{K}$. Decimos que f es una ***aplicación lineal*** si

- $f(\boldsymbol{v}+\boldsymbol{w}) = f(\boldsymbol{v}) + f(\boldsymbol{w})$,
- $f(\alpha\boldsymbol{v}) = \alpha f(\boldsymbol{v})$,

para todo $\boldsymbol{v}, \boldsymbol{w} \in V$ y todo escalar $\alpha \in \mathbb{K}$.

Las dos condiciones anteriores se pueden combinar bajo la expresión

$$f(\alpha\boldsymbol{v} + \beta\boldsymbol{w}) = \alpha f(\boldsymbol{v}) + \beta f(\boldsymbol{w}),$$

para todos los escalares α, β y vectores $\boldsymbol{v}, \boldsymbol{w} \in V$.

Ejemplo 4.2.1.

- La aplicación traza : $\mathbb{K}^{n\times n} \to \mathbb{K}$ definida por

$$\text{traza}(A) = \sum_{i=1}^{n} a_{ii}$$

  es lineal. Se denomina ***traza*** de la matriz A.

- Sea $A_{m\times n}$ una matriz. La aplicación $f : \mathbb{K}^n \to \mathbb{K}^m$definida por $f(\boldsymbol{v}) = A\boldsymbol{v}$ es una aplicación lineal. Este ejemplo será el más importante para nosotros.

Sea f una aplicación lineal de V en V'. El conjunto

$$\text{im}(f) = \{f(\boldsymbol{v}) \mid \boldsymbol{v} \in V\} \subset V'$$

se denomina ***imagen*** de f.

Imagen de una aplicación lineal

La imagen de cualquier aplicación lineal $f : V \to V'$ es un subespacio de V'.

PRUEBA: Sea f una aplicación lineal de V en V'. Si $\boldsymbol{w}_1$ y $\boldsymbol{w}_2$ son vectores de la imagen de f, entonces existen $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ tales que $f(\boldsymbol{v}_i) = \boldsymbol{w}_i, i = 1,2$. Tenemos que comprobar que para $\alpha, \beta \in \mathbb{K}$, el vector $\boldsymbol{w} = \alpha\boldsymbol{w}_1 + \beta\boldsymbol{w}_2$ está en la imagen de f. Basta considerar $\boldsymbol{v} = \alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2$, y se tiene que $f(\boldsymbol{v}) = \boldsymbol{w}$. □

Tenemos así que toda matriz $A \in \mathbb{K}^{m\times n}$ genera un subespacio en $\mathbb{K}^m$ como imagen de la función lineal $f(\boldsymbol{x}) = A\boldsymbol{x}$ de $\mathbb{K}^n$ en $\mathbb{K}^m$. De manera análoga, la traspuesta A^t define un subespacio en $\mathbb{K}^n$ como la imagen de la aplicación lineal $g(\boldsymbol{y}) = A^t\boldsymbol{y}$.

Espacios columna y fila

El ***espacio de columnas*** de una matriz $A_{m\times n}$ es la imagen de la aplicación lineal $f(\boldsymbol{x}) = A\boldsymbol{x}$. Se notará por $\mathrm{Col}(A)$.
El ***espacio de filas*** de una matriz $A_{m\times n}$ es la imagen de la aplicación lineal $g(\boldsymbol{y}) = A^t\boldsymbol{y}$. Se notará por $\mathrm{Fil}(A)$.

Recordemos que una expresión de la forma $A\boldsymbol{x}$ es una combinación lineal de las columnas de A. Si escribimos $\boldsymbol{x} = \begin{pmatrix} \xi_1 & \xi_2 & \dots & \xi_n \end{pmatrix}^t$, entonces

$$A\boldsymbol{x} = \begin{pmatrix} A_{*1} & A_{*2} & \dots & A_{*n} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} = \sum_{j=1}^{n} \xi_j A_{*j}$$

muestra que el conjunto de las imágenes $A\boldsymbol{x}$ es el mismo que el conjunto de las combinaciones lineales de las columnas de A. Por tanto, si f es la aplicación lineal definida por $f(\boldsymbol{x}) = A\boldsymbol{x}$, entonces $\mathrm{im}(f)$ no es más que el espacio generado por las columnas de A, es decir, $\mathrm{Col}(A)$.

Es interesante saber si dos matrices dadas tiene el mismo espacio columna o no.

Igualdad de espacios columna

Para dos matrices A y B del mismo orden,

- $\mathrm{Col}(A^t) = \mathrm{Col}(B^t) \Leftrightarrow A \overset{\mathrm{f}}{\sim} B$.
- $\mathrm{Col}(A) = \mathrm{Col}(B) \Leftrightarrow A \overset{\mathrm{c}}{\sim} B$.

PRUEBA: Supongamos que $\mathrm{Col}(A^t) = \mathrm{Col}(B^t)$. Entonces cada fila de B se puede expresar como combinación lineal de las filas de A. Esto implica que

$$\begin{pmatrix} A \\ B \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} E_A \\ \mathbf{0} \end{pmatrix}.$$

Análogamente,

$$\begin{pmatrix} B \\ A \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} E_B \\ \mathbf{0} \end{pmatrix}.$$

Como

$$\begin{pmatrix} A \\ B \end{pmatrix} \overset{\text{f}}{\sim} \begin{pmatrix} B \\ A \end{pmatrix},$$

se sigue que $E_A = E_B$, y entonces $A \overset{\text{f}}{\sim} B$. Recíprocamente, si $A \overset{\text{f}}{\sim} B$, existe una matriz no singular P tal que $PA = B$. Para verificar que $\mathrm{Col}(A^t) = \mathrm{Col}(B^t)$, consideramos

$$\begin{aligned} \boldsymbol{a} \in \mathrm{Col}(A^t) &\Leftrightarrow \boldsymbol{a}^t = \boldsymbol{y}^t A = \boldsymbol{y}^t P^{-1} PA \text{ para algún } \boldsymbol{y} \\ &\Leftrightarrow \boldsymbol{a}^t = \boldsymbol{z}^t B \text{ para } \boldsymbol{z}^t = \boldsymbol{y}^t P^{-1} \\ &\Leftrightarrow \boldsymbol{a} \in \mathrm{Col}(B^t). \end{aligned}$$

□

Ejemplo 4.2.2. Dos conjuntos $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_r\}$ y $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_s\}$ en $\mathbb{R}^n$ generan el mismo subespacio si y solamente si las filas no nulas de E_A y E_B coinciden, donde A y B son las matrices que contienen los $\boldsymbol{a}_i$ y $\boldsymbol{b}_i$ como *filas*. Esto es un corolario de lo anterior, pues las filas nulas son irrelevantes a la hora de considerar el espacio de filas de una matriz, y sabemos que $A \overset{\text{f}}{\sim} B$ si y solamente si $E_A = E_B$.

Consideremos el caso

$$\mathscr{A} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \\ 1 \\ 4 \end{pmatrix} \right\}, \mathscr{B} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right\},$$

y veamos que generan el mismo subespacio. Para ello, consideramos

$$A = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = E_A,$$

y

$$B = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} = E_B.$$

Como las filas no nulas de E_A y E_B coinciden, tenemos el resultado.

Generación de los espacios de fila y columna

Sea A una matriz $m \times n$, y U cualquier forma escalonada por filas derivada de A. Entonces

- Las filas no nulas de U generan $\text{Col}(A^t)$, el espacio de filas de A.
- Las columnas básicas de A generan $\text{Col}(A)$, el espacio de columnas de A.

PRUEBA: La primera es inmediata, por la equivalencia por filas. Para la segunda, recordemos que todas las columnas de A se expresan como combinación lineal de las columnas básicas, pues así ocurre en la forma escalonada reducida por filas. Por tanto, las columnas básicas generan el espacio $\text{Col}(A)$. □

Ejemplo 4.2.3. Calculemos un conjunto de generadores para $\text{Col}(A)$ y $\text{Col}(A^t)$ (espacios de filas y columnas de A), donde

$$A = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 1 & 3 \\ 3 & 6 & 1 & 4 \end{pmatrix}.$$

Para ello,

$$A \to E_A = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Entonces $\text{Col}(A)$ está generado por las columnas básicas de A, esto es,

$$\text{Col}(A) = \langle \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \rangle,$$

y $\text{Col}(A^t)$ está generado por las filas no nulas de E_A, es decir,

$$\text{Col}(A^t) = \langle \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \rangle.$$

Espacio nulo

- Para una matriz A de orden $m \times n$, el conjunto $\text{null}(A) = \{\boldsymbol{x}_{n\times 1} \mid A\boldsymbol{x} = \boldsymbol{0}\} \subset \mathbb{K}^n$ se denomina ***espacio nulo*** o ***núcleo*** de A. Es el conjunto de soluciones del sistema homogéneo $A\boldsymbol{x} = \boldsymbol{0}$.
- El conjunto $\text{null}(A^t) = \{\boldsymbol{y}_{m\times 1} \mid A^t\boldsymbol{y} = \boldsymbol{0}\} \subset \mathbb{K}^m$ se denomina ***espacio nulo a la izquierda*** de A, porque es el conjunto de soluciones del sistema homogéneo $\boldsymbol{y}^t A = \boldsymbol{0}$.

Ejemplo 4.2.4. Consideremos la matriz

$$A = \begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

Entonces el cálculo de $\text{null}(A)$ se reduce a obtener el conjunto de soluciones del sistema lineal homogéneo $A\boldsymbol{x} = \boldsymbol{0}$. Para ello, calculamos

$$A \xrightarrow{\text{rref}} E_A = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \end{pmatrix},$$

de donde el conjunto de soluciones es el espacio generado por el vector

$$\boldsymbol{h}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

Núcleo de una aplicación

Sea $f : V \to V'$ una aplicación lineal entre espacios vectoriales. El conjunto

$$\ker(f) = \{\boldsymbol{v} \in V \mid f(\boldsymbol{v}) = \boldsymbol{0}\}$$

se denomina ***núcleo*** de la aplicación lineal f.

Ejemplo 4.2.5. Sea $V = \mathbb{R}^{2\times 2}$ el conjunto de las matrices de orden 2×2 y consideremos la aplicación $f : V \to \mathbb{R}$ definida por la traza de la matriz. Entonces $\ker(f)$ es el conjunto de matrices 2×2 de la forma

$$\begin{pmatrix} a & b \\ c & -a \end{pmatrix}.$$

Es fácil ver que $\ker(f)$ es un subespacio vectorial de f, pues si $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \ker(f)$ y $\alpha_1, \alpha_2 \in \mathbb{K}$, entonces

$$\begin{aligned} f(\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2) &= f(\alpha_1\boldsymbol{v}_1) + f(\alpha_2\boldsymbol{v}_2) = \alpha_1 f(\boldsymbol{v}_1) + \alpha_2 f(\boldsymbol{v}_2) \\ &= \alpha_1\mathbf{0} + \alpha_2\mathbf{0} = \mathbf{0}, \end{aligned}$$

por lo que $\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 \in \ker(f)$.

Si $A_{m\times n}$ es una matriz, y f es la aplicación lineal definida por $f(\boldsymbol{x}) = A\boldsymbol{x}$, vemos que $\ker(f) = \text{null}(A)$.

Espacio nulo trivial

Si A es una matriz $m \times n$, entonces $\text{null}(A) = \mathbf{0}$ si y solamente si $\text{rango}(A) = n$.

PRUEBA: La solución trivial $\boldsymbol{x} = \mathbf{0}$ es la única solución de $A\boldsymbol{x} = \mathbf{0}$ si y solamente si el rango de A es igual al número de incógnitas. □

4.3. Independencia lineal

Independencia lineal

Un conjunto de vectores $\mathscr{L} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ se dice ***linealmente independiente*** si la única solución para los escalares α_i en la ecuación homogénea

$$\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \cdots + \alpha_n\boldsymbol{v}_n = \mathbf{0}$$

es la solución trivial $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. En otro caso se dice que es un ***conjunto linealmente dependiente***.

Las relaciones de dependencia entre vectores salen a la luz al calcular la forma escalonada reducida por filas.

Ejemplo 4.3.1. Vamos a determinar si el conjunto

$$\mathscr{L} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix} \right\}$$

es linealmente independiente. Aplicamos la definición, y buscamos si existe una solución no trivial de

$$\alpha_1\begin{pmatrix}1\\2\\1\end{pmatrix}+\alpha_2\begin{pmatrix}1\\0\\2\end{pmatrix}+\alpha_3\begin{pmatrix}5\\6\\7\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Esto es equivalente a estudiar las soluciones del sistema lineal homogéneo

$$\begin{pmatrix}1&1&5\\2&0&6\\1&2&7\end{pmatrix}\begin{pmatrix}\alpha_1\\\alpha_2\\\alpha_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Si A es la matriz de coeficientes del sistema, entonces la forma escalonada reducida por filas es

$$E_A=\begin{pmatrix}1&0&3\\0&1&2\\0&0&0\end{pmatrix}.$$

Esto significa que existen soluciones no triviales, y $\mathscr{L}$ es un conjunto linealmente dependiente. En particular, E_A nos indica que $A_{*3}=3A_{*1}+2A_{*2}$ y

$$3\begin{pmatrix}1\\2\\1\end{pmatrix}+2\begin{pmatrix}1\\0\\2\end{pmatrix}-\begin{pmatrix}5\\6\\7\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Ejemplo 4.3.2. Supongamos que en un conjunto de vectores $\mathscr{L}=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r\}$, uno de ellos se expresa como combinación lineal de los restantes. Por ejemplo,

$$\boldsymbol{v}_r=\alpha_1\boldsymbol{v}_1+\alpha_2\boldsymbol{v}_2+\cdots+\alpha_{r-1}\boldsymbol{v}_{r-1}.$$

Entonces el conjunto $\mathscr{L}$ es linealmente dependiente. En efecto, la relación anterior implica que

$$\alpha_1\boldsymbol{v}_1+\alpha_2\boldsymbol{v}_2+\cdots+\alpha_{r-1}\boldsymbol{v}_{r-1}-\boldsymbol{v}_r=\boldsymbol{0},$$

y el coeficiente que afecta a $\boldsymbol{v}_r$ es no nulo.

Independencia lineal y matrices

Sea A una matriz $m \times n$.

- Cada una de las siguientes sentencias es equivalente a decir que las columnas de A forman un conjunto linealmente independiente.
  - $\text{null}(A) = \mathbf{0}$.
  - $\text{rango}(A) = n$.
- Cada una de las siguientes sentencias es equivalente a decir que las filas de A forman un conjunto linealmente independiente.
  - $\text{null}(A^t) = \mathbf{0}$.
  - $\text{rango}(A) = m$.
- Cuando A es una matriz cuadrada, cada una de las siguientes afirmaciones es equivalente a decir que A es no singular.
  - Las columnas de A forman un conjunto linealmente independiente.
  - Las filas de A forman un conjunto linealmente independiente.

PRUEBA: Por definición, las columnas de A forman un conjunto linealmente independiente cuando el único conjunto de escalares α_i que satisface la ecuación homogénea

$$\mathbf{0} = \alpha_1 A_{*1} + \alpha_2 A_{*2} + \cdots + \alpha_n A_{*n} = \begin{pmatrix} A_{*1} & A_{*2} & \dots & A_{*n} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

es la solución trivial $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Pero esto significa que $\text{null}(A) = \mathbf{0}$, que es equivalente a $\text{rango}(A) = n$. El resto sigue cambiando A por A^t. □

Ejemplo 4.3.3. Matrices de Vandermonde. Las matrices de la forma

$$V_{m\times n} = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{pmatrix},$$

en donde $x_i \neq x_j$ para todo $i \neq j$ se llaman ***matrices de Vandermonde***. Las columnas de V constituyen un conjunto linealmente independiente cuando $n \leq m$. Para ver esto, recordemos que es equivalente a probar que $\text{null}(V) = \mathbf{0}$. Si

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

entonces, para cada $i = 1, 2, \dots, m$,

$$\alpha_0 + x_i\alpha_1 + x_i^2\alpha_2 + \cdots + x_i^{n-1}\alpha_{n-1} = 0.$$

Esto implica que el polinomio

$$p(x) = \alpha_0 + x\alpha_1 + x^2\alpha_2 + \cdots + x^{n-1}\alpha_{n-1}$$

tiene m raíces distintas, en concreto, las x_i. Pero $\deg(p(x)) \leq n-1$, y si $p(x)$ no es el polinomio nulo entonces $p(x)$ tiene, a lo más, $n-1 < m$ raíces. Por tanto, el sistema se verifica si y solamente si $\alpha_i = 0$ para todo i, lo que significa que las columnas de V forman un conjunto linealmente independiente.

Ejemplo 4.3.4. Dado un conjunto de m puntos $\mathscr{L} = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, en donde los x_i son distintos dos a dos, existe un único polinomio

$$L(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_{m-1} t^{m-1} \tag{4.3.1}$$

de grado $m-1$ tal que $L(x_i) = y_i, i = 1, 2, \dots, m$. En efecto, los coeficientes α_i deben satisfacer el sistema

$$\begin{array}{rcccl} \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \cdots + \alpha_{m-1} x_1^{m-1} & = & L(x_1) & = & y_1, \\ \alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2 + \cdots + \alpha_{m-1} x_2^{m-1} & = & L(x_2) & = & y_2, \\ & & \vdots & & \vdots \\ \alpha_0 + \alpha_1 x_m + \alpha_2 x_m^2 + \cdots + \alpha_{m-1} x_m^{m-1} & = & L(x_m) & = & y_m. \end{array}$$

Si lo escribimos en forma matricial,

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{m-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{m-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{m-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{m-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix},$$

se tiene que la matriz de coeficientes es una matriz de Vandermonde, y como los valores x_i son distintos dos a dos, dicha matriz es no singular. Por tanto, el sistema tiene solución única, y existe un único conjunto de coeficientes para el polinomio $L(t)$ en 4.3.1. Si queremos ser más específicos, la solución es

$$L(t) = \sum_{i=1}^{m} \left(y_i \frac{\prod_{j \neq i}^{m} (t - x_j)}{\prod_{j \neq i}^{m} (x_i - x_j)} \right).$$

Es fácil de verificar que el lado derecho es un polinomio de grado $m-1$ que pasa por los puntos de $\mathscr{L}$ y, por tanto, tiene que coincidir con nuestra solución única. El polinomio $L(t)$ se conoce como el ***polinomio de interpolación de Lagrange*** de grado $m-1$.

Nos encontraremos con sistemas similares cuando estudiemos ajuste de curvas mediante mínimos cuadrados.

Si rango$(A_{m\times n}) < n$, entonces las columnas de A forman un conjunto linealmente dependiente. Para tales matrices queremos extraer un ***subconjunto maximal linealmente independiente*** de columnas, es decir, un conjunto linealmente independiente al que no le podemos añadir otra columna de la matriz A que mantenga dicho carácter. Aunque hay varias formas de realizar tal selección, las columnas básicas constituyen una solución.

Subconjuntos maximales independientes

Si rango$(A_{m\times n}) = r$ entonces:

- Cualquier subconjunto maximal independiente de columnas de A contiene exactamente r columnas.
- Cualquier subconjunto maximal independiente de filas de A contiene exactamente r filas.
- En particular, las r columnas básicas de A constituyen un subconjunto maximal independiente de columnas de A.

PRUEBA: Recordemos que las relaciones que tengan las columnas de A existen entre las columnas de E_A. Esto garantiza que un conjunto de columnas de A es linealmente independiente si y solamente si las columnas en las posiciones correspondientes de E_A son un conjunto independiente. Sea

$$C = \begin{pmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \dots & \boldsymbol{c}_k \end{pmatrix}$$

una matriz que contiene un subconjunto independiente de columnas de E_A tal que rango$(C) = k$. Como cada columna de E_A es una combinación de las r columnas básicas de E_A (vectores $\boldsymbol{e}_i$), existen escalares β_{ij} tales que $\boldsymbol{c}_j = \sum_{i=1}^{r} \beta_{ij}\boldsymbol{e}_i$ para $j = 1, 2, \dots, k$. Estas ecuaciones se pueden escribir en forma matricial como

$$\begin{pmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \dots & \boldsymbol{c}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \dots & \boldsymbol{e}_r \end{pmatrix}_{m\times r} \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1k} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2k} \\ \vdots & \vdots & \dots & \vdots \\ \beta_{r1} & \beta_{r2} & \dots & \beta_{rk} \end{pmatrix}_{r\times k}$$

o bien

$$C_{m\times k} = \begin{pmatrix} I_r \\ \mathbf{0} \end{pmatrix} B_{r\times k} = \begin{pmatrix} B_{r\times k} \\ \mathbf{0} \end{pmatrix}, \text{ donde } B_{r\times k} = (\beta_{ij}).$$

Por tanto, en C hay, a lo más, r filas no nulas, de donde $r \geq$ rango$(C) = k$, y cualquier subconjunto independiente de columnas de E_A, y por tanto de A, no puede contener más de r vectores. Como las r columnas básicas de E_A forman un conjunto independiente, las r columnas básicas de A también. La segunda parte de la proposición se deduce de rango(A) = rango(A^t). □

Cuestiones básicas de independencia

Para un conjunto no vacío de vectores $\mathscr{L} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n\}$ de un espacio V, se tiene que:

- Si $\mathscr{L}$ contiene un subconjunto linealmente dependiente, entonces $\mathscr{L}$ es linealmente dependiente.
- Si $\mathscr{L}$ es linealmente independiente, entonces todo subconjunto de $\mathscr{L}$ es linealmente independiente.
- Si $\mathscr{L}$ es linealmente independiente y $\boldsymbol{v} \in V$, entonces el conjunto extensión $\mathscr{L}_{\text{ext}} = \mathscr{L} \cup \{\boldsymbol{v}\}$ es linealmente independiente si y solamente si $\boldsymbol{v} \notin \langle \mathscr{L} \rangle$.
- Si $\mathscr{L} \subset \mathbb{K}^m$ y $n > m$ entonces $\mathscr{L}$ es linealmente dependiente.

PRUEBA:

- Supongamos que $\mathscr{L}$ contiene un subconjunto linealmente dependiente, y por conveniencia, supongamos que dicho conjunto está formado por $\mathscr{L}' = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$. Por definición de dependencia lineal, existen unos escalares $\alpha_1, \ldots, \alpha_k$, no todos nulos, tales que

$$\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_k \boldsymbol{u}_k = \boldsymbol{0}.$$

  Entonces podemos escribir

$$\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_k \boldsymbol{u}_k + 0\boldsymbol{u}_{k+1} + \cdots + 0\boldsymbol{u}_n = \boldsymbol{0},$$

  que es una combinación lineal no trivial de los elementos de $\mathscr{L}$.

- Es consecuencia inmediata de lo anterior.

- Si $\mathscr{L}_{\text{ext}}$ es linealmente independiente, entonces $\boldsymbol{v} \notin \langle \mathscr{L} \rangle$, pues en otro caso tendríamos una expresión de la forma

$$\boldsymbol{v} = \alpha_1 \boldsymbol{u}_1 + \cdots + \alpha \boldsymbol{u}_n,$$

  que implicaría la dependencia lineal de $\mathscr{L}_{\text{ext}}$. Recíprocamente, supongamos ahora que $\boldsymbol{v} \notin \langle \mathscr{L} \rangle$, y consideremos una combinación lineal de los elementos de $\mathscr{L}_{\text{ext}}$ de la forma

$$\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_n \boldsymbol{u}_n + \alpha_{n+1} \boldsymbol{v} = \boldsymbol{0}.$$

  Como $\boldsymbol{v}$ no se puede expresar en función de los elementos de $\mathscr{L}$, tenemos que $\alpha_{n+1} = 0$. Entonces nos queda

$$\alpha_1 \boldsymbol{u}_1 + \cdots + \alpha_n \boldsymbol{u}_n = \boldsymbol{0},$$

  y la independencia lineal de $\mathscr{L}$ implica que $\alpha_1 = \cdots = \alpha_n = 0$.

- Si colocamos las columnas de $\mathscr{L}$ en una matriz $A_{m \times n}$, entonces $\text{rango}(A) \leq m < n$.

□

4.4. Bases y dimensión

Un conjunto generador de un espacio vectorial puede contener vectores redundantes, de forma que el espacio podría ser generado por un número menor de vectores. Se trata de determinar cuántos y cuáles hacen falta.

Base de un espacio vectorial

Un conjunto linealmente independiente y generador de un espacio vectorial V se denomina ***base*** de V.

Todo espacio vectorial tiene una base, y una vez que se ha encontrado una, podemos encontrar tantas como queramos.

Ejemplo 4.4.1.

- Los vectores unitarios $\mathscr{S} = \{e_1, e_2, \ldots, e_n\}$ de $\mathbb{R}^n$ forman una base de $\mathbb{R}^n$. La llamamos ***base estándar*** de $\mathbb{K}^n$.
- Si A es una matriz $n \times n$ no singular, entonces el conjunto de filas, así como el de columnas de A, forman una base de $\mathbb{K}^n$.

Ejemplo 4.4.2. Consideremos la matriz

$$A = \begin{pmatrix} 1 & 1 & 5 \\ 2 & 0 & 6 \\ 1 & 2 & 7 \end{pmatrix},$$

cuya forma escalonada reducida por filas es

$$E_A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Sabemos que las columnas básicas de A, que son A_{*1} y A_{*2}, generan el espacio $\mathrm{Col}(A)$. Además, son independientes, pues si existe una expresión $\alpha_1 A_{*1} + \alpha_2 A_{*2} = \mathbf{0}$, para ciertos escalares $\alpha_1, \alpha_2 \in \mathbb{R}$, entonces dicha relación se tiene entre las columnas correspondientes de E_A, esto es, $\alpha_1 e_1 + \alpha_2 e_2 = \mathbf{0}$; esto implica que $\alpha_1 = \alpha_2 = 0$. Por tanto, las columnas básicas de A forman una base de $\mathrm{Col}(A)$.

Ejemplo 4.4.3. Vamos a calcular una base del espacio nulo de la matriz

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

Procedemos a calcular su forma escalonada reducida por filas E_A, que nos permitirá obtener un conjunto generador de null(A):

$$A \xrightarrow{\text{rref}} E_A = \begin{pmatrix} 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

que da lugar al sistema homogéneo

$$\begin{cases} x_1 & -x_2 & & & +x_5 & = & 0, \\ & & x_3 & & +x_5 & = & 0, \\ & & & x_4 & -x_5 & = & 0. \end{cases}$$

Las variables básicas son x_1, x_3 y x_4; despejamos para obtener

$$\begin{cases} x_1 & = & x_2 & -x_5, \\ x_2 & = & x_2, & \\ x_3 & = & & -x_5, \\ x_4 & = & & x_5, \\ x_5 & = & & x_5. \end{cases} \quad \text{y entonces } \boldsymbol{x} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} = x_2\boldsymbol{h}_1 + x_5\boldsymbol{h}_2.$$

Podemos escribir que null$(A) = \langle \boldsymbol{h}_1, \boldsymbol{h}_2 \rangle$. Observemos que el conjunto $\{\boldsymbol{h}_1, \boldsymbol{h}_2\}$ es linealmente independiente, pues una expresión de la forma $\alpha_1\boldsymbol{h}_1 + \alpha_2\boldsymbol{h}_2 = \boldsymbol{0}$ implica que, al fijarnos en las segunda y quita componentes, que $\alpha_1 = 0, \alpha_2 = 0$. Por tanto, $\{\boldsymbol{h}_1, \boldsymbol{h}_2\}$ es una base de null(A).

Los espacio vectoriales que tienen bases con una cantidad infinita de elementos se llaman ***espacios infinito dimensionales***. Los que tienen una base finita se denominan ***espacios finito dimensionales***. Nosotros nos restringiremos a los de dimensión finita, que en realidad se reducen a $\mathbb{R}^n$ y $\mathbb{C}^n$.

Caracterización de una base

Sea V un subespacio de $\mathbb{K}^m$ y $\mathscr{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\} \subset V$. Son equivalentes:

1. $\mathscr{B}$ es una base de V.
2. $\mathscr{B}$ es un conjunto generador minimal de V.
3. $\mathscr{B}$ es un conjunto linealmente independiente maximal.

PRUEBA: 1) $\Rightarrow$ 2). Aquí minimal significa que no hay bases de tamaño inferior. Procedemos por reducción al absurdo. Supongamos que $\mathscr{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\}$ es una base de V, con $k < n$. Cada $\boldsymbol{b}_j$ se puede expresar como combinación lineal de los vectores $\boldsymbol{c}_i$. Así, existen escalares α_{ij} tales que

$$\boldsymbol{b}_j = \sum_{i=1}^{k} \alpha_{ij}\boldsymbol{c}_i, \text{ para } j = 1, 2, \ldots, n. \tag{4.4.1}$$

Si construimos las matrices

$$B_{m\times n} = \begin{pmatrix} \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_n \end{pmatrix}, C_{m\times k} = \begin{pmatrix} \boldsymbol{c}_1 & \ldots & \boldsymbol{c}_k \end{pmatrix},$$

las expresiones de 4.4.1 se pueden escribir de manera matricial como

$$B = CA, \text{ donde } A_{k\times n} = (\alpha_{ij}).$$

El rango de una matriz no puede exceder sus dimensiones, y como $k < n$, tenemos que $\text{rango}(A) \leq k < n$. Entonces $\text{null}(A) \neq \mathbf{0}$. Si $\boldsymbol{z} \neq 0$ es tal que $A\boldsymbol{z} = \mathbf{0}$, entonces $B\boldsymbol{z} = \mathbf{0}$. Pero esto es imposible, porque las columnas de B son linealmente independientes, y por tanto $\text{null}(B) = \mathbf{0}$. Así, la hipótesis inicial de la existencia de una base con menos de n elementos es falsa.

2) $\Rightarrow$ 1). Debemos probar la independencia lineal de $\mathscr{B}$. Supongamos que no es así, y entonces uno de los vectores $\boldsymbol{b}_i$ se podría expresar como combinación lineal de los restantes vectores $\boldsymbol{b}_j$. Entonces el conjunto

$$\mathscr{B}' = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{i-1}, \boldsymbol{b}_{i+1}, \ldots, \boldsymbol{b}_n\}$$

seguiría siendo conjunto generador, pero con menos vectores que $\mathscr{B}$, que, por hipótesis, era minimal.

3) $\Rightarrow$ 1). Si $\mathscr{B}$ fuera un conjunto maximal linealmente independiente, pero no fuera base, existiría un vector $\boldsymbol{v} \in V$, pero que $\boldsymbol{v} \notin \langle\mathscr{B}\rangle$. Entonces el conjunto extendido

$$\mathscr{B} \cup \{\boldsymbol{v}\} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n, \boldsymbol{v}\}$$

sería linealmente independiente, en contra de la propiedad de maximal de $\mathscr{B}$.

1) $\Rightarrow$ 3). Supongamos que $\mathscr{B}$ es una base de V, pero que no es maximal. Si existe $\mathscr{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k\} \subset V$, con $k > n$ un conjunto linealmente independiente, entonces todo vector de V se podría expresar como combinación lineal de los vectores $\boldsymbol{c}_i$, por su carácter maximal. Entonces $\mathscr{C}$ es base, pero según 2), el conjunto $\mathscr{C}$ tendría que ser un conjunto minimal de generadores, y $\mathscr{B}$ es uno más pequeño. Por tanto, $\mathscr{B}$ es un conjunto maximal de vectores linealmente independientes. □

Aunque un espacio V puede tener muchas bases diferentes, lo anterior garantiza que todas ellas tienen el mismo número de elementos, que se denomina ***dimensión*** de V.

Dimensión

La ***dimensión*** de un espacio vectorial V se define como

$$\begin{aligned}\dim V = &\text{ número de vectores de cualquier base de } V\\ = &\text{ número de vectores de cualquier}\\ &\text{conjunto generador minimal de } V\\ = &\text{ número de vectores de cualquier}\\ &\text{conjunto linealmente independiente maximal de } V\end{aligned}$$

Si V es un espacio de dimensión n, entonces todo conjunto independiente $\mathscr{L} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\} \subset V$ que contiene n vectores es una base de V.

Una forma de pensar en la dimensión es en términos de *grados de libertad*. En el espacio trivial $\mathscr{Z}$ no hay grados de libertad (dimensión cero), en una recta hay un grado, en un plano dos, etc.

Los resultados anteriores permiten la demostración de la existencia de una base para cualquier espacio vectorial. La idea es usar una iteración. Por ejemplo, sea V un espacio vectorial. Si $V = \mathbf{0}$, hemos acabado. Si no, existe $\boldsymbol{v}_1 \in V$ no nulo. Si $V = \langle \boldsymbol{v}_1 \rangle$, fin del proceso. En otro caso, existe $\boldsymbol{v}_2 \notin \langle \boldsymbol{v}_1 \rangle$. Y así de forma reiterada podemos construir una base de cualquier espacio de dimensión finita.

Ejemplo 4.4.4. Si $\mathscr{L}_r = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$ es un conjunto linealmente independiente en un espacio vectorial V de dimensión n, entonces podemos encontrar vectores extensión $\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$ de V tales que

$$\mathscr{L}_n = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$$

es una base de V. Veamos un procedimiento para encontrar una extensión. Sea $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ cualquier base de V, y formemos la matriz

$$A = \begin{pmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_r & \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_n \end{pmatrix}.$$

Es claro que $\mathrm{Col}(A) = V$, por lo que las columnas básicas de A forman una base de V. Observemos que $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$ son columnas básicas de A, porque ninguna de ellas es combinación lineal de las anteriores. Por tanto, las restantes $n - r$ columnas básicas deben ser un subconjunto de $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$, digamos que $\{\boldsymbol{b}_{j_1}, \boldsymbol{b}_{j_2}, \ldots, \boldsymbol{b}_{j_{n-r}}\}$. Entonces una base de V que extiende a $\mathscr{L}$ es

$$\mathscr{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{b}_{j_1}, \ldots, \boldsymbol{b}_{j_{n-r}}\}.$$

Por ejemplo, para extender el conjunto

$$\mathscr{L} = \left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -2 \end{pmatrix} \right\}$$

a una base de $\mathbb{R}^4$, añadimos la base estándar $\{e_1, e_2, e_3, e_4\}$ a los vectores de $\mathscr{L}$, y reducimos:

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 2 & -2 & 0 & 0 & 0 & 1 \end{pmatrix} \to E_A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1/2 \end{pmatrix}.$$

Entonces $\{A_{*1}, A_{*2}, A_{*4}, A_{*5}\}$ son las columnas básicas de A, y

$$\mathscr{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

es una base de $\mathbb{R}^4$ que contiene a $\mathscr{L}$.

Dimensión de un subespacio

Para espacios vectoriales L_1 y L_2 tales que $L_1 \subset L_2$, se verifica que

- $\dim(L_1) \leq \dim(L_2)$.
- Si $\dim(L_1) = \dim(L_2)$ entonces $L_1 = L_2$.

PRUEBA: Sea $\dim L_1 = m$ y $\dim L_2 = n$. Si $m > n$, entonces una base de L_1 es un conjunto linealmente independiente con más de n vectores. Pero $n = \dim L_2$ es es el tamaño de un conjunto maximal linealmente independiente dentro de L_2. Por tanto, $m \leq n$.

Si $m = n$ pero $L_1 \neq L_2$, existe un vector $\boldsymbol{w} \in L_2$ que no está en L_1. Si $\mathscr{B}$ es una base de L_1, entonces $\mathscr{B} \cup \{\boldsymbol{w}\}$ es un conjunto linealmente independiente, subconjunto de L_2, con $m+1 = n+1$ elementos. Esto es imposible, porque la dimensión de L_2 es n, que es el tamaño de un conjunto maximal linealmente independiente. Por tanto, $L_1 = L_2$. □

Subespacios fundamentales: bases y dimensión

Sea A una matriz de orden $m \times n$, y rango$(A) = r$.

- $\dim(\text{Col}(A)) = r$.
- $\dim(\text{null}(A)) = n - r$.
- $\dim(\text{Col}(A^t)) = r$.
- $\dim(\text{null}(A^t)) = m - r$.

Sea P una matriz no singular tal que $PA = U$ es una forma escalonada por filas, y sea $\mathscr{H}$ el conjunto de $\boldsymbol{h}_i$ que aparecen en la solución general de $A\boldsymbol{x} = \mathbf{0}$.

- Las columnas básicas de A forman una base de Col(A).
- Las filas no nulas de U forman una base de Col(A^t).
- El conjunto $\mathscr{H}$ es una base de null(A).
- Las últimas $m - r$ filas de P forman una base de null(A^t).

Para matrices con entradas complejas, lo anterior queda igual si cambiamos A^t por A^*.

PRUEBA: El conjunto de columnas de A genera Col(A), pero no tiene que ser una base por las posibles dependencias entre las columnas. Sin embargo, el conjunto de columnas *básicas* es también un conjunto generador, y forman un conjunto independiente: ninguna columna básica puede depender linealmente de las otras, pues entonces lo mismo ocurriría con las columnas correspondientes de la forma escalonada reducida por filas, donde es evidente que no sucede. Por tanto, el conjunto de columnas básicas de A forma una base de Col(A), y $\dim \text{Col}(A) = \text{rango}(A) = r$.

Análogamente, el conjunto de filas de A genera Col(A^t), pero puede haber dependencias entre ellas. Recordemos que si

$$U = \begin{pmatrix} C_{r\times n} \\ \mathbf{0} \end{pmatrix}$$

es una forma escalonada equivalente a A, entonces las filas de C generan Col(A^t). Como rango$(C) = r$, las filas de C son linealmente independientes. Entonces

rango$(A^t) = \dim \text{Col}(A^t) = r$, que es lo que se conoce como igualdad de rangos de los espacios de fila y columna.

Veamos ahora los espacios nulos. Recordemos que el conjunto $\mathscr{H}$ que contenía los $\boldsymbol{h}_i$ que aparecían en la solución general de $A\boldsymbol{x} = \boldsymbol{0}$ generaban null(A). Además, son independientes. Recordemos que los vectores $\boldsymbol{h}_i$ tienen un 1 en la posición f_i asociada a la variable libre, y todos los demás vectores $\boldsymbol{h}_j$ tienen un cero en esa posición. Entonces, si consideramos una expresión

$$\alpha_1 \boldsymbol{h}_1 + \alpha_2 \boldsymbol{h}_2 + \cdots + \alpha_{n-r} \boldsymbol{h}_{n-r} = \boldsymbol{0},$$

obtenemos igualdades de la forma $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_{n-r} = 0$. Por tanto,

$$\dim(\text{null}(A)) = n - r.$$

□

Se sigue inmediatamente el siguiente resultado:

Teorema de la dimensión

$$\dim(\text{Col}(A)) + \dim(\text{null}(A)) = n$$

para todas las matrices $m \times n$.

Dimensión de la suma

Si L_1 y L_2 son subespacios de un espacio vectorial V, entonces

$$\dim(L_1 + L_2) + \dim(L_1 \cap L_2) = \dim(L_1) + \dim(L_2).$$

PRUEBA: La estrategia es construir una base de $L_1 + L_2$ y contar el número de vectores que contiene. Sea $\mathscr{L} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$ una base de $L_1 \cap L_2$. Existe una extensión $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ a una base de L_1, y otra $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ a una base de L_2. Entonces

$$\mathscr{B}_1 = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\} \text{ es una base de } L_1$$

y

$$\mathscr{B}_2 = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} \text{ es una base de } L_2.$$

Sabemos que $\mathscr{B}_1 \cup \mathscr{B}_2$ es un conjunto generador de $L_1 + L_2$, y queremos probar que es linealmente independiente. Consideremos para ello

$$\sum_{i=1}^{t} \alpha_i \boldsymbol{z}_i + \sum_{j=1}^{n} \beta_j \boldsymbol{u}_j + \sum_{k=1}^{m} \gamma_k \boldsymbol{v}_k = \boldsymbol{0},$$

que podemos escribir como

$$\sum_{k=1}^{m} \gamma_k \boldsymbol{v}_k = -\left(\sum_{i=1}^{t} \alpha_i \boldsymbol{z}_i + \sum_{j=1}^{n} \beta_j \boldsymbol{u}_j \right) \in L_1.$$

Es claro que el lado izquierdo de la anterior igualdad está en L_2, por lo que existen escalares δ_i tales que

$$\sum_{k=1}^{m} \gamma_k \boldsymbol{v}_k = \sum_{i=1}^{t} \delta_i \boldsymbol{z}_i, \text{ o de forma equivalente } \sum_{k=1}^{m} \gamma_k \boldsymbol{v}_k - \sum_{i=1}^{t} \delta_i \boldsymbol{z}_i = \boldsymbol{0}.$$

Como $\mathscr{B}_2$ es una base, todos los coeficientes de la expresión anterior son nulos: $\gamma_k = 0, k = 1, \ldots, m, \delta_i = 0, i = 1, \ldots, t$. Entonces nos queda que

$$\sum_{i=1}^{t} \alpha_i \boldsymbol{z}_i + \sum_{j=1}^{n} \beta_j \boldsymbol{u}_j = \boldsymbol{0}.$$

De nuevo, $\mathscr{B}_1$ es base, y todos los coeficientes son nulos. Hemos probado así que los vectores de $\mathscr{B}_1 \cup \mathscr{B}_2$ forman una base de $L_1 + L_2$, y

$$\dim(L_1 + L_2) = t + m + n = (t + m) + (t + n) - t = \dim L_1 + \dim L_2 - \dim(L_1 \cap L_2).$$

□

4.5. Transformaciones lineales

Transformación lineal

Sean U y V espacios vectoriales sobre un cuerpo $\mathbb{K}$ ($\mathbb{R}$ o $\mathbb{C}$ para nosotros).

- Una ***transformación lineal*** de U en V es una aplicación lineal $T : U \to V$.
- Un ***endomorfismo*** de U es una aplicación lineal de U en sí mismo.

Ejemplo 4.5.1.

- Si $A \in \mathbb{R}^{m\times n}$ y $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, la función $T(\boldsymbol{x}) = A\boldsymbol{x}$ es una transformación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$. T es un endomorfismo de $\mathbb{R}^n$ si A es de orden $n \times n$.

- La rotación Q de un vector $\boldsymbol{u}$ en $\mathbb{R}^2$ un ángulo θ en el sentido contrario a las agujas del reloj se puede describir como una multiplicación matricial. Si $\boldsymbol{u} = (x, y)^t$, entonces

$$Q(\boldsymbol{u}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

- La proyección de un vector $\boldsymbol{v} \in \mathbb{R}^3$ en $(x, y, 0)^t$, vector del plano xy, está definida por la matriz

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

- La simetría R que aplica cada vector $\boldsymbol{v} = (x, y, z)^t \in \mathbb{R}^3$ en $R(\boldsymbol{v}) = (x, y, -z)^t$ se puede representar por la matriz

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Las transformaciones lineales entre espacios de dimensión finita siempre se pueden representar por una matriz. Para probarlo, es necesario el concepto de coordenada.

Sea $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ una base de un espacio vectorial U, y tomemos $\boldsymbol{u} \in U$. Entonces, por ser sistema generador, existen unos escalares $\alpha_i, i = 1, \ldots, n$ tales que

$$\boldsymbol{u} = \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \cdots + \alpha_n\boldsymbol{u}_n.$$

Los escalares α_i están unívocamente determinados, pues si

$$\boldsymbol{u} = \sum_{i=1}^{n} \alpha_i\boldsymbol{u}_i = \sum_{i=1}^{n} \beta_i\boldsymbol{u}_i,$$

entonces

$$\boldsymbol{0} = \sum_{i=1}^{n} (\alpha_i - \beta_i)\boldsymbol{u}_i,$$

y, por la independencia lineal de $\mathscr{B}$, se tiene que $\alpha_i = \beta_i$ para todo $i = 1, \ldots, n$.

Coordenadas de un vector

Sea $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ una base de un espacio vectorial U, y sea $\boldsymbol{u} \in U$. Los coeficientes α_i en la expresión $\boldsymbol{u} = \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \cdots + \alpha_n\boldsymbol{u}_n$ se llaman las ***coordenadas*** de $\boldsymbol{u}$ ***respecto de*** la base $\mathscr{B}$. Lo notaremos por

$$[\boldsymbol{u}]_{\mathscr{B}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Ejemplo 4.5.2. Consideremos en $\mathbb{R}^3$ la base formada por los vectores $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, donde

$$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}.$$

Para calcular las coordenadas del vector $\boldsymbol{e}_1$ con respecto a la base $\mathscr{B}$, tenemos que encontrar los escalares $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ tales que

$$\boldsymbol{e}_1 = \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \alpha_3\boldsymbol{u}_3.$$

Esto se traduce a resolver el sistema de ecuaciones $A\boldsymbol{\alpha} = \boldsymbol{e}_1$, donde

$$A = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{pmatrix}, \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Se tiene que

$$\begin{pmatrix} A & \boldsymbol{e}_1 \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{pmatrix},$$

por lo que

$$[\boldsymbol{e}_1]_{\mathscr{B}} = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix}.$$

De ahora en adelante, $\mathscr{S} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ denotará la base estándar de vectores, en el orden natural, para $\mathbb{R}^n$ o $\mathbb{C}^n$. Si no se hace mención de otra base, suponemos que estamos usando la base estándar.

Es importante hacer notar la diferencia entre dos expresiones. Dado un vector $\boldsymbol{v}$ de $\mathbb{K}^n$, tiene una expresión como n-upla, que coincide con sus coordenadas respecto de la base estándar. Por otro lado, fijada una base $\mathscr{B}$ de K^n, le corresponde otra n-upla que hemos notado como $[\boldsymbol{v}]_{\mathscr{B}}$, que son sus coordenadas respecto de la base $\mathscr{B}$.

Así, por ejemplo, si $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ es una base de un espacio vectorial V de dimensión finita n, las coordenadas de $\boldsymbol{u}_i$ respecto de la base $\mathscr{B}$ es la n-upla que tiene ceros en todas las posiciones salvo la i-ésima, que contiene el valor 1. Esto es lo que se conoce como el morfismo de coordenadas entre un espacio vectorial de dimensión finita V y $\mathbb{K}^n$.

Matriz de una aplicación lineal

Sean $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ y $\mathscr{B}' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ bases de U y V, respectivamente. La ***matriz de coordenadas*** de una aplicación lineal $T : U \to V$ con respecto al par $(\mathscr{B}, \mathscr{B}')$ es la matriz de orden $m \times n$

$$[T]_{\mathscr{B}\mathscr{B}'} = \left(\begin{array}{c|c|c|c} [T(\boldsymbol{u}_1)]_{\mathscr{B}'} & [T(\boldsymbol{u}_2)]_{\mathscr{B}'} & \ldots & [T(\boldsymbol{u}_n)]_{\mathscr{B}'} \end{array} \right).$$

En otras palabras, si $T(\boldsymbol{u}_j) = \alpha_{1j}\boldsymbol{v}_1 + \alpha_{2j}\boldsymbol{v}_2 + \cdots + \alpha_{mj}\boldsymbol{v}_m$, entonces

$$[T(\boldsymbol{u}_j)]_{\mathscr{B}'} = \begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{mj} \end{pmatrix} \text{ y } [T]_{\mathscr{B}\mathscr{B}'} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \ldots & \alpha_{mn} \end{pmatrix}.$$

Cuando T es un endomorfismo de U y una sola base implicada, usaremos la notación $[T]_{\mathscr{B}}$ en lugar de $[T]_{\mathscr{B}\mathscr{B}}$. Esta matriz será cuadrada.

Ejemplo 4.5.3. Consideremos la aplicación $f : \mathbb{R}^3 \to \mathbb{R}^3$ definida como

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}.$$

Vamos calcular la matriz de f respecto a diferentes bases en origen y destino. Por ejemplo, sean

$$\mathscr{B} = \left\{ \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

y

$$\mathscr{B}' = \left\{ \boldsymbol{v}_1 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}.$$

Para calcular $[f]_{\mathscr{B}\mathscr{B}'}$ procedemos como sigue:

$$f(\boldsymbol{u}_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = -1\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3 \Rightarrow [f(\boldsymbol{u}_1)]_{\mathscr{B}'} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix},$$

$$f(\boldsymbol{u}_2) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = -1\boldsymbol{v}_1 + 1\boldsymbol{v}_2 + 0\boldsymbol{v}_3 \Rightarrow [f(\boldsymbol{u}_2)]_{\mathscr{B}'} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix},$$

$$f(\boldsymbol{u}_3) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = -1\boldsymbol{v}_1 + 1\boldsymbol{v}_2 + 0\boldsymbol{v}_3 \Rightarrow [f(\boldsymbol{u}_1)]_{\mathscr{B}'} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Entonces

$$[f]_{\mathscr{B}\mathscr{B}'} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ahora consideramos la misma base en origen y destino. Tomemos la base

$$\mathscr{C} = \left\{ \boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \boldsymbol{w}_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}.$$

Tenemos que

$$f(\boldsymbol{w}_1) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 1\boldsymbol{w}_1 + 1\boldsymbol{w}_2 - 1\boldsymbol{w}_3 \Rightarrow [f(\boldsymbol{w}_1)]_{\mathscr{B}} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix},$$

$$f(\boldsymbol{w}_2) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = 0\boldsymbol{w}_1 + 3\boldsymbol{w}_2 - 2\boldsymbol{w}_3 \Rightarrow [f(\boldsymbol{w}_2)]_{\mathscr{B}} = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix},$$

$$f(\boldsymbol{w}_3) = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = 0\boldsymbol{w}_1 + 3\boldsymbol{w}_2 - 2\boldsymbol{w}_3 \Rightarrow [f(\boldsymbol{w}_3)]_{\mathscr{B}} = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix},$$

de donde

$$[f]_{\mathscr{C}} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 3 & 3 \\ -1 & -2 & -2 \end{pmatrix}.$$

Si consideramos la base estándar $\mathscr{S}$, es fácil ver que

$$[f]_{\mathscr{S}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

En el centro del álgebra lineal se encuentra la identificación entre la teoría de las transformaciones lineales y la teoría de matrices. Esto se debe al siguiente resultado, que expresa la acción de un operador lineal sobre un vector como como el producto de una matriz por un vector columna.

Imagen de un vector como producto

Sea $T : U \to V$ una aplicación lineal, y $\mathscr{B}$ y $\mathscr{B}'$ bases respectivas de U y V. Para cada $\boldsymbol{u} \in U$, la acción de T sobre $\boldsymbol{u}$ está dada por la multiplicación matricial

$$[T(\boldsymbol{u})]_{\mathscr{B}'} = [T]_{\mathscr{B}\mathscr{B}'}[\boldsymbol{u}]_{\mathscr{B}}.$$

PRUEBA: Sean $\mathscr{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ y $\mathscr{B}' = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. Si $\boldsymbol{u} = \sum_{j=1}^{n} \xi_j \boldsymbol{u}_j$, y $T(\boldsymbol{u}_j) = \sum_{i=1}^{m} \alpha_{ij} \boldsymbol{v}_j$, entonces

$$[\boldsymbol{u}]_{\mathscr{B}} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} \text{ y } [T]_{\mathscr{B}\mathscr{B}'} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \ldots & \alpha_{mn} \end{pmatrix}.$$

Podemos escribir

$$\begin{aligned} T(\boldsymbol{u}) = T(\sum_{j=1}^{n} \xi_j \boldsymbol{u}_j) &= \sum_{j=1}^{n} \xi_j T(\boldsymbol{u}_j) \\ &= \sum_{j=1}^{n} \xi_j \sum_{i=1}^{m} \alpha_{ij} \boldsymbol{v}_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_{ij} \xi_j \right) \boldsymbol{v}_i. \end{aligned}$$

En otras palabras, las coordenadas de $T(\boldsymbol{u})$ respecto a $\mathscr{B}'$ son los términos $\sum_{j=1}^{n} \alpha_{ij} \xi_j$ para $i = 1, 2, \ldots, m$. Por tanto,

$$[T(\boldsymbol{u})]_{\mathscr{B}'} = \begin{pmatrix} \sum_{j=1}^{n} \alpha_{1j} \xi_j \\ \sum_{j=1}^{n} \alpha_{2j} \xi_j \\ \vdots \\ \sum_{j=1}^{n} \alpha_{mj} \xi_j \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \ldots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix} = [T]_{\mathscr{B}\mathscr{B}'}[\boldsymbol{u}]_{\mathscr{B}}.$$

□

Ejemplo 4.5.4. Vamos a considerar de nuevo el ejemplo 4.5.3, con las bases

$$\mathscr{B} = \left\{ \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

y

$$\mathscr{B}' = \left\{ \boldsymbol{v}_1 = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}.$$

Partimos de la aplicación $f : \mathbb{R}^3 \to \mathbb{R}^3$ dada por

$$f\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}.$$

y del vector

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \in \mathbb{R}^3.$$

Nuestro objetivo es obtener $[f(\boldsymbol{u})]_{\mathscr{B}'}$ y lo vamos a hacer de dos formas.

Método 1. El vector $\boldsymbol{u}$ no lo están dando como elemento de $\mathbb{R}^3$, por lo que las componentes que lo definen coinciden con sus coordenadas respecto de la base estándar $\mathscr{S}$ de $\mathbb{R}^3$. Por otro lado, la definición de f también se basa en la expresión de los elementos de $\mathbb{R}^3$ en función de la base estándar. Por tanto, es inmediato que

$$f(\boldsymbol{u}) = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

son las coordenadas de $f(\boldsymbol{u})$ respecto de la base $\mathscr{S}$. Queremos calcular las coordenadas de este vector respecto de la base $\mathscr{B}'$. Para ello, tenemos que resolver el sistema

$$\begin{pmatrix} -1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Su solución es $x_1 = -1, x_2 = -1, x_3 = 0$, por lo que

$$[f(\boldsymbol{u})]_{\mathscr{B}'} = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}.$$

Método 2. Calculamos en el ejemplo anterior que

$$[f]_{\mathscr{B}\mathscr{B}'} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ahora necesitamos la coordenadas de $\boldsymbol{u}$ respecto de $\mathscr{B}$ para aplicar la fórmula que se ha demostrado en el teorema. Resolvemos el sistema

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

cuya solución es $x_1 = 2, x_2 = -2, x_3 = 1$. Entonces

$$[f(\boldsymbol{u})]_{\mathscr{B}'} = [f]_{\mathscr{B},\mathscr{B}'}[\boldsymbol{u}]_{\mathscr{B}},$$

$$[f(\boldsymbol{u})]_{\mathscr{B}'} = \begin{pmatrix} -1 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}.$$

La relación con las operaciones matriciales va más allá.

Conexiones con el álgebra matricial

- Si $T, L : U \to V$ son aplicaciones lineales, y $\mathscr{B}$ y $\mathscr{B}'$ son bases respectivas de U y V, entonces
  - $[\alpha T]_{\mathscr{B}\mathscr{B}'} = \alpha[T]_{\mathscr{B}\mathscr{B}'}$, para escalares α.
  - $[T + L]_{\mathscr{B}\mathscr{B}'} = [T]_{\mathscr{B}\mathscr{B}'} + [L]_{\mathscr{B}\mathscr{B}'}$.
- Si $T : U \to V$ y $L : V \to \mathscr{W}$, y $\mathscr{B}, \mathscr{B}'$ y $\mathscr{B}''$ son bases respectivas de U, V y $\mathscr{W}$, entonces $L \circ T$ es una aplicación lineal de U en $\mathscr{W}$ y

$$[L \circ T]_{\mathscr{B}\mathscr{B}''} = [L]_{\mathscr{B}'\mathscr{B}''}[T]_{\mathscr{B}\mathscr{B}'}.$$

- Si $T : U \to U$ es una aplicación lineal con inversa T^{-1}, entonces para toda base $\mathscr{B}$ de U se verifica que

$$[T^{-1}]_{\mathscr{B}} = [T]_{\mathscr{B}}^{-1}.$$

PRUEBA: Las tres primeras propiedades se deducen inmediatamente del resultado anterior. Por ejemplo, para calcular la matriz de la composición, sea $\boldsymbol{u}$ un vector arbitrario de U. Podemos escribir, por un lado,

$$[(L\circ T)(\boldsymbol{u})]_{\mathscr{B}''} = [L\circ T]_{\mathscr{B}\mathscr{B}''}[\boldsymbol{u}]_{\mathscr{B}},$$

y también

$$[(L\circ T)(\boldsymbol{u})]_{\mathscr{B}''} = [L(T(\boldsymbol{u}))]_{\mathscr{B}''} = [L]_{\mathscr{B}'\mathscr{B}''}[T(\boldsymbol{u})]_{\mathscr{B}'} = [L]_{\mathscr{B}'\mathscr{B}''}[T]_{\mathscr{B}\mathscr{B}'}[\boldsymbol{u}]_{\mathscr{B}}.$$

Por tanto, para todo $\boldsymbol{u}\in U$, se tiene la igualdad

$$[L\circ T]_{\mathscr{B}\mathscr{B}''}[\boldsymbol{u}]_{\mathscr{B}} = [L]_{\mathscr{B}'\mathscr{B}''}[T]_{\mathscr{B}\mathscr{B}'}[\boldsymbol{u}]_{\mathscr{B}},$$

lo que implica la igualdad de las matrices

$$[L\circ T]_{\mathscr{B}\mathscr{B}''} = [L]_{\mathscr{B}'\mathscr{B}''}[T]_{\mathscr{B}\mathscr{B}'}.$$

Para probar la referente a la inversa, observemos que si $\dim U = n$, entonces la matriz de la aplicación identidad respecto de cualquier base de U es la matriz identidad I_n. Entonces, por la propiedad de la composición,

$$I_n = [\mathrm{id}]_{\mathscr{B}} = [T\circ T^{-1}]_{\mathscr{B}} = [T]_{\mathscr{B}}[T^{-1}]_{\mathscr{B}},$$

y esto significa que $[T^{-1}]_{\mathscr{B}} = [T]_{\mathscr{B}}^{-1}$. □

Nota 4.5.5. Desde el punto de vista histórico, la composición de aplicaciones lineales es la que dio lugar a la definición de matrices y su producto. Por ejemplo, consideremos transformaciones geométricas lineales en el plano que dejen invariante el origen. Entonces son de la forma

$$\left\{\begin{array}{rcl} x' &=& ax+by,\\ y' &=& cx+dy,\end{array}\right.,\left\{\begin{array}{rcl} x'' &=& a'x'+b'y',\\ y'' &=& c'x'+d'y',\end{array}\right.$$

La composición se obtiene sustituyendo los valores de x', y' en la segunda transformación. Entonces queda

$$\left\{\begin{array}{rcl} x'' &=& a'(ax+by)+b'(cx+dy) = (a'a+b'c)x+(a'b+b'd)y,\\ y'' &=& c'(ax+by)+d'(cx+dy) = (c'a+d'c)x+(c'b+d'd)y.\end{array}\right.$$

Observemos que los coeficientes de esta transformación son los elementos de la matriz producto

$$\begin{pmatrix} a' & b'\\ c' & d'\end{pmatrix}\begin{pmatrix} a & b\\ c & d\end{pmatrix}.$$

Ejemplo 4.5.6. Consideremos las aplicaciones lineales $T:\mathbb{R}^3\to\mathbb{R}^2$ y $L:\mathbb{R}^2\to\mathbb{R}^2$ definidas como

$$T\begin{pmatrix} x\\ y\\ z\end{pmatrix}=\begin{pmatrix} x+y\\ y-z\end{pmatrix}, L\begin{pmatrix} u\\ v\end{pmatrix}=\begin{pmatrix} 2u-v\\ u\end{pmatrix}.$$

La composición $C=L\circ T:\mathbb{R}^3\to\mathbb{R}^2$ es la transformación lineal

$$C\begin{pmatrix} x\\ y\\ z\end{pmatrix}=L\begin{pmatrix} x+y\\ y-z\end{pmatrix}=\begin{pmatrix} 2(x+y)-(y-z)\\ x+y\end{pmatrix}=\begin{pmatrix} 2x+y+z\\ x+y\end{pmatrix}.$$

En forma matricial, con respecto a las bases estándar de $\mathbb{R}^2$ y $\mathbb{R}^3$, nos queda

$$[C]_{\mathscr{S}_3\mathscr{S}_2}=\begin{pmatrix} 2 & 1 & 1\\ 1 & 1 & 0\end{pmatrix}, [L]_{\mathscr{S}_2}=\begin{pmatrix} 2 & -1\\ 1 & 0\end{pmatrix}, \text{ y } [T]_{\mathscr{S}_3\mathscr{S}_2}=\begin{pmatrix} 1 & 1 & 0\\ 0 & 1 & -1\end{pmatrix}.$$

Es fácil ver que

$$[C]_{\mathscr{S}_3\mathscr{S}_2}=[L]_{\mathscr{S}_2}[T]_{\mathscr{S}_3\mathscr{S}_2}.$$

Por otro lado,

$$[L^{-1}]_{\mathscr{S}_2}=[L]^{-1}_{\mathscr{S}_2}=\begin{pmatrix} 2 & -1\\ 1 & 0\end{pmatrix}^{-1}=\begin{pmatrix} 0 & 1\\ -1 & 2\end{pmatrix}.$$

4.6. Cambio de base

Por su propia naturaleza, la representación matricial de una aplicación lineal depende de las coordenadas. Sin embargo, existen propiedades de estos operadores que es conveniente estudiar respecto a unas bases especiales, y que permiten determinar características intrínsecas de los mismos que son independientes de la base elegida. En esta sección vamos a ver cómo se relacionan las matrices con un cambio de base. Nos centraremos en lo que le ocurre a la matriz de una endomorfismo.

Sean $\mathscr{B}=\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_n\}$ y $\mathscr{B}'=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_n\}$ bases de V. Definimos la aplicación $T:V\to V$ dada por $T(\boldsymbol{v}_i)=\boldsymbol{u}_i$. Es claro que T es invertible, pues basta considerar $T^{-1}(\boldsymbol{u}_i)=\boldsymbol{v}_i$.

Observemos que la aplicación identidad tiene como matriz respecto a estas bases

$$[\mathrm{id}]_{\mathscr{B}\mathscr{B}'}=\begin{pmatrix} [\boldsymbol{u}_1]_{\mathscr{B}'} & | & [\boldsymbol{u}_2]_{\mathscr{B}'} & | & \ldots & [\boldsymbol{u}_n]_{\mathscr{B}'}\end{pmatrix}=P(\mathscr{B},\mathscr{B}'),$$

que denominamos *matriz de paso* de $\mathscr{B}$ a $\mathscr{B}'$. Entonces, dado $\boldsymbol{v}\in V$,

$$[\boldsymbol{v}]_{\mathscr{B}'}=[\mathrm{id}(\boldsymbol{v})]_{\mathscr{B}'}=[\mathrm{id}]_{\mathscr{B}\mathscr{B}'}[\boldsymbol{v}]_{\mathscr{B}}=P(\mathscr{B},\mathscr{B}')[\boldsymbol{v}]_{\mathscr{B}}.$$

Esta expresión es la que relaciona las coordenadas del vector $\boldsymbol{v}$ con respecto a las bases $\mathscr{B}$ y $\mathscr{B}'$.

La matriz $P(\mathscr{B},\mathscr{B}')$ es no singular, pues

$$P(\mathscr{B},\mathscr{B}') = [T]_{\mathscr{B}'},$$

y esta aplicación es invertible. Además,

$$P(\mathscr{B},\mathscr{B}')^{-1} = P(\mathscr{B}',\mathscr{B}).$$

Podemos resumir lo anterior en el siguiente recuadro.

Ecuaciones del cambio de base

Sean $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ y $\mathscr{B}' = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ bases de V, y consideremos la aplicación $T : V \to V$ definida por $T(\boldsymbol{v}_i) = \boldsymbol{u}_i$. Entonces

$$P(\mathscr{B},\mathscr{B}') = [T]_{\mathscr{B}'} = \left(\begin{array}{ccccc} [\boldsymbol{u}_1]_{\mathscr{B}'} & | & [\boldsymbol{u}_2]_{\mathscr{B}'} & | & \ldots \quad [\boldsymbol{u}_n]_{\mathscr{B}'} \end{array} \right)$$

se denomina ***matriz de paso*** de $\mathscr{B}$ a $\mathscr{B}'$, y verifica que

- $[\boldsymbol{v}]_{\mathscr{B}'} = P(\mathscr{B},\mathscr{B}')[\boldsymbol{v}]_{\mathscr{B}}$.
- $P(\mathscr{B},\mathscr{B}')$ es no singular y

$$(P(\mathscr{B},\mathscr{B}'))^{-1} = P(\mathscr{B}',\mathscr{B}).$$

Ejemplo 4.6.1. Consideremos en $V = \mathbb{R}^3$ las bases

$$\mathscr{B} = \{\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\},$$

$$\mathscr{B}' = \{\boldsymbol{u}'_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \boldsymbol{u}'_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \boldsymbol{u}'_3 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}\}$$

Vamos a calcular la matriz de paso de $\mathscr{B}'$ a $\mathscr{B}$. Sabemos que tiene la forma

$$M(\mathscr{B}',\mathscr{B}) = \left(\begin{array}{ccc} [\boldsymbol{u}'_1]_{\mathscr{B}} & [\boldsymbol{u}'_2]_{\mathscr{B}} & [\boldsymbol{u}'_3]_{\mathscr{B}} \end{array} \right),$$

por lo que tenemos que calcular las coordenadas de cada vector $\boldsymbol{u}'_i, i = 1,2,3$ con respecto a la base $\mathscr{B}$. Esto es equivalente a encontrar los escalares $\alpha_i, \beta_i, \gamma_1, i = 1,2,3$ tales que

$$\begin{aligned}\boldsymbol{u}'_1 &= \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \alpha_3\boldsymbol{u}_3,\\ \boldsymbol{u}'_2 &= \beta_1\boldsymbol{u}_1 + \beta_2\boldsymbol{u}_2 + \beta_3\boldsymbol{u}_3,\\ \boldsymbol{u}'_3 &= \gamma_1\boldsymbol{u}_1 + \gamma_2\boldsymbol{u}_2 + \gamma_3\boldsymbol{u}_3,\end{aligned}$$

que implica la resolución de tres sistemas de ecuaciones:

$$\begin{bmatrix} 1\\ -1\\ 1\end{bmatrix} = \begin{bmatrix} 1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix},$$

$$\begin{bmatrix} 1\\ 2\\ 2\end{bmatrix} = \begin{bmatrix} 1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix},$$

$$\begin{bmatrix} 1\\ 3\\ 3\end{bmatrix} = \begin{bmatrix} 1 & 1 & 1\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\begin{pmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\end{pmatrix}$$

Todos tienen la misma matriz de coeficientes, por lo que podemos resolverlos de forma simultánea:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 1 & 1\\ 0 & 1 & 1 & -1 & 2 & 3\\ 0 & 0 & 1 & 1 & 2 & 3\end{array}\right] \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 2 & -1 & -2\\ 0 & 1 & 0 & -2 & 0 & 0\\ 0 & 0 & 1 & 1 & 2 & 3\end{bmatrix}.$$

Por tanto, la matriz del cambio de base de $\mathscr{B}'$ a $\mathscr{B}$ o matriz de paso es

$$M(\mathscr{B}',\mathscr{B}) = \begin{bmatrix} 2 & -1 & -2\\ -2 & 0 & 0\\ 1 & 2 & 3\end{bmatrix}.$$

Podríamos aplicar un procedimiento similar para el cálculo de la matriz del cambio de base de $\mathscr{B}$ a $\mathscr{B}'$, o bien recordar que es la inversa de la matriz anterior:

$$M(\mathscr{B},\mathscr{B}') = M(\mathscr{B}',\mathscr{B})^{-1} = \begin{bmatrix} 0 & -1/2 & 0\\ 3 & 4 & 2\\ -2 & -5/2 & -1\end{bmatrix}.$$

Ejemplo 4.6.2. Sea A un endomorfismo de V, y sean $\mathscr{B}$ y $\mathscr{B}'$ dos bases de V. Vamos a estudiar la relación entre $[A]_{\mathscr{B}}$ y $[A]_{\mathscr{B}'}$. Para cualquier vector $\boldsymbol{v} \in V$ se tiene que

$$[A(\boldsymbol{v})]_{\mathscr{B}} = [A]_{\mathscr{B}}[\boldsymbol{v}]_{\mathscr{B}}, [A(\boldsymbol{v})]_{\mathscr{B}'} = [A]_{\mathscr{B}'}[\boldsymbol{v}]_{\mathscr{B}'},$$

y para cualquier vector $\boldsymbol{w} \in V$ sabemos que

$$[\boldsymbol{w}]_{\mathscr{B}} = P(\mathscr{B}', \mathscr{B})[\boldsymbol{w}]_{\mathscr{B}'}.$$

Aplicamos esto a los vectores $\boldsymbol{v}$ y $A\boldsymbol{v}$. Entonces

$$[A(\boldsymbol{v})]_{\mathscr{B}} = P(\mathscr{B}', \mathscr{B})[A\boldsymbol{v}]_{\mathscr{B}'} = P(\mathscr{B}', \mathscr{B})[A]_{\mathscr{B}'}[\boldsymbol{v}]_{\mathscr{B}'} = P(\mathscr{B}', \mathscr{B})[A]_{\mathscr{B}'}P(\mathscr{B}, \mathscr{B}')[\boldsymbol{v}]_{\mathscr{B}}.$$

Como esto es cierto para todo vector $\boldsymbol{v}$, se sigue que

$$[A]_{\mathscr{B}} = P(\mathscr{B}', \mathscr{B})[A]_{\mathscr{B}'}P(\mathscr{B}, \mathscr{B}') = P^{-1}[A]_{\mathscr{B}'}P,$$

donde $P = P(\mathscr{B}, \mathscr{B}')$ es la matriz del cambio de base de $\mathscr{B}$ a $\mathscr{B}'$.

Semejanza de matrices

Dos matrices $B_{n\times n}$ y $C_{n\times n}$ se dicen ***semejantes*** cuando existe una matriz Q no singular tal que $B = Q^{-1}CQ$.

Lo anterior indica que las matrices de un endomorfismo respecto a diferentes bases son semejantes. Recíprocamente, si dos matrices son semejantes, entonces son las matrices de una aplicación lineal respecto a diferentes bases. Por tanto, *matrices semejantes representan el mismo endomorfismo.*

Ejemplo 4.6.3. Consideremos el operador lineal $f : \mathbb{R}^2 \to \mathbb{R}^2$ definido como

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ -2x+3y \end{pmatrix},$$

junto a las bases

$$\mathscr{S} = \left\{\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}, \mathscr{B}' = \left\{\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right\}.$$

La matriz $[f]_{\mathscr{S}}$ se obtiene a partir de

$$f(\boldsymbol{e}_1) = \begin{pmatrix} 0 \\ -2 \end{pmatrix} = 0\boldsymbol{e}_1 + (-2)\boldsymbol{e}_2,$$

$$f(\boldsymbol{e}_2) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 1\boldsymbol{e}_1 + 3\boldsymbol{e}_2,$$

luego

$$[f]_{\mathscr{S}} = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}.$$

La matriz del cambio de base es

$$P = P(\mathscr{B}', \mathscr{S}) = \left(\begin{array}{c|c} [\boldsymbol{v}_1]_{\mathscr{S}} & [\boldsymbol{v}_2]_{\mathscr{S}} \end{array} \right) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

y si aplicamos la fórmula que relaciona la matriz de la aplicación f con respecto a cada base, obtenemos

$$\begin{aligned} [f]_{\mathscr{B}'} &= P(\mathscr{S}, \mathscr{B}')[f]_{\mathscr{S}} P(\mathscr{B}', \mathscr{S}) = P^{-1}[f]_{\mathscr{S}} P \\ &= \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

La nueva matriz es diagonal, mientras que la original no lo es. Esto muestra que no siempre la base estándar es la mejor elección para dar la representación matricial. Más adelante estudiaremos cómo buscar bases respecto de las cuales la matriz de una aplicación lineal sea lo más sencilla posible.

Capítulo 5

Determinantes

Al comienzo del curso hacíamos referencia al antiguo tablero chino para contar, en el que cañas de bambú coloreadas se manipulaban de acuerdo a ciertas reglas para resolver un sistema de ecuaciones lineales. El tablero chino parece que se usaba por el 200 a.C., y mantuvo su mecanismo durante un milenio. El tablero y las reglas para usarlo llegaron a Japón, donde Seki Kowa (1642-1708), un gran matemático japonés, sintetizó las antiguas ideas chinas de manipulación de rectángulos. Kowa formuló el concepto de lo que hoy llamamos determinante para facilitar la resolución de sistemas lineales. Se piensa que su definición data de poco antes de 1683.

Alrededor de los mismos años, entre 1678 y 1693, Gottfried W. Leibniz (1646-1716), un matemático alemán, desarrollaba su propio concepto de determinante de forma independiente, junto con aplicaciones de manipulación de rectángulos de números para resolver sistemas de ecuaciones lineales. El trabajo de Leibniz solamente trata sistemas de tres ecuaciones con tres incógnitas, mientras que Seki Kowa dio un tratamiento general para sistemas de n ecuaciones con n incógnitas. Parece que tanto Kowa como Leibniz desarrollaron lo que se llamó posteriormente regla de Cramer, pero no en la misma forma ni notación. Estos dos hombres tuvieron algo en común: sus ideas sobre la resolución de sistemas lineales nunca fueron adoptadas por la comunidad matemática de su tiempo, y sus descubrimientos se desvanecieron rápidamente en el olvido. Al final, el concepto de determinante se redescubrió, y la materia ha sido intensamente tratada en el periodo de 1750 a 1900. Durante el mismo, los determinantes se convirtieron en la mayor herramienta usada para analizar y resolver sistemas lineales, mientras que la teoría de matrices permanecía relativamente poco desarrollada. Pero las matemáticas, como un río, están siempre cambiando su curso, y grandes afluentes se pueden secar y convertirse en riachuelos, mientras que pequeños arroyuelos se convierten en poderosos torrentes. Esto es precisamente lo que ocurrió con las matrices y determinantes. El estudio y

uso de los determinantes llevó al álgebra matricial de Cayley, y hoy las matrices y el álgebra lineal están en la corriente principal de la matemática aplicada, mientras que el papel de los determinantes ha sido relegado a una zona de remanso, por seguir con la analogía fluvial. Sin embargo, todavía es importante comprender qué es un determinante y aprender sus propiedades fundamentales. Nuestro objetivo no es aprender determinantes por su propio interés, sino que exploraremos aquellas propiedades que son útiles en el posterior desarrollo de la teoría de matrices y sus aplicaciones.

Existen varias aproximaciones para la definición del determinante de una matriz, de las que destacamos dos vías. La primera es de tipo inductivo, en donde el determinante de una matriz de orden n se calcula a partir de los determinantes de matrices de orden $n-1$. La segunda se basa en una definición directa, que precisa el concepto de permutación y de su clasificación en tipo par o impar. Desarrollaremos ambas, con un teorema de unicidad que garantiza la igualdad de ambas definiciones. La elección de una forma u otra depende del aspecto teórico en el que se quiera insistir.

5.1. Definición inductiva

Sea A una matriz de orden n con coeficientes en un anillo K. Habitualmente usaremos un cuerpo K, pero también será necesario trabajar con matrices cuyos coeficientes son polinomios. Si $n = 1$, entonces $A = (a_{11})$ y definimos su determinante como $\det(A) = a_{11}$. Si $n > 1$ y $A = (a_{ij})$, llamamos **cofactor** del elemento a_{ij} al número $\hat{A}_{ij} = (-1)^{i+j} \det(A_{ij})$, donde A_{ij} es la matriz de orden $n-1$ que se obtiene al eliminar la fila i-ésima y la columna j-ésima de la matriz A. Se define entonces la función determinante como $\det : \mathscr{M}_{n\times n}(K) \to K$ dada por

$$\det(A) = a_{11}\hat{A}_{11} + a_{21}\hat{A}_{21} + \cdots + a_{n1}\hat{A}_{n1},$$

es decir, la suma de los productos de cada elemento de la primera columna por su cofactor correspondiente.

Determinante: forma inductiva

Sea $A = (a_{ij})$ una matriz cuadrada de orden n.

- Si $n = 1$, entonces $\det(A) = a_{11}$.
- Si $n > 1$, entonces

$$\det(A) = a_{11}\hat{A}_{11} + a_{21}\hat{A}_{21} + \cdots + a_{n1}\hat{A}_{n1},$$

donde $\hat{A}_{ij} = (-1)^{i+j}\det(A_{ij})$ es el cofactor del elemento a_{ij}.

Ejemplo 5.1.1. 1. Para una matriz 2×2 se tiene

$$\begin{aligned}\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} &= a_{11}\hat{A}_{11} + a_{21}\hat{A}_{21} \\ &= a_{11}a_{22} - a_{21}a_{12}.\end{aligned}$$

2. Para la matriz identidad,

$$\det(I_n) = 1 \cdot \det(I_{n-1}),$$

y por inducción $\det(I_n) = 1$.

3. Para una matriz 3×3 se tiene

$$\begin{aligned}\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} &= a_{11}\hat{A}_{11} + a_{21}\hat{A}_{21} + a_{31}\hat{A}_{31} \\ &= a_{11}\det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21}\det\begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} \\ &\quad + a_{31}\det\begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{21}(a_{12}a_{33} + a_{32}a_{13}) \\ &\quad + a_{31}(a_{12}a_{23} - a_{22}a_{13}) \\ &= a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} \\ &\quad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.\end{aligned}$$

Esta es la que se conoce como **regla de Sarrus**, que admite la siguiente representación gráfica para los sumandos positivos y negativos:

4. El determinante de una matriz triangular superior es el producto de sus elementos diagonales.

Ejemplo 5.1.2.

$$\begin{aligned}\det\begin{pmatrix}1&2&3&4\\0&1&2&4\\2&3&4&5\\2&0&1&2\end{pmatrix} &= 1\cdot\det\begin{pmatrix}1&2&4\\3&4&5\\0&1&2\end{pmatrix} - 0\cdot\det\begin{pmatrix}2&3&4\\3&4&5\\0&1&2\end{pmatrix}\\ &\quad + 2\cdot\det\begin{pmatrix}2&3&4\\1&2&4\\0&1&2\end{pmatrix} - 2\cdot\det\begin{pmatrix}2&3&4\\1&2&4\\3&4&5\end{pmatrix}\\ &= 3-4-2=-3.\end{aligned}$$

De los ejemplos anteriores, observamos que esta forma de calcular el valor de un determinante es muy costoso en cuanto al número de operaciones, pues una matriz de orden n contiene $n!$ sumandos, cada uno de ellos con un producto de n elementos. Por ellos, vamos a estudiar qué propiedades posee esta función que permita su evaluación de una forma más simple.

5.2. Propiedades

Lineal en las filas

La aplicación determinante es lineal en las filas.

PRUEBA:

- Sea A una matriz de orden n, cuya fila A_{i*} se expresa como suma de dos vectores fila $B_{i*}+C_{i*}$. Debemos probar que

$$\det\begin{pmatrix}A_{1*}\\ \vdots\\ B_{i*}+C_{i*}\\ \vdots\\ A_{n*}\end{pmatrix} = \det\begin{pmatrix}A_{1*}\\ \vdots\\ B_{i*}\\ \vdots\\ A_{n*}\end{pmatrix} + \det\begin{pmatrix}A_{1*}\\ \vdots\\ C_{i*}\\ \vdots\\ A_{n*}\end{pmatrix}.$$

  La prueba es por inducción sobre n. Para $n=1$ no haya nada que probar. Si $n>1$, en la primera propiedad, sea A la matriz inicial, y B y C las matrices que aparecen a la derecha de la igualdad. Llamemos $a_{ij}=b_{ij}+c_{ij}$

los elementos de la fila i-ésima de la matriz A. Es fácil ver que el cofactor de a_{i1} en la matriz A es igual a los cofactores de los elementos b_{i1} y c_{i1} en las matrices B y C, respectivamente: $\hat{A}_{i1} = \hat{B}_{i1} = \hat{C}_{i1}$. Tenemos entonces

$$\det(A) = a_{11}\hat{A}_{11} + \cdots + (b_{i1} + c_{i1})\hat{A}_{i1} + \cdots + a_{n1}\hat{A}_{n1}.$$

Cada uno de los cofactores $\hat{A}_{t1}$, con $t \neq i$ es un determinante de orden $n-1$ y una fila expresada como suma de otras dos. Entonces, por hipótesis de inducción, $\hat{A}_{t1} = \hat{B}_{t1} + \hat{C}_{t1}$ para cada $t \neq i$. Obtenemos así que

$$\begin{aligned}\det(A) &= a_{11}(\hat{B}_{11} + \hat{C}_{11}) + \cdots + (b_{i1} + c_{i1})\hat{A}_{i1} + \cdots + a_{n1}(\hat{B}_{n1} + \hat{C}_{n1}) \\ &= a_{11}\hat{B}_{11} + \cdots + b_{i1}\hat{B}_{i1} + \cdots + a_{n1}\hat{B}_{n1} \\ &+ a_{11}\hat{C}_{11} + \cdots + c_{i1}\hat{C}_{i1} + \cdots + a_{n1}\hat{C}_{n1} \\ &= \det(B) + \det(C).\end{aligned}$$

- Sea A una matriz cuya fila A_{i*} es el producto de un escalar por un vector fila kB_{i*}. Hay que probar que

$$\det\begin{pmatrix} A_{1*} \\ \vdots \\ kB_{i*} \\ \vdots \\ A_{n*} \end{pmatrix} = k\det\begin{pmatrix} A_{1*} \\ \vdots \\ B_{i*} \\ \vdots \\ A_{n*} \end{pmatrix}.$$

Aplicamos inducción sobre n. Si $n = 1$, es trivial. Si $n > 1$, llamemos C a la matriz que aparece a la derecha de la igualdad. Entonces $a_{ij} = kb_{ij}$ y el cofactor del elemento a_{i1} es igual al cofactor del elemento c_{i1}. Para $t \neq i$, el cofactor del elemento a_{t1} es $k\hat{A}_{t1}$, por la hipótesis de inducción. Entonces

$$\begin{aligned}\det(A) &= a_{11}\hat{A}_{11} + \cdots + kb_{i1}\hat{A}_{i1} + \cdots + a_{n1}\hat{A}_{n1} \\ &= a_{11}k\hat{A}_{11} + \cdots + kb_{i1}\hat{A}_{i1} + \cdots + a_{n1}\hat{A}_{n1} \\ &= k\det(C).\end{aligned}$$

□

Una consecuencia inmediata de lo anterior es que si A tiene una fila de ceros, entonces $\det(A) = 0$; supongamos que la fila A_{i*} contiene únicamente valores nulos. En tal caso, podemos sacar el factor cero y tenemos el resultado.

Alternada

Si A es una matriz con dos filas iguales, entonces $\det(A) = 0$. Si B es la matriz que resulta de intercambiar dos filas de la matriz A, entonces $\det(B) = -\det(A)$.

PRUEBA:

- **Paso 1**. Las filas iguales de la matriz A son consecutivas. Hacemos la prueba por inducción. Para $n = 2$ es una sencilla comprobación. Supongamos entonces $n > 2$. Si las filas iguales son A_{i*} y A_{i+1*}, entonces $a_{i1} = a_{i+1,1}$ y

$$\det(A) = a_{11}\hat{A}_{11} + \cdots + a_{i1}\hat{A}_{i1} + a_{i1}\hat{A}_{i+1,1} + \cdots + a_{n1}\hat{A}_{n1}.$$

  Para $t \neq i$, el cofactor $\hat{A}_{ti}$ corresponde a una matriz de orden $n-1$ con dos filas iguales, por lo que $\hat{A}_{ti} = 0$ para $t \neq i$. Por otro lado,

$$\hat{A}_{i1} = (-1)^{i+1}\det(A_{i1}),\ \hat{A}_{i+1,1} = (-1)^{i+2}\det(A_{i+1,1}),$$

  y las matrices A_{i1} y $A_{i+1,1}$ son iguales. En consecuencia, $\hat{A}_{i1} = -\hat{A}_{i+1,1}$ y $\det(A) = 0$.

- **Paso 2**. Intercambio de dos filas consecutivas. Por el paso 1, sabemos que

$$\begin{aligned} 0 &= \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i*} + A_{i+1,*} \\ A_{i*} + A_{i+1,*} \\ \vdots \\ A_{n*} \end{pmatrix} \\ &= \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i*} \\ A_{i*} \\ \vdots \\ A_{n*} \end{pmatrix} + \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i*} \\ A_{i+1,*} \\ \vdots \\ A_{n*} \end{pmatrix} + \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i+1,*} \\ A_{i+1,*} \\ \vdots \\ A_{n*} \end{pmatrix} + \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i+1,*} \\ A_{i*} \\ \vdots \\ A_{n*} \end{pmatrix} \\ &= 0 + \det(A) + 0 + \det(B), \end{aligned}$$

  donde B es la matriz que se obtiene a partir de A por el intercambio de la fila i con la fila $i+1$. Por tanto, $\det(B) = -\det(A)$.

- **Paso 3**. Las filas iguales de la matriz A no son necesariamente consecutivas. Supongamos que la matriz A tiene iguales las filas A_{i*} y A_{t*}. Por el paso 2, podemos realizar intercambios de filas hasta que sean consecutivas. Cada intercambio supone una alteración del signo del determinante, pero el resultado final es una matriz con determinante nulo, por el paso 1.

- **Paso 4**. Intercambio de dos filas. Se procede de igual manera que en el paso 2.

□

Una consecuencia de lo anterior es que el determinante de una matriz A no cambia si una fila A_{t*} es sustituida por $A_{t*}+kA_{i*}$, donde k es un escalar y $t \neq i$. En efecto,

$$\det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{t*}+kA_{i*} \\ \vdots \\ A_{i*} \\ \vdots \\ A_{n*} \end{pmatrix} = \det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{t*} \\ \vdots \\ A_{i*} \\ A_{n*} \end{pmatrix} + \det\begin{pmatrix} A_{1*} \\ \vdots \\ kA_{i*} \\ A_{i*} \\ \vdots \\ A_{n*} \end{pmatrix}$$

$$= \det(A) + k\det\begin{pmatrix} A_{1*} \\ \vdots \\ A_{i*} \\ \vdots \\ A_{i*} \\ \vdots \\ A_{n*} \end{pmatrix} = \det(A),$$

pues la última matriz tiene dos filas iguales.

Estas propiedades se pueden expresar en términos de las acciones de las matrices elementales.

Efecto de las transformaciones elementales

Sea A una matriz de orden n. Entonces

- $\det(P_{ij}A) = -\det(A)$ si P_{ij} es una matriz de permutación.
- $\det(T_i(\alpha)A) = \alpha\det(A)$ si $\alpha \in K$.
- $\det(T_{ij}(\alpha)A) = \det(A)$ si $\alpha \in K$.

Si tomamos $A = I_n$, entonces

- $\det(P_{ij}) = -1$.
- $\det(T_i(\alpha)) = \alpha$.
- $\det(T_{ij}(\alpha)) = 1$.

En resumen, si E es una matriz elemental, entonces $\det(EA) = \det(E)\det(A)$.

Cuando hablamos del cálculo de la inversa de una matriz, decíamos que las matrices que la poseen se denominan no singulares. Esta nomenclatura es clásica y está asociada al valor del determinante de la matriz.

Existencia de inversa y determinante

Sea K un cuerpo y $A \in \mathcal{M}_{n\times n}(K)$. Entonces A tiene inversa si y solamente si $\det(A) \neq 0$.

PRUEBA: Supongamos que A tiene inversa. Entonces A es producto de matrices elementales $A = E_1 \cdot E_k$ y

$$\det(A) = \det(E_1 \cdots E_{k-1}E_k) = \det(E_1 \cdots E_{k-1})\det(E_k) = \cdots = \det(E_1)\cdots\det(E_k) \neq 0,$$

pues las matrices elementales tienen determinante no nulo. Recíprocamente, supongamos que A no tiene inversa. En tal caso, su forma escalonada reducida por filas E_A tiene una fila de ceros, y $E_A = E_1 \cdots E_k A$, con E_i matrices elementales. Entonces

$$0 = \det(E_A) = \det(E_1 \cdots E_k)\det(A).$$

La matriz $E_1 \cdots E_k$ tiene inversa, por lo que su determinante es no nulo, de donde $\det(A) = 0$. □

Las matrices con determinante nulo se denominan **singulares**. Supongamos que A es una matriz cuadrada de orden n singular. Entonces A no tiene inversa, por lo que rango$(A) < n$. Esto implica que una columna de A se expresa como combinación lineal de las restantes.

Otra importante propiedad de la función determinante es su relación con el producto de matrices cuadradas.

Producto de matrices y determinante

- Si A y B son matrices cuadradas, entonces $\det(AB) = \det(A)\det(B)$.
- $\det\begin{pmatrix} A & B \\ \mathbf{0} & D \end{pmatrix} = \det(A)\det(D)$ si A y D son cuadradas.

PRUEBA:

- Supongamos que A es singular. Entonces $\det(A) = 0$, y además sabemos que $A = E_1 \cdots E_k E_A$, donde $E_1, \ldots, E_k$ son matrices elementales y E_A es la forma escalonada reducida por filas de A, que tiene una fila de ceros (al menos la última). Pero en este caso la última fila de $E_A B$ también es de ceros, luego $\det(E_A B) = 0$. Por tanto

$$\begin{aligned}\det(AB) &= \det(E_1 \cdots E_k E_A B) = \det(E_1) \cdots \det(E_k)\det(E_A B) = 0 \\ &= \det(A)\det(B).\end{aligned}$$

  Si A es no singular, entonces $A = E_1 E_2 \ldots E_k$ es producto de matrices elementales. Sabemos que la regla del producto es válida para estas matrices, por lo que

$$\begin{aligned}\det(AB) &= \det(E_1 E_2 \ldots E_k B) = \det(E_1)\det(E_2)\ldots\det(E_k)\det(B) \\ &= \det(E_1 E_2 \ldots E_k)\det(B) = \det(A)\det(B).\end{aligned}$$

- Veamos ahora la segunda parte. Existen P_A y P_D matrices no singulares, producto de elementales, tales que $A = P_A E_A$ y $D = P_D E_D$, las respectivas formas escalonadas reducidas por filas. Tenemos la siguiente identidad:

$$\begin{pmatrix} A & B \\ \mathbf{0} & D \end{pmatrix} = \begin{pmatrix} P_A & \mathbf{0} \\ \mathbf{0} & P_D \end{pmatrix} \begin{pmatrix} E_A & P_A^{-1} B \\ \mathbf{0} & E_D \end{pmatrix}.$$

Si A es singular, entonces E_A contiene columnas proporcionales, y la segunda matriz de la derecha también, por lo que es singular, y

$$\det\begin{pmatrix} A & B \\ \mathbf{0} & D \end{pmatrix} = 0 = \det(A)\det(D).$$

Análogamente, si D es singular, la segunda matriz de la derecha contiene ahora una fila de ceros, y entonces también es singular. Si ninguna es singular, entonces $E_A = I_r, E_D = I_s$, y $\det(A) = \det(P_A), \det(D) = \det(P_D)$, con

$$\det\begin{pmatrix} A & B \\ \mathbf{0} & D \end{pmatrix} = \det\begin{pmatrix} P_A & \mathbf{0} \\ \mathbf{0} & P_D \end{pmatrix}\det\begin{pmatrix} I_r & P_A^{-1}B \\ \mathbf{0} & I_s \end{pmatrix}.$$

La segunda matriz de la derecha es triangular, con determinante igual a 1. Si $P_A = E_1 \dots E_k$ y $P_D = F_1 \dots F_l$, la primera matriz de la derecha se puede escribir como

$$\begin{pmatrix} P_A & \mathbf{0} \\ \mathbf{0} & P_D \end{pmatrix} = \begin{pmatrix} E_1 & \\ & I \end{pmatrix} \cdots \begin{pmatrix} E_k & \\ & I \end{pmatrix} \begin{pmatrix} I & \\ & F_1 \end{pmatrix} \cdots \begin{pmatrix} I & \\ & F_l \end{pmatrix}.$$

Todos estos factores son matrices elementales, y se puede aplicar la regla del producto. Por tanto, tenemos el resultado.

□

Por último, el tratamiento que hemos hecho por filas se puede hacer por columnas.

Matriz traspuesta y determinante

Sea A una matriz cuadrada. Entonces $\det(A) = \det(A^t)$.

PRUEBA: Si A es singular, entonces A^t es singular, de donde $\det(A) = 0 = \det(A^t)$. Si A no es singular, entonces podemos expresar A como producto de matrices elementales $A = E_1 \cdots E_k$, de donde $A^t = E_k^t \cdots E_1^t$ es producto de matrices elementales. Para cada matriz elemental se tiene que $\det(E_i) = \det(E_i^t)$, por lo que $\det(A) = \det(A^t)$. □

El resultado anterior permite extender las propiedades anteriores a las columnas de una matriz.

Extensión de propiedades a las columnas

- La aplicación determinante es lineal en las columnas.
- Si A es una matriz con dos columnas iguales, entonces $\det(A) = 0$. Si B es la matriz que resulta de intercambiar dos columnas de la matriz A, entonces $\det(B) = -\det(A)$.
- $\det(AP_{ij}) = -\det(A)$ si P_{ij} es una matriz de permutación.
- $\det(AT_i(\alpha)) = \alpha \det(A)$ si $\alpha \in K$.
- $\det(AT_{ij}(\alpha)) = \det(A)$ si $\alpha \in K$.
- El determinante de una matriz se puede obtener mediante el desarrollo por cualquiera de sus filas o columnas:

$$\begin{aligned}\det(A) &= a_{i1}\hat{A}_{i1} + a_{i2}\hat{A}_{i2} + \cdots + a_{in}\hat{A}_{in} \\ &= a_{1j}\hat{A}_{1j} + a_{2j}\hat{A}_{2j} + \cdots + a_{nj}\hat{A}_{nj}.\end{aligned}$$

PRUEBA: Solamente hay que hacer un pequeño razonamiento para la última propiedad. Si $A = (a_{ij})$ y B es la matriz que se obtiene al intercambiar la primera columna con la columna j-ésima, entonces

$$B = \begin{pmatrix} a_{1j} & a_{12} & \cdots & a_{11} & \cdots & a_{1n} \\ a_{2j} & a_{22} & \cdots & a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ a_{nj} & a_{n2} & cdots & a_{n1} & \cdots & a_{nn} \end{pmatrix} \text{ y } \det(B) = -\det(A).$$

Por la definición,

$$\det(B) = a_{1j}\hat{B}_{11} + a_{2j}\hat{B}_{21} + \cdots + a_{nj}\hat{B}_{n1},$$

y los cofactores de B verifican que $\hat{B}_{k1} = -\hat{A}_{kj}$ para todo $j = 1, \ldots, n$. Por tanto,

$$\det(B) = -a_{1j}\hat{A}_{1j} - a_{2j}\hat{A}_{2j} - \cdots - a_{nj}\hat{A}_{nj},$$

y al igualar con $-\det(A)$ tenemos el resultado. Para las filas, usamos la igualdad del determinante con el de la matriz traspuesta. □

Nota 5.2.1. Habíamos visto que si $\det(A) = 0$, entonces una columna de A es combinación lineal de las restantes. De forma análoga, si $\det(A) = 0$, también una fila de A es combinación lineal de las restantes.

5.3. Rango y determinantes

En esta sección veremos cómo el rango de una matriz cualquiera puede calcularse usando determinantes. Para ello necesitamos las siguientes definiciones

Submatrices

Sea A una matriz de orden $m \times n$ y elegimos un subconjunto $I \subset \{1,\dots,m\}$, y un subconjunto $J \in \{1,\dots,n\}$. Se defina la **submatriz** de A determinada por I y J como la matriz $[A]_{I,J}$ cuyas entradas son las entradas $[A]_{ij}$ de A tales que $i \in I$ y $j \in J$. En otras palabras, es la matriz que se obtiene de A al eliminar las filas que no están en I, y las columnas que no están en J.

Ejemplo 5.3.1.

1. En la definición de cofactor asociado al elemento a_{ij} de una matriz cuadrada A tomábamos la submatriz que se obtenía al eliminar la fila i y la columna j de la matriz, que se obtiene con $I = \{1,\dots,n\} - \{i\}, J = \{1,\dots,n\} - \{j\}$.

2. En la matriz
$$A = \begin{pmatrix} 1 & -1 & 0 & 2 \\ 2 & -1 & 1 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$
la submatriz de A que se obtiene para $I = \{1,3\}, J = \{3,4\}$ es
$$[A]_{I,J} = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

Menores

Dada una matriz A, $m \times n$, los ***menores*** de A son los determinantes de las submatrices cuadradas de A. Es decir, los escalares
$$m_{I,J} = \det([A]_{I,J}),$$
donde $I \subset \{1,\dots,m\}$ y $J \subset \{1,\dots,n\}$ son dos conjuntos con el mismo número de elementos.

Diremos que $m_{I,J}$ es un menor de **orden** s si la submatriz $[A]_{I,J}$ tiene orden $s \times s$, es decir, si I y J tienen s elementos.

Ejemplo 5.3.2. En la matriz

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

para $I = \{1,3\}, J = \{3,4\}$ se obtiene el menor de orden 2

$$\det \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} = -2.$$

Recordemos que el rango de una matriz A es igual al número de columnas básicas, o al número de filas distintas de cero de cualquier forma escalonada, en particular de su forma escalonada reducida por filas. Con respecto a las submatrices, el rango se comporta de la siguiente manera:

Rango y submatrices

Si $[A]_{I,J}$ es una submatriz de A, entonces $\text{rango}([A]_{I,J}) \leq \text{rango}(A)$.

PRUEBA: Supongamos que A es una matriz de orden $m \times n$ y rango r. Sabemos que una serie de operaciones elementales por filas transforman A en E_A, su forma escalonada reducida por filas. Si llamamos $M = \{1,\ldots,m\}$, podemos considerar la submatriz $[A]_{M,J}$, que estará formada por algunas columnas de A. Si aplicamos a $[A]_{M,J}$ las mismas operaciones elementales, obtendremos las columnas correspondientes de la matriz E, es decir, obtendremos una matriz que tendrá a lo más r filas no nulas. Esto implica que la forma escalonada reducida por filas de $[A]_{M,J}$ tiene a lo más r filas no nulas, luego $\text{rango}(A_{M,J}) \leq r = \text{rango}(A)$.

Hemos probado entonces que, dada una matriz A de rango r, toda submatriz formada por columnas completas de A tiene a lo sumo rango r. Pero observemos que $([A]_{I,J})^t$ es una submatriz de $([A]_{M,J})^t$ formada por columnas completas, luego

$$\text{rango}([A]_{I,J}) = \text{rango}(([A]_{I,J})^t) \leq \text{rango}(([A]_{M,J})^t) = \text{rango}([A]_{M,J}) \leq \text{rango}(A),$$

como queríamos demostrar. □

Veamos que existe otra definición de rango, a partir de los menores de una matriz A.

Rango y menores

El rango de A es igual al mayor orden alcanzado por los menores no nulos de A. Es decir, rango$(A) = r$ si y sólo si:

- Existe un menor $m_{I,J} \neq 0$ de orden r.
- Todos los menores de A de orden mayor que r son nulos.

PRUEBA: Supongamos que A es una matriz de orden $m \times n$, que tiene rango r. Sabemos que hay unas operaciones elementales por filas que transforman A en su forma escalonada reducida por filas E.

Sean $p_1, \ldots, p_r$ las columnas básicas de A, y definamos $J = \{p_1, \ldots, p_r\}$ y $M = \{1, \ldots, m\}$. La submatriz $[A]_{M,J}$ está formada por las columnas básicas de A. Si aplicamos a $[A]_{M,J}$ las mismas operaciones elementales que a A, obtendremos las columnas básicas de E, es decir, una matriz $m \times r$ cuyas primeras r filas forman la matriz identidad, y las restantes $m - r$ filas son de ceros. Como esta matriz es una reducida por filas con r pivotes, $[A]_{M,J}$ tiene rango r.

Consideremos ahora $([A]_{M,J})^t$. Sabemos que también tiene rango r, luego el mismo argumento anterior nos dice que tomando sus columnas básicas obtenemos una matriz de rango r. Sean $q_1, \ldots, q_r$ sus columnas básicas. Observemos que tomar las columnas $q_1, \ldots, q_r$ de $([A]_{M,J})^t$ es lo mismo que definir $I = \{q_1, \ldots, q_r\}$ y considerar la matriz $([A]_{I,J})^t$. Esta matriz tendrá entonces rango r, luego la matriz $[A]_{I,J}$ también lo tendrá. Pero entonces $A_{I,J}$ es una submatriz de A, de orden $r \times r$ y rango r, luego su determinante es distinto de cero. Es decir, el menor $m_{I,J}$, de orden r, es no nulo.

Consideremos ahora un menor $m_{I',J'}$ de orden $k > r$. Como una submatriz no puede tener mayor rango que la matriz original, $[A]_{I',J'}$ es una matriz de orden $k \times k$ y rango menor o igual a $r < k$, luego su determinante debe ser cero. Es decir, $m_{I',J'} = 0$.

Hemos demostrado entonces que para todo r, una matriz de rango r admite un menor no nulo de orden r, y todos sus menores de orden mayor que r son nulos. Por tanto, el mayor orden alcanzado por los menores no nulos de A es igual al rango de A, como queríamos demostrar. □

Más adelante veremos que para calcular el rango de una matriz usando menores, no es necesario ver que se anulan *todos* los menores de un cierto orden.

5.4. Método del orlado

Recordemos que el rango de una matriz A es r si y sólo si existe un menor no nulo de orden r, y todos los menores de orden mayor que r son nulos. En principio, paras demostrar que una matriz tiene rango r habría que encontrar un menor no nulo de orden r, y calcular todos los menores de orden $r+1$ para ver que se anulan. En esta sección veremos que esto no es necesario. Si encontramos una submatriz cuadrada M de A tal que $\det(M) \neq 0$, y vemos que se anulan todos los menores de orden $r+1$ que corresponden a submatrices de A *que contienen a* M, entonces $\text{rango}(A) = r$.

Teorema principal del método del orlado

Sea A una matriz $m \times n$ con un menor no nulo de orden r, es decir, con una submatriz cuadrada M de orden r tal que $\det(M) \neq 0$. Supongamos que todas las submatrices cuadradas de orden $r+1$ que contienen a M tienen determinante 0. Entonces $\text{rango}(A) = r$.

PRUEBA: Observemos que intercambiar filas o columnas de una matriz no cambia su rango, ni el hecho de que sus menores de un cierto orden sean o no sean todos nulos. Por tanto, podemos suponer que la submatriz M está formada por las primeras r filas y columnas de A. Es decir, podremos escribir A de la forma:

$$A = \left(\begin{array}{c|c} M & A_1 \\ \hline A_2 & A_3 \end{array}\right).$$

Estamos suponiendo que $\det(M) \neq 0$, luego existe su inversa M^{-1}, y podemos considerar el siguiente producto:

$$\left(\begin{array}{c|c} M^{-1} & \mathbf{0} \\ \hline -A_2M^{-1} & I_{m-r} \end{array}\right)\left(\begin{array}{c|c} M & A_1 \\ \hline A_2 & A_3 \end{array}\right) = \left(\begin{array}{c|c} I_r & M^{-1}A_1 \\ \hline \mathbf{0} & -A_2M^{-1}A_1 + A_3 \end{array}\right).$$

La igualdad anterior es de la forma $PA = B$, donde P es una matriz cuadrada $m \times m$ tal que $\det(P) = \det(M^{-1}) \neq 0$. Por tanto, $\text{rango}(A) = \text{rango}(B)$, y sólo tenemos que probar que $\text{rango}(B) = r$. Bastará demostrar que $-A_2M^{-1}A_1 + A_3$ es la matriz nula, es decir, que $A_2M^{-1}A_1 = A_3$.

La fila i de la matriz $A_2M^{-1}A_1$ es $[A_2M^{-1}A_1]_{i*} = [A_2]_{i*}M^{-1}A_1$, y la columna j de esta fila, es decir, el elemento (i,j) de la matriz $A_2M^{-1}A_1$ es

$$[A_2M^{-1}A_1]_{ij} = [[A_2]_{i*}M^{-1}A_1]_{*j} = [A_2]_{i*}M^{-1}[A_1]_{*j}.$$

Por tanto, tenemos que demostrar que

$$\boxed{[A_2]_{i*}M^{-1}[A_1]_{*j} = [A_3]_{ij}}$$

para todo $i = 1,\dots,m-r$ y para todo $j = 1,\dots,n-r$.

Ahora consideremos la submatriz de A formada por las filas $1,\dots,r$ y $r+i$, y por las columnas $1,\dots,r$ y $r+j$. Es la siguiente:

$$\left(\begin{array}{c|c} M & [A_1]_{*j} \\ \hline [A_2]_{i*} & [A_3]_{ij} \end{array}\right).$$

Como esta submatriz es cuadrada de orden $r+1$ y contiene a M, tiene determinante cero. Observemos que se tiene siguiente producto:

$$\left(\begin{array}{c|c} M^{-1} & \mathbf{0} \\ \hline -[A_2]_{i*}M^{-1} & 1 \end{array}\right)\left(\begin{array}{c|c} M & [A_1]_{*j} \\ \hline [A_2]_{i*} & [A_3]_{ij} \end{array}\right) = \left(\begin{array}{c|c} I_r & M^{-1}[A_1]_{*j} \\ \hline \mathbf{0} & -[A_2]_{i*}M^{-1}[A_1]_{*j} + [A_3]_{ij} \end{array}\right).$$

Llamemos P, Q y R a las matrices que aparecen en esta fórmula. Tenemos entonces $PQ = R$. Por hipótesis $\det(Q) = 0$, y por otra parte el determinante de R es claramente $-[A_2]_{i*}\,M^{-1}\,[A_1]_{*j} + [A_3]_{ij}$. Por tanto, $-[A_2]_{i*}\,M^{-1}\,[A_1]_{*j} + [A_3]_{ij} = \det(R) = \det(P)\det(Q) = \det(P)\cdot 0 = 0$.

Luego $[A_2]_{i*}M^{-1}[A_1]_{*j} = [A_3]_{ij}$, como queríamos demostrar. □

A partir de este resultado, el **método del orlado** para calcular el rango de una matriz A es el siguiente. Por abuso del lenguaje, diremos que un menor $\det(M')$ contiene a $\det(M)$ si M es una submatriz de M'.

Método del orlado

Sea A una matriz $m \times n$. El método del orlado para calcular el rango de A es como sigue. Si $A = \mathbf{0}$, entonces $\text{rango}(A) = 0$. En caso contrario, tomamos un menor $\det(M)$ de orden r no nulo. Se realizan los siguientes pasos:

1. Halle, si existe, un menor no nulo de orden $r+1$ que contenga a M.
2. Si no lo hay, $\text{rango}(A) = r$.
3. Si lo hay, repetir el paso 1 aumentando r una unidad, y reemplazando M por el menor encontrado.

Ejemplo 5.4.1. Consideremos la matriz

$$A = \begin{pmatrix} 3 & 6 & 5 & 9 \\ 1 & 1 & 2 & 4 \\ 1 & -2 & 3 & 7 \end{pmatrix}.$$

El menor de A asociado a las dos primeras filas y las dos primeras columnas es

$$M = \det \begin{pmatrix} 3 & 6 \\ 1 & 1 \end{pmatrix} = -3 \neq 0.$$

Realicemos el orlado de esta submatriz con la tercera fila:

$$M_1 = \det \begin{pmatrix} 3 & 6 & 5 \\ 1 & 1 & 2 \\ 1 & -2 & 3 \end{pmatrix} = 0,$$

$$M_2 = \det \begin{pmatrix} 3 & 6 & 9 \\ 1 & 1 & 4 \\ 1 & -2 & 7 \end{pmatrix} = 0.$$

Por tanto, rango$(A) = 2$.

5.5. Regla de Cramer

Figura 5.1: Gabriel Cramer, (1704-1752)

Regla de Cramer

Sea $A_{n\times n}\boldsymbol{x} = \boldsymbol{b}$ un sistema con A matriz no singular y $\boldsymbol{u}$ su solución. Entonces la i-ésima componente u_i del vector $\boldsymbol{u}$ verifica que

$$x_i = \frac{\det(A_i)}{\det(A)},$$

donde $A_i = \left(\begin{array}{c|c|c|c|c|c|c} A_{*1} & \dots & A_{*i-1} & \boldsymbol{b} & A_{*i+1} & \dots & A_{*n} \end{array} \right)$. Esto es, A_i es la matriz que se obtiene de A cambiando la columna A_{*i} por $\boldsymbol{b}$.

PRUEBA: Consideremos las matrices $I_i(\boldsymbol{u}), i = 1,\dots,n$, que se obtienen al sustituir la columna i-ésima de la matriz identidad I por el vector $\boldsymbol{u}$. Si $A\boldsymbol{u} = \boldsymbol{b}$, la definición de multiplicación de matrices implica que

$$\begin{aligned} A\cdot I_i(\boldsymbol{u}) &= A\begin{pmatrix} \boldsymbol{e}_1 & \dots & \boldsymbol{u} & \dots & \boldsymbol{e}_n \end{pmatrix} \\ &= \begin{pmatrix} A\boldsymbol{e}_1 & \dots & A\boldsymbol{u} & \dots & A\boldsymbol{e}_n \end{pmatrix} \\ &= \begin{pmatrix} A_{*1} & \dots & \boldsymbol{b} & \dots & A_{*n} \end{pmatrix} = A_i. \end{aligned}$$

Por la propiedad multiplicativa de los determinantes,

$$\det(A)\det(I_i(\boldsymbol{u})) = \det(A_i).$$

Observemos que $\det(I_i(\boldsymbol{u})) = u_i$, pues basta realizar el desarrollo del determinante por la i-ésima fila. De aquí, $\det(A)\cdot u_i = \det(A_i)$ y tenemos el resultado por el carácter no singular de A.

□

La regla de Cramer tiene únicamente un interés teórico, pues no se aplica en cálculo numérico.

Ejemplo 5.5.1. Consideremos el sistema $A\boldsymbol{x} = \boldsymbol{b}$, con

$$A = \begin{pmatrix} 1 & 4 & 5 \\ 4 & 18 & 26 \\ 3 & 16 & 30 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 6 \\ 0 \\ -6 \end{pmatrix}.$$

Según la regla de Cramer,

$$x_1 = \frac{\det(A_1)}{\det(A)} = \frac{\begin{vmatrix} 6 & 4 & 5 \\ 0 & 18 & 26 \\ -6 & 16 & 30 \end{vmatrix}}{6} = \frac{660}{6} = 110,$$

$$x_2 = \frac{\det(A_2)}{\det(A)} = \frac{\begin{vmatrix} 1 & 6 & 5 \\ 4 & 0 & 26 \\ 3 & -6 & 30 \end{vmatrix}}{6} = \frac{-216}{6} = -36,$$

$$x_3 = \frac{\det(A_3)}{\det(A)} = \frac{\begin{vmatrix} 1 & 4 & 6 \\ 4 & 18 & 0 \\ 3 & 16 & -6 \end{vmatrix}}{6} = \frac{48}{6} = 8.$$

5.6. Cofactores y matriz inversa

Recordemos que el cofactor de $A_{n\times n}$ asociado con la posición (i, j) se define como

$$\hat{A}_{ij} = (-1)^{i+j} \det(A_{ij}),$$

donde A_{ij} es la submatriz de orden $n-1$ que se obtiene borrando la fila i-ésima y la columna j-ésima de la matriz A. La matriz de cofactores se notará por $\hat{A} = (\hat{A}_{ij})$.

Ejemplo 5.6.1. Los cofactores de la matriz

$$A = \begin{pmatrix} 1 & 4 & 5 \\ 4 & 18 & 26 \\ 3 & 16 & 30 \end{pmatrix}$$

son

$$\begin{aligned}
\hat{A}_{11} &= (-1)^2 \det\begin{pmatrix} 18 & 26 \\ 16 & 30 \end{pmatrix} = 124, \\
\hat{A}_{12} &= (-1)^3 \det\begin{bmatrix} 4 & 26 \\ 3 & 30 \end{bmatrix} = -42, \\
\hat{A}_{13} &= (-1)^4 \det\begin{bmatrix} 4 & 18 \\ 3 & 16 \end{bmatrix} = 10, \\
&\vdots \\
\hat{A}_{33} &= (-1)^6 \det\begin{bmatrix} 1 & 4 \\ 4 & 18 \end{bmatrix} = 2.
\end{aligned}$$

La matriz de cofactores es igual entonces a

$$\hat{A} = \begin{bmatrix} 124 & -42 & 10 \\ -40 & 15 & -4 \\ 14 & -6 & 2 \end{bmatrix}.$$

Inversa y matriz adjunta

Se define la matriz adjunta de $A_{n\times n}$ como $\text{Adj}(A) = \hat{A}^t$, la traspuesta de la matriz de cofactores. Si A es no singular, entonces

$$A^{-1} = \frac{\text{Adj}(A)}{\det(A)}.$$

PRUEBA: El elemento $[A^{-1}]_{ij}$ es la i-ésima componente de la solución del sistema $A\boldsymbol{x} = \boldsymbol{e}_j$, donde $\boldsymbol{e}_j$ es el j-ésimo vector de la base estándar. Por la regla de Cramer,

$$[A^{-1}]_{ij} = x_i = \frac{\det(A_i)}{\det(A)},$$

donde A_i es la matriz que se obtiene al cambiar la columna i-ésima de A por e_j, y el desarrollo de $\det(A_i)$ por la columna i-ésima implica que

$$\det(A_i) = \det \begin{pmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & 1 & \dots & a_{jn} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{pmatrix} = \hat{A}_{ji}.$$

$$\uparrow$$

columna i

□

Ejemplo 5.6.2. El cálculo de la inversa de la matriz del ejemplo 5.6.1

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 18 & 26 \\ 3 & 16 & 30 \end{bmatrix}$$

mediante la matriz adjunta

$$\mathrm{Adj}(A) = \hat{A}^t = \begin{bmatrix} 124 & -40 & 14 \\ -42 & 15 & -6 \\ 10 & -4 & 2 \end{bmatrix}$$

nos da

$$A^{-1} = \frac{\mathrm{Adj}(A)}{\det(A)} = \frac{1}{6} \begin{bmatrix} 124 & -40 & 14 \\ -42 & 15 & -6 \\ 10 & -4 & 2 \end{bmatrix} = \begin{bmatrix} \frac{62}{3} & -\frac{20}{3} & 7/3 \\ -7 & 5/2 & -1 \\ 5/3 & -2/3 & 1/3 \end{bmatrix}.$$

5.7. Determinantes y permutaciones

Denotaremos S_n al conjunto de las permutaciones de un conjunto de n elementos, es decir, al conjunto de las aplicaciones biyectivas

$$\sigma : \{1, \dots, n\} \longrightarrow \{1, \dots, n\}.$$

Dada una permutación $\sigma \in S_n$, una de las características de σ que va a tener más importancia en este tema será su ***número de inversiones***, $ni(\sigma)$, que es el

número de pares (i, j) tales que $i < j$ y $\sigma(i) > \sigma(j)$. Hay una forma gráfica muy útil para describir el número de inversiones de σ: Si disponemos dos copias del conjunto $\{1, \ldots, n\}$, donde los números están alineados verticalmente como en el ejemplo de la figura 5.2, la permutación σ viene representada por n segmentos, o líneas, que unen el elemento i de la primera copia al elemento $\sigma(i)$ de la segunda. Es claro que $ni(\sigma)$ es simplemente el número de cruces (intersecciones) que aparecen en esta representación. (Nota: se debe evitar que más de dos rectas se corten en el mismo punto, para poder contar los cruces de forma más clara.)

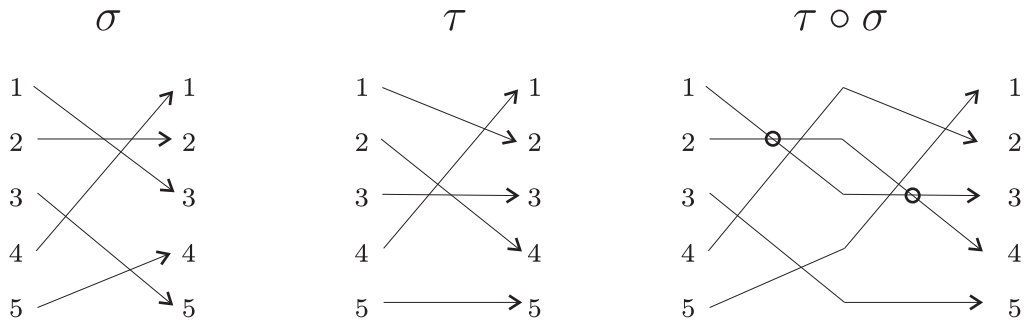

Figura 5.2: Ejemplo de permutaciones $\sigma, \tau \in S_5$ tales que $ni(\sigma) = 5$, $ni(\tau) = 4$ y $ni(\tau \circ \sigma) = 7$.

A partir de $ni(\sigma)$, se obtiene el ***signo*** (o la ***paridad***) de la permutación σ, dado simplemente por la fórmula:

$$sg(\sigma) = (-1)^{ni(\sigma)}.$$

Observemos que $sg(\sigma) = 1$ si $ni(\sigma)$ es un número par, y $sg(\sigma) = -1$ si $ni(\sigma)$ es un número impar. En el primer caso diremos que σ es una ***permutación par***, mientras que en el segundo diremos que σ es una ***permutación impar***.

Composición de permutaciones y signo

La composición de dos permutaciones pares, o de dos permutaciones impares, es una permutación par. La composición de una permutación par y una impar, es una permutación impar.

PRUEBA: Sean $\sigma, \tau \in S_n$ dos permutaciones, que representaremos gráficamente de forma consecutiva, como en la figura 5.2. Observemos que la composición $\tau \circ \sigma$ se obtiene siguiendo las líneas desde la primera copia de $\{1, \ldots, n\}$ a

la tercera. Un par (i, j) será una inversión de $\tau \circ \sigma$ si las líneas correspondientes se cruzan una sóla vez (o bien en σ o bien en τ). Por el contrario, (i, j) no será una inversión de $\tau \circ \sigma$ si las líneas correspondientes no se cruzan, o si se cruzan dos veces (los dos cruces marcados en la figura 5.2). Si llamamos m al número de pares (i, j) tales que sus líneas correspondientes se cruzan dos veces, este argumento nos dice que $ni(\tau \circ \sigma) = ni(\sigma) + ni(\tau) - 2m$. De aquí se deduce inmediatamente el resultado sobre la paridad de $\tau \circ \sigma$. □

Una consecuencia inmediata del resultado anterior es la siguiente propiedad:

Permutación inversa y signo

Dada $\sigma \in S_n$, se tiene $sg(\sigma) = sg(\sigma^{-1})$.

PRUEBA: Al ser $\sigma^{-1} \circ \sigma = \mathrm{Id}$ una permutación par (la identidad es la única permutación sin inversiones), el resultado anterior implica que σ y σ^{-1} tienen la misma paridad. □

Más adelante necesitaremos algún que otro resultado sobre permutaciones, pero ahora ya podemos dar la definición principal de este tema.

Determinante por permutaciones

Sea A una matriz cuadrada de orden $n \times n$. Se define el **determinante** por permutaciones de A, denotado $|A|$, como:

$$|A| = \sum_{\sigma \in S_n} sg(\sigma) \cdot [A]_{1\sigma(1)} [A]_{2\sigma(2)} \cdots [A]_{n\sigma(n)}.$$

Observemos que $|A|$ es una suma de $n!$ términos, puesto que hay un término por cada permutación de S_n. Cada sumando contiene un producto de la forma $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$. Estos n escalares son n entradas de la matriz A que están en n filas distintas, y también en n columnas distintas. Recíprocamente, si elegimos n entradas de A que estén en filas distintas y en columnas distintas, el producto de esos n escalares aparecerá en uno de los sumandos que definen $|A|$: la permutación correspondiente será la que asocia, a cada fila, la columna donde está la entrada elegida. Por tanto, $|A|$ es la suma de todos estos posibles productos, cada uno de ellos con un signo determinado por la permutación correspondiente.

Ejemplo 5.7.1. Si A es una matriz de orden 2×2, sólo hay dos permutaciones posibles en S_2, la identidad y la que permuta los elementos 1 y 2. La primera es par, y da lugar al sumando $a_{11}a_{22}$. La segunda es impar, y da lugar al sumando $-a_{12}a_{21}$. Por tanto, en el caso $n = 2$, tenemos la conocida fórmula:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Ejemplo 5.7.2. Si A es una matriz de orden 3×3, y como hay 6 permutaciones en S_3, la fórmula que define $|A|$ consta de 6 sumandos, tres de ellos pares y tres impares. Concretamente:

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

Dos resultados inmediatos son los siguientes. No necesitamos dar la demostración, al ser evidentes a partir de la definición del determinante.

Fila o columna de ceros

Si una matriz cuadrada A tiene una fila o una columna de ceros, entonces $|A| = 0$.

Determinante de la matriz identidad

Si I es la matriz identidad de orden $n \times n$, entonces $|I| = 1$.

5.7.1. Efecto por trasposición y operaciones elementales

Veamos ahora el efecto que producen las operaciones elementales por filas en el determinante de una matriz.

Determinante y operaciones elementales de tipo I

Si la matriz cuadrada B se obtiene de A al intercambiar dos filas (o dos columnas), entonces:

$$|B| = -|A|.$$

PRUEBA: Supongamos que B se obtiene de A al intercambiar las filas i y j, con $i < j$. Tendremos:

$$\begin{aligned} |B| &= \sum_{\sigma\in S_n} sg(\sigma)\cdot [B]_{1\sigma(1)}\cdots [B]_{i\sigma(i)}\cdots [B]_{j\sigma(j)}\cdots [B]_{n\sigma(n)} \\ &= \sum_{\sigma\in S_n} sg(\sigma)\cdot [A]_{1\sigma(1)}\cdots [A]_{j\sigma(i)}\cdots [A]_{i\sigma(j)}\cdots [A]_{n\sigma(n)} \\ &= \sum_{\sigma\in S_n} sg(\sigma)\cdot [A]_{1\sigma(1)}\cdots [A]_{i\sigma(j)}\cdots [A]_{j\sigma(i)}\cdots [A]_{n\sigma(n)}. \end{aligned}$$

Si llamamos τ a la permutación (llamada trasposición) que intercambia los elementos i y j, dejando invariantes los demás, se tiene:

$$|B| = \sum_{\sigma\in S_n} sg(\sigma)\cdot [A]_{1\sigma(\tau(1))}\cdots [A]_{i\sigma(\tau(i))}\cdots [A]_{j\sigma(\tau(j))}\cdots [A]_{n\sigma(\tau(n))}.$$

Ahora observemos que una trasposición siempre es impar, como se puede ver en su representación gráfica: hay $j-i-1$ líneas que empiezan entre i y j, y todas ellas se cruzan con las líneas i y j; además tenemos el cruce de la línea i con la línea j, y no hay ningún otro cruce. Por tanto, $ni(\tau) = 2(j-i-1)+1$, luego τ es impar. Esto implica que $sg(\sigma) = -sg(\sigma\circ\tau)$, y así tenemos:

$$\begin{aligned} |B| &= \sum_{\sigma\in S_n} -sg(\sigma\circ\tau)\cdot [A]_{1\sigma(\tau(1))}\cdots [A]_{i\sigma(\tau(i))}\cdots [A]_{j\sigma(\tau(j))}\cdots [A]_{n\sigma(\tau(n))} \\ &= -\sum_{\sigma\in S_n} sg(\sigma\circ\tau)\cdot [A]_{1\sigma(\tau(1))}\cdots [A]_{i\sigma(\tau(i))}\cdots [A]_{j\sigma(\tau(j))}\cdots [A]_{n\sigma(\tau(n))}. \end{aligned}$$

El conjunto de las permutaciones $\sigma\circ\tau$ cuando $\sigma\in S_n$ es, de nuevo, el conjunto S_n de todas las permutaciones. Luego en este sumatorio $\sigma\circ\tau$ toma como valor todas las posibles permutaciones y la suma da precisamente $|A|$. Por tanto se obtiene $|B| = -|A|$, como queríamos demostrar.

Supongamos ahora que B se obtiene de A al permutar dos columnas. En este caso B^t se obtiene de A^t al permutar dos filas, luego $|B| = |B^t| = -|A^t| = -|A|$. □

Un corolario inmediato de este resultado es el siguiente:

Determinante de matrices con filas o columnas iguales

Si A es una matriz cuadrada con dos filas o dos columnas iguales, entonces

$$|A| = 0.$$

PRUEBA: En este caso, A se obtiene de sí misma al intercambiar dos filas o dos columnas (aquellas que son iguales). Por tanto, $|A| = -|A|$, luego $|A| = 0$. □

Nota 5.7.3. En la demostración anterior, hemos supuesto que $2 \neq 0$ en el cuerpo k. En efecto, $|A| = -|A|$ es equivalente $2|A| = 0$, que implica $|A| = 0$ siempre que $2 \neq 0$. Hay cuerpos en los que esto no ocurre (por ejemplo $\mathbb{Z}/\mathbb{Z}2$), aunque incluso en este caso el resultado sigue siendo cierto. Un buen ejercicio es demostrar este resultado directamente con la definición de determinante. En concreto, supongamos que $A_{i*} = A_{j*}$, con $i \neq j$. Entonces para $1 \leq k \leq n$ tenemos que $a_{ik} = a_{jk}$. Si $\sigma \in S_n$, sea $\sigma' = \sigma \circ (i\,j)$. Vamos a probar que

$$a_{1\sigma(1)} \cdots a_{n\sigma(n)} = a_{1\sigma'(1)} \cdots a_{n\sigma'(n)}.$$

Esto se verifica porque $\sigma(k) = \sigma'(k)$ si $k \neq i, j$, de donde $a_{k\sigma(k)} = a_{k\sigma'(k)}$, mientras que $\sigma'(i) = \sigma(j)$ y $\sigma'(j) = \sigma(i)$, por lo que $a_{i\sigma'(i)} = a_{i\sigma(j)} = a_{j\sigma(j)}$ y $a_{j\sigma'(j)} = a_{j\sigma(i)} = a_{i\sigma(i)}$.

Pero $sg(\sigma') = -sg(\sigma)$, luego

$$sg(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} + sg(\sigma') a_{1\sigma'(1)} \cdots a_{n\sigma'(n)} = 0.$$

La aplicación $\sigma \mapsto \sigma'$ establece una correspondencia biyectiva entre las permutaciones pares e impares de S_n, por lo que

$$\begin{aligned} |A| &= \sum_{\sigma \in S_n} sg(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \\ &= \sum_{sg(\sigma)=1} (sg(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} + sg(\sigma') a_{1\sigma'(1)} \cdots a_{n\sigma'(n)}) \\ &= 0. \end{aligned}$$

Observemos que la prueba es independiente de la característica del cuerpo.

Para estudiar el efecto que una operación elemental de tipo II o III induce en el determinante de una matriz, veremos un resultado más general, que incluye a ambos.

Determinante y combinaciones lineales de filas

Sean $B, A_1, \ldots, A_t$ matrices $n \times n$ tales que:

- La fila i de B es una combinación lineal de las filas i de $A_1, \ldots, A_t$, es decir, existen unos escalares $\alpha_1, \ldots, \alpha_t$ tales que

$$[B]_{i*} = \alpha_1 [A_1]_{i*} + \cdots \alpha_t [A_t]_{i*}.$$

- Para todo $k \neq i$, las filas k de $B, A_1, \ldots, A_t$ son todas iguales.

Entonces

$$|B| = \alpha_1 |A_1| + \cdots + \alpha_t |A_t|.$$

PRUEBA: Para todo $k \neq i$, sabemos que la entrada (k, j) de cualquiera de las matrices estudiadas es $[B]_{k,j}$. Para la fila i, denotaremos $[A_m]_{i,j}$ a la entrada (i, j) de la matriz A_m, para $m = 1, \ldots, t$. Entonces tenemos $[B]_{i,j} = \alpha_1 [A_1]_{i,j} + \cdots + \alpha_t [A_t]_{i,j}$, para todo $j = 1, \ldots, n$.

La fórmula del determinante de A queda:

$$\begin{aligned}
|B| &= \sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots [B]_{i\sigma(i)} \cdots [B]_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots \left(\alpha_1 [A_1]_{i\sigma(i)} + \cdots + \alpha_t [A_t]_{i\sigma(i)}\right) \cdots [B]_{n\sigma(n)} \\
&= \left(\sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots (\alpha_1 [A_1]_{i\sigma(i)}) \cdots [B]_{n\sigma(n)} \right) + \cdots \\
&\qquad \cdots + \left(\sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots (\alpha_t [A_t]_{i\sigma(i)}) \cdots [B]_{n\sigma(n)} \right) \\
&= \alpha_1 \left(\sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots [A_1]_{i\sigma(i)} \cdots [B]_{n\sigma(n)} \right) + \cdots \\
&\qquad \cdots + \alpha_t \left(\sum_{\sigma \in S_n} sg(\sigma) \cdot [B]_{1\sigma(1)} \cdots [A_t]_{i\sigma(i)} \cdots [B]_{n\sigma(n)} \right) \\
&= \alpha_1 |A_1| + \cdots \alpha_t |A_t|,
\end{aligned}$$

como queríamos demostrar. □

Ejemplo 5.7.4. A partir de la combinación lineal $(2, 3, -1) = 2(1, 5, 1) + (-1)(0, 7, 3)$,

se deduce la igualdad:

$$\begin{vmatrix} 13 & 57 & 36 \\ 2 & 3 & -1 \\ 248 & 504 & 311 \end{vmatrix} = 2 \begin{vmatrix} 13 & 57 & 36 \\ 1 & 5 & 1 \\ 248 & 504 & 311 \end{vmatrix} + (-1) \begin{vmatrix} 13 & 57 & 36 \\ 0 & 7 & 3 \\ 248 & 504 & 311 \end{vmatrix}.$$

Determinante y operaciones elementales de tipo II

Si la matriz cuadrada B se obtiene de A al multiplicar una fila (o columna) por un escalar α, entonces:

$$|B| = \alpha |A|.$$

PRUEBA: Inmediato a partir de los resultados anteriores, tomando $t = 1$, $\alpha_1 = \alpha$ y $A_1 = A$. □

Determinante y operaciones elementales de tipo III

Si la matriz cuadrada B se obtiene de A al sumar a una fila (o columna) un múltiplo de otra, entonces:

$$|B| = |A|.$$

PRUEBA: Supongamos que B se obtiene de A al sumar, a la fila i, la fila j multiplicada por $\alpha \in k$. Denotemos $A_1 = A$, y sea A_2 la matriz que se obtiene de A al sustituir la fila i for la j, dejando la fila j como está: es decir, las filas i y j de A_2 son ambas iguales a la fila j de A. En este caso, B, A_1 y A_2 satisfacen las hipótesis de los resultados anteriores: las filas k con $k \neq i$ son todas iguales, y la fila i de B es combinación lineal de las filas i de A_1 y A_2. Concretamente:

$$[B]_{i*} = [A]_{i*} + \alpha [A]_{j*} = [A_1]_{i*} + \alpha [A_2]_{i*},$$

por lo que se tiene $|B| = |A_1| + \alpha |A_2|$. Como A_2 es una matriz con dos filas iguales, $|A_2| = 0$, luego $|B| = |A_1| = |A|$.

El caso de las operaciones elementales por columnas se demuestra igual. □

5.8. Igualdad de ambas definiciones

El objetivo de esta sección es probar que el determinante de una matriz definido por recurrencia y el definido por permutaciones coinciden. Una forma es comprobar que el determinante por permutaciones es igual al desarrollo por cofactores de una columna de la matriz. Vamos a seguir otro método, basado en la unicidad de las propiedades del determinante que hemos probado tanto para las dos definiciones. Por ello, necesitamos una definición para estas funciones.

Función determinante

Sea K un anillo y n un entero positivo. Una función $\delta_n : \mathscr{M}_{n\times n} \to K$ es una **función determinante** si satisface las siguientes propiedades:

1. $\delta_n(I) = 1$, donde I es la matriz identidad.
2. $\delta_n(A) = 0$ si A tiene una fila de ceros.
3. $\delta_n(P_{ij}A) = -\delta_n(A)$, donde P_{ij} es una matriz de permutación.
4. $\delta_n(T_i(\alpha)A) = \alpha\delta_n(A)$, donde $T_i(\alpha)$ es una matriz elemental de tipo II.
5. $\delta_n(T_{ij}(\alpha)A) = \delta_n(A)$, donde $T_{ij}(\alpha)$ es una matriz elemental de tipo III.

Ya sabemos que el determinante por recurrencia y el determinante por permutaciones son funciones determinante. Para probar que coinciden, basta ver que una función determinante está unívocamente determinada.

Unicidad de la función determinante

Sea K un cuerpo. Para cada entero positivo n existe a lo más una única función determinante $\delta_n : \mathscr{M}_{n\times n} \to K$. En consecuencia, $\det(A) = |A|$.

PRUEBA: Sean δ_n, γ_n dos funciones determinante y llamemos $\beta = \delta_n - \gamma_n$. La función β satisface las siguientes condiciones:

- $\beta(I) = 0$.

- $\beta(A) = 0$ si A tiene una fila de ceros.
- $\beta(P_{ij}A) = -\beta(A)$, donde P_{ij} es una matriz de permutación.
- $\beta(T_i(\alpha)A) = \alpha\beta(A)$, donde $T_i(\alpha)$ es una matriz elemental de tipo II.
- $\beta(T_{ij}(\alpha)A) = \beta(A)$, donde $T_{ij}(\alpha)$ es una matriz elemental de tipo III.

Lo anterior prueba que si E es una matriz elemental, entonces $\beta(A)$ y $\beta(EA)$ son ambos nulos o ambos distintos de cero. Dada una matriz A de orden n, existen matrices elementales $E_1, \ldots, E_r$ tales que $E_1 \cdots E_r A = E_A$, donde E_A es la forma escalonada reducida por filas de A. Si A tiene rango n, entonces $E_A = I$ y $\beta(E_1 \cdots E_r A) = 0$, de donde $\beta(A) = 0$. Si A es de rango inferior, entonces E_A contiene una fila de ceros, por lo que $\beta(E_A) = 0$ y $\beta(A) = 0$ □

Nota 5.8.1. La definición de la función determinante se puede hacer sobre anillos, como $\mathbb{Z}$ o $K[X_1, \ldots, X_n]$, donde sigue siendo válida la regla del producto (Adkins:Weintraub). Si el anillo es un dominio, podemos trabajar en el cuerpo de fracciones para obtener algunos resultados. Sin embargo, hay algunas propiedades de los vectores que no se verifican en los anillos. Por ejemplo, el determinante de una matriz puede ser cero, pero eso no implica que una columna o fila sea combinación lineal de las restantes. Por ejemplo, la matriz con coeficientes enteros

$$A = \begin{pmatrix} 12 & 18 \\ -6 & -9 \end{pmatrix}$$

es singular, pero ninguna columna es múltiplo entero de la otra, aunque sí son $\mathbb{Z}$-linealmente independientes.

Cuando se consideran espacios de funciones, hay que tener cuidado con algunas propiedades relativas a la independencia lineal. Por ejemplo, sea $n > 1$ y U un intervalo abierto de $\mathbb{R}$. Sea R el conjunto de funciones $f : U \to \mathbb{R}$que son diferenciables $n-1$ veces al menos. Dada $f \in R$, notaremos por Df su derivada y $D^h f$ su derivada h-ésima. Dadas $f_1, \ldots, f_n \in R$, la función

$$W(f_1, \ldots, f_n)(t) = \det \begin{pmatrix} f_1(t) & f_2(t) & \ldots & f_n(t) \\ (Df_1)(t) & (Df_2)(t) & \ldots & (Df_n)(t) \\ \vdots & \vdots & & \vdots \\ (D^{n-1}f_1)(t) & (Dn-1f_2)(t) & \ldots & (D^{n-1}f_n)(t) \end{pmatrix}$$

se denomina Wronskiano de $f_1, \ldots, f_n$. Se puede probar que si $W(f_1, \ldots, f_n)(t) \neq 0$ para algún $t \in U$, entonces el conjunto $\{f_1, \ldots, f_n\}$ es $\mathbb{R}$-linealmente independiente. El recíproco es falso. Sea U un intervalo que contiene al origen consideremos las funciones $f_1(t) = t^3, f_2(t) = |t|^3$. Entonces $\{f_1, f_2\}$ es un conjunto $\mathbb{R}$-linealmente independiente, pero $W(f_1, f_2)(t) = 0$ para todo $t \in U$.

5.9. Coste de cálculo del determinante

Con la definición por permutaciones, el determinante de una matriz de orden n se calcula mediante $n!$ sumandos, cada uno de ellos con $n-1$ productos. Por tanto, el número de operaciones necesarias son $n!-1$ sumas y $(n-1)\cdot n!$ productos, lo que hace un total de $n(n!)-1$ operaciones. Para una matriz de orden 25, son necesarias del orden de $3{,}9\times 10^{26}$ operaciones. Un ordenador que realice 10^{15} operaciones por segundo (un petaflop) necesitará 3170 años para completar el cálculo.

Veamos qué ocurre si usamos el desarrollo por cofactores de una fila o columna. Usemos, Por ejemplo, el desarrollo por la primera columna: tenemos

$$\det(A)=\sum_{i=1}^{n} a_{i1}\hat{A}_{i1}.$$

Llamemos p_n al número de operaciones necesarias mediante este método. Es claro que $p_1=1$, y para $n=2$ tenemos dos productos y una suma, luego $p_2=3$. Supongamos calculado p_{n-1}, el coste de un determinante de orden $n-1$. Entonces el desarrollo por cofactores necesita el cálculo de n determinantes de orden $n-1$, n productos y $n-1$ sumas. Tenemos así la ecuación

$$p_n = np_{n-1}+n+(n-1) = np_{n-1}+2n-1, \text{ si } n>2.$$

Los valores de p_n crecen de manera proporcional a $n!$. Por ejemplo, $p_{15}\approx 3{,}55\times 10^{12}$, $p_{25}\approx 4{,}21\times 10^{25}$. Es ligeramente menos costoso que el desarrollo por permutaciones.

En el caso de un cuerpo, las transformaciones elementales nos permiten realizar el cálculo del determinante con un coste del orden de $2/3n^3$, tal como se tiene para la eliminación gaussiana. Esto funciona correctamente cuando se trabaja en un cuerpo, pero en anillos más generales no se tiene la posibilidad de dividir por un escalar no nulo, como es el caso de $\mathbb{Z}$ o de los anillos de polinomios. Existen variantes de la eliminación gaussiana que evitan divisiones y permiten, con un pequeño coste añadido, tratar estos casos.

Ejemplo 5.9.1. El desarrollo de un determinante de orden n por los elementos de una fila y sus adjuntos es muy costoso, pues el número de términos crece rápidamente y los cálculos son muy complicados. El método óptimo para resolverlo es a través de la eliminación gaussiana, que alcanza una complejidad de $2/3n^3$ operaciones. Para ciertos determinantes no muy grandes, y con entradas números enteros se puede usar el método de condensación pivotal de Chiò.

Sea $A_{n\times n}=(a_{ij})$ una matriz, donde $a_{11}\neq 0$ (se puede hacer análogamente para cualquier elemento no nulo de la matriz). Multiplicamos cada fila, excepto

Figura 5.3: F. Chiò (1813-1871)

la primera, por a_{11}. Entonces

$$a_{11}^{n-1}\det(A)=\det\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n}\\ a_{21}a_{11} & a_{22}a_{11} & \dots & a_{2n}a_{11}\\ \vdots & \vdots & & \vdots\\ a_{n1}a_{11} & a_{n2}a_{11} & \dots & a_{nn}a_{11}\end{pmatrix}.$$

Ahora restamos a la segunda fila la primera multiplicada por a_{21}, a la tercera la primera multiplicada por a_{31}, hasta la n-ésima, que le restamos la primera multiplicada por a_{n1}. Nos queda

$$\begin{aligned} a_{11}^{n-1}\det(A) &= \det\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n}\\ 0 & a_{22}a_{11}-a_{21}a_{12} & \dots & a_{2n}a_{11}-a_{21}a_{1n}\\ \vdots & \vdots & & \vdots\\ 0 & a_{n2}a_{11}-a_{12}a_{n1} & \dots & a_{nn}a_{11}-a_{1n}a_{n1}\end{pmatrix}\\ &= \det\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n}\\ 0 & \begin{vmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix} & \begin{vmatrix} a_{11} & a_{13}\\ a_{21} & a_{23}\end{vmatrix} & \dots & \begin{vmatrix} a_{11} & a_{1n}\\ a_{21} & a_{2n}\end{vmatrix}\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & \begin{vmatrix} a_{11} & a_{12}\\ a_{n1} & a_{n2}\end{vmatrix} & \begin{vmatrix} a_{11} & a_{13}\\ a_{n1} & a_{n3}\end{vmatrix} & \dots & \begin{vmatrix} a_{11} & a_{1n}\\ a_{n1} & a_{nn}\end{vmatrix}\end{pmatrix}\\ &= a_{11}\det(B). \end{aligned}$$

Entonces $\det(A)=\frac{1}{a_{11}^{n-2}}\det(B)$. Si el elemento fuera el (i_0,j_0), hay que incluir un factor $(-1)^{i_0+j_0}$ por el último desarrollo. Se suele buscar un elemento que sea

igual a 1, para simplificar las operaciones. Por ejemplo,

$$\begin{aligned}
\det\begin{pmatrix} 1 & 2 & 3 & 4 \\ 8 & 7 & 6 & 5 \\ 1 & 8 & 2 & 7 \\ 3 & 6 & 4 & 5 \end{pmatrix} &= 1\cdot\det\begin{pmatrix} \begin{vmatrix} 1 & 2 \\ 8 & 7 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 8 & 6 \end{vmatrix} & \begin{vmatrix} 1 & 4 \\ 8 & 5 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 1 & 8 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} & \begin{vmatrix} 1 & 4 \\ 1 & 7 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 3 & 6 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 3 & 4 \end{vmatrix} & \begin{vmatrix} 1 & 4 \\ 3 & 5 \end{vmatrix} \end{pmatrix} \\
&= \det\begin{pmatrix} -9 & -18 & -27 \\ 6 & -1 & 3 \\ 0 & -5 & -7 \end{pmatrix} = (-9)\cdot\det\begin{pmatrix} 1 & 2 & 3 \\ 6 & -1 & 3 \\ 0 & -5 & -7 \end{pmatrix} \\
&= (-9)\cdot\det\begin{pmatrix} \begin{vmatrix} 1 & 2 \\ 6 & -1 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 6 & 3 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 0 & -5 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 0 & -7 \end{vmatrix} \end{pmatrix} \\
&= (-9)\cdot\det\begin{pmatrix} -13 & -15 \\ -5 & -7 \end{pmatrix} = (-9)(91-75) = -144.
\end{aligned}$$

Este método consume más operaciones que la eliminación gaussiana. Sin embargo, es uno de los que se utilizan dentro del contexto de eliminación gaussiana libre de fracciones, usada en cálculo simbólico o para matrices con entradas en dominios no cuerpos.

La resolución de un sistema de ecuaciones mediante la regla de Cramer presenta también desventajas con respecto a la eliminación gaussiana. Mediante la eliminación, sabemos que un sistema de ecuaciones tiene un coste de resolución del orden de $\frac{2}{3}n^3$; como hemos comentado antes, este es el orden del cálculo de un determinante. La regla de Cramer supone el cálculo de $n+1$ determinantes: uno para el determinante de la matriz de coeficientes del sistema, y n con los determinantes en los que se ha cambiado una columna. Por ello, la regla de Cramer tiene un coste del orden de $\frac{2}{3}n^4$, que es sensiblemente superior. Además, la eliminación gaussiana permite la resolución simultánea de varios sistemas con la misma matriz de coeficientes, pero diferentes términos independientes. La regla de Cramer obliga a recalcular n determinantes.

Capítulo 6

Producto escalar y ortogonalidad

6.1. Normas vectoriales

Una gran parte del Álgebra Lineal es Geometría, porque la materia creció por la necesidad de generalizar la geometría básica de $\mathbb{R}^2$ y $\mathbb{R}^3$ a espacios de dimensión superior. La aproximación habitual es poner en coordenadas conceptos geométricos de $\mathbb{R}^2$ y $\mathbb{R}^3$, y extender enunciados relativos a pares o ternas a n-uplas en $\mathbb{R}^n$ y $\mathbb{C}^n$.

Por ejemplo, la longitud de un vector $\boldsymbol{u} \in \mathbb{R}^2$ o $\boldsymbol{v} \in \mathbb{R}^3$ se obtiene del teorema de Pitágoras calculando la longitud de la hipotenusa de un triángulo rectángulo. Esta medida de la longitud

$$\|\boldsymbol{u}\| = \sqrt{x^2 + y^2} \text{ y } \|\boldsymbol{v}\| = \sqrt{x^2 + y^2 + z^2},$$

se denomina ***norma euclídea*** en $\mathbb{R}^2$ y $\mathbb{R}^3$, y se extiende de manera obvia a dimensiones superiores.

Norma vectorial euclídea

Para un vector $\boldsymbol{x}_{n\times 1}$, la ***norma euclídea*** de $\boldsymbol{x}$ se define como

- $\|\boldsymbol{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = \sqrt{\boldsymbol{x}^t\boldsymbol{x}}$, cuando $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$.
- $\|\boldsymbol{x}\| = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \sqrt{\boldsymbol{x}^*\boldsymbol{x}}$, cuando $\boldsymbol{x} \in \mathbb{C}^{n\times 1}$.

Por ejemplo, si

$$\boldsymbol{u} = \begin{pmatrix} 0 \\ -1 \\ 2 \\ -2 \\ 4 \end{pmatrix} \text{ y } \boldsymbol{v} = \begin{pmatrix} i \\ 2 \\ 1-i \\ 0 \\ 1+i \end{pmatrix},$$

entonces

$$\|\boldsymbol{u}\| = \sqrt{\sum u_i^2} = \sqrt{\boldsymbol{u}^t\boldsymbol{u}} = \sqrt{0+1+4+4+16} = 5,$$
$$\|\boldsymbol{v}\| = \sqrt{\sum |v_i|^2} = \sqrt{\boldsymbol{v}^*\boldsymbol{v}} = \sqrt{1+4+2+0+2} = 3.$$

Observemos lo siguiente:

- La versión compleja de $\|\boldsymbol{x}\|$ incluye a la versión real como un caso especial porque $|z|^2 = z^2$ cuando z es un número real. Recordemos que si $z = a+ib$ entonces $\bar{z} = a-ib$, y el módulo de z es $|z| = \sqrt{a^2+b^2} = \sqrt{\bar{z}z}$. El hecho de que $|z|^2$ es un número real asegura que $\|\boldsymbol{x}\|$ es un número real, incluso cuando $\boldsymbol{x}$ contenga entradas complejas.

- La definición de norma euclídea garantiza que para todos los escalares α,

$$\|\boldsymbol{x}\| \geq 0, \|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}, \text{ y } \|\alpha\boldsymbol{x}\| = |\alpha|\,\|\boldsymbol{x}\|.$$

- Dado un vector $\boldsymbol{x} \neq \boldsymbol{0}$, es frecuente obtener otro vector que tenga la misma dirección de $\boldsymbol{x}$, pero longitud unidad. Para ello, ***normalizamos*** el vector $\boldsymbol{x}$ con $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$. Entonces

$$\|\boldsymbol{u}\| = \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \right\| = \frac{1}{\|\boldsymbol{x}\|}\|\boldsymbol{x}\| = 1.$$

Producto escalar estándar

Los escalares definidos por

$$\boldsymbol{y}^t\boldsymbol{x} = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}, \boldsymbol{y}^*\boldsymbol{x} = \sum_{i=1}^{n} \bar{y}_i x_i \in \mathbb{C}$$

se denominan ***productos escalares estándar*** para $\mathbb{R}^n$ y $\mathbb{C}^n$, respectivamente.

Desigualdad CBS

$$|x^* y| \leq \|x\| \, \|y\|, \text{ para } x, y \in \mathbb{C}^n.$$

La igualdad se da si y solamente si $y = \alpha x$, con $\alpha = x^* y / x^* x$.

PRUEBA: Podemos suponer $x \neq 0$. Sea $\alpha = x^* y / x^* x$, y observemos que $x^*(\alpha x - y) = 0$. Entonces

$$\begin{aligned} 0 \leq \|\alpha x - y\|^2 &= (\alpha x - y)^*(\alpha x - y) = \bar{\alpha} x^*(\alpha x - y) - y^*(\alpha x - y) \\ &= -y^*(\alpha x - y) = y^* y - \alpha y^* x = \frac{\|y\|^2 \|x\|^2 - (x^* y)(y^* x)}{\|x\|^2}. \end{aligned}$$

Como $y^* x = \overline{(x^* y)}$, se sigue que $(x^* y)(y^* x) = |x^* y|^2$, luego

$$0 \leq \frac{\|y\|^2 \|x\|^2 - |x^* y|^2}{\|x\|^2}.$$

Como $\|x\|^2 > 0$ tenemos que $0 \leq \|y\|^2 \|x\|^2 - |x^* y|^2$ y se sigue la desigualdad. □

Desigualdad triangular

$$\|x + y\| \leq \|x\| + \|y\| \text{ para todo } x, y \in \mathbb{C}^n.$$

PRUEBA:

$$\begin{aligned} \|x + y\|^2 &= (x + y)^*(x + y) = x^* x + x^* y + y^* x + y^* y \\ &= \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}(x^* y). \end{aligned}$$

Por la desigualdad CBS, tenemos que

$$2\,\mathrm{Re}(x^* y) \leq 2\left|x^* y\right| \leq 2\|x\| \, \|y\|.$$

Volviendo a la ecuación anterior, nos queda

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

que es lo que queríamos. □

p-normas

Para $p \geq 1$, la p-norma de $\boldsymbol{x} \in \mathbb{C}^n$ se define como

$$\boldsymbol{x}_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

En la práctica, se usan solamente tres p-normas:

- $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$,
- $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$,
- $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$.

Normas vectoriales

Una ***norma*** para un espacio vectorial $\mathcal{V}$ real o complejo es una función $\|\cdot\|$ de $\mathcal{V}$ en $\mathbb{R}$ que verifica las siguientes condiciones:

1. $\|\boldsymbol{x}\| \geq 0$ y $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$.
2. $\|\alpha \boldsymbol{x}\| = |\alpha| \, \|\boldsymbol{x}\|$.
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Dos normas $\|\cdot\|_a, \|\cdot\|_b$ son *equivalentes* si existen $c_1, c_2 > 0$ tales que $c_1 \|\boldsymbol{v}\|_b \leq \|\boldsymbol{b}\|_a \leq c_2 \|\boldsymbol{v}\|_b$ para todo vector $\boldsymbol{v}$. ¿De dónde viene esta definición? Las normas vectoriales son necesarias para definir el límite de sucesiones de vectores. Una sucesión $\{\boldsymbol{v}_k\} \subset \mathcal{V}$ se dice que converge a $\boldsymbol{x}$ si la sucesión de números reales $\|\boldsymbol{x}_k - \boldsymbol{x}\|$ converge a cero. Esta convergencia depende de la norma, y podríamos tener convergencia con respecto a una norma y no con otra. Sin embargo, en los espacios vectoriales de dimensión finita, todas las normas son equivalentes [**?**]. Esto implica que la convergencia de una sucesión respecto de una norma implica la convergencia con respecto a cualquier otra norma, y al mismo límite.

6.2. Espacios con producto escalar

La norma euclídea, que apareció primero, es un concepto que depende de coordenadas. Aislando sus propiedades podemos dar una definición independiente de las mismas.

Definición de producto escalar

Un ***producto escalar*** sobre un espacio vectorial real o complejo es una función que aplica cada par de vectores $\boldsymbol{x}, \boldsymbol{y}$ en un escalar real o complejo $\boldsymbol{x} \bullet \boldsymbol{y}$ con las siguientes propiedades:

- $\boldsymbol{x} \bullet \boldsymbol{x}$ es real, con $\boldsymbol{x} \bullet \boldsymbol{x} \geq 0$, y $\boldsymbol{x} \bullet \boldsymbol{x} = 0$ si y solamente si $\boldsymbol{x} = \mathbf{0}$.
- $\alpha \boldsymbol{x} \bullet \boldsymbol{y} = \alpha(\boldsymbol{x} \bullet \boldsymbol{y})$ para todos los escalares α.
- $(\boldsymbol{x} + \boldsymbol{y}) \bullet \boldsymbol{z} = \boldsymbol{x} \bullet \boldsymbol{z} + \boldsymbol{y} \bullet \boldsymbol{z}$.
- $\boldsymbol{x} \bullet \boldsymbol{y} = \overline{\boldsymbol{y} \bullet \boldsymbol{x}}$. Para espacios reales, queda $\boldsymbol{x} \bullet \boldsymbol{y} = \boldsymbol{y} \bullet \boldsymbol{x}$.

Observemos que la última propiedad implica que

$$\boldsymbol{x} \bullet (\alpha \boldsymbol{y}) = \overline{(\alpha \boldsymbol{y}) \bullet \boldsymbol{x}} = \overline{\alpha(\boldsymbol{y} \bullet \boldsymbol{x})} = \overline{\alpha}(\boldsymbol{x} \bullet \boldsymbol{y}),$$

y

$$\boldsymbol{x} \bullet (\boldsymbol{y} + \boldsymbol{z}) = \overline{(\boldsymbol{y} + \boldsymbol{z}) \bullet \boldsymbol{x}} = \overline{\boldsymbol{y} \bullet \boldsymbol{x} + \boldsymbol{z} \bullet \boldsymbol{x}} = \overline{\boldsymbol{y} \bullet \boldsymbol{x}} + \overline{\boldsymbol{z} \bullet \boldsymbol{x}} = \boldsymbol{x} \bullet \boldsymbol{y} + \boldsymbol{x} \bullet \boldsymbol{z}.$$

Los espacios vectoriales reales con un producto escalar se denominan ***espacios euclídeos***. Los espacios vectoriales complejos con un producto escalar se denominan ***espacios unitarios***.

Los productos escalares estándares son

- $\boldsymbol{x} \bullet \boldsymbol{y} = \boldsymbol{y}^t \boldsymbol{x}$ para $\mathbb{R}^n$,
- $\boldsymbol{x} \bullet \boldsymbol{y} = \boldsymbol{y}^* \boldsymbol{x}$ para $\mathbb{C}^n$.

A partir de un producto escalar se define una norma $\|\boldsymbol{x}\| = (\boldsymbol{x} \bullet \boldsymbol{x})^{1/2}$. En general, si no especificamos la norma, nos referimos a $\|\cdot\|_2$.

Relación entre A, A^*A y AA^*

Sea A una matriz de orden $m \times n$. Entonces

1. $\text{null}(A) = \text{null}(A^*A)$.
2. $\text{rango}(A) = \text{rango}(A^*A) = \text{rango}(AA^*)$.
3. $\text{Col}(A^*) = \text{Col}(A^*A), \text{Col}(A) = \text{Col}(AA^*)$.

PRUEBA: Si $A\boldsymbol{v} = \boldsymbol{0}$, entonces $(A^*A)\boldsymbol{v} = \boldsymbol{0}$, por lo que $\text{null}(A) \subset \text{null}(A^*A)$. Recíprocamente, si $(A^*A)\boldsymbol{v} = \boldsymbol{0}$, entonces

$$\|A\boldsymbol{v}\|^2 = (A\boldsymbol{v}) \bullet (A\boldsymbol{v}) = \boldsymbol{v}^* A^* A\boldsymbol{v} = 0$$

de donde $A\boldsymbol{v} = \boldsymbol{0}$, y tenemos la igualdad $\text{null}(A) = \text{null}(A^*A)$.

Entonces

$$\begin{aligned} \text{rango}(A) &= \dim(\text{Col}(A)) = n - \dim(\text{null}(A)) = n - \dim(\text{null}(A^*A)) \\ &= \dim(\text{Col}(A^*A)) = \text{rango}(A^*A), \end{aligned}$$

y si aplicamos este resultado a A^*, nos queda $\text{rango}(A^*) = \text{rango}(AA^*)$.

Si $\boldsymbol{b} \in \text{Col}(A^*A)$, entonces $\boldsymbol{b} = A^*A\boldsymbol{u} = A^*(A\boldsymbol{u})$, de donde $\boldsymbol{b} \in \text{Col}(A^*)$. Así, $\text{Col}(A^*A) \subset \text{Col}(A^*)$, y como son de la misma dimensión, son iguales. La otra igualdad se deduce de la anterior al aplicarla a A^*. □

6.3. Vectores ortogonales

Ortogonalidad

En un espacio vectorial $\mathcal{V}$ con producto escalar, dos vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$ se dicen ***ortogonales*** si $\boldsymbol{x} \bullet \boldsymbol{y} = 0$. Lo notaremos como $\boldsymbol{x} \perp \boldsymbol{y}$.

En $\mathbb{R}^n$ con el producto escalar estándar, $\boldsymbol{x} \perp \boldsymbol{y} \Leftrightarrow \boldsymbol{y}^t\boldsymbol{x} = 0$. En $\mathbb{C}^n$ con el producto escalar estándar, $\boldsymbol{x} \perp \boldsymbol{y} \Leftrightarrow \boldsymbol{y}^*\boldsymbol{x} = 0$.

Conjuntos ortonormales

Un conjunto $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ es un ***conjunto ortogonal*** si $\boldsymbol{u}_i \perp \boldsymbol{u}_j$ cuando $i \neq j$, y $\boldsymbol{u}_i \neq \boldsymbol{0}$ para todo i.
Un conjunto $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ es un ***conjunto ortonormal*** si $\boldsymbol{u}_i \perp \boldsymbol{u}_j$ cuando $i \neq j$ y $\|\boldsymbol{u}_i\| = 1$ para todo i.

Es fácil ver que un conjunto ortogonal es linealmente independiente. En efecto, consideremos un conjunto ortogonal $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$, y una combinación lineal

$$\alpha_1 \boldsymbol{u}_1 + \alpha_2 \boldsymbol{u}_2 + \cdots + \alpha_n \boldsymbol{u}_n = \boldsymbol{0}.$$

Si realizamos el producto escalar a ambos lados de esta igualdad por $\boldsymbol{u}_i$, para cada $i = 1, \ldots, n$, nos queda

$$0 = \left(\sum_{j=1}^{n} \alpha_j \boldsymbol{u}_j \right) \bullet \boldsymbol{u}_i = \sum_{j=1}^{n} \alpha_j (\boldsymbol{u}_j \bullet \boldsymbol{u}_i).$$

Todos los sumandos de la derecha son nulos, salvo $\boldsymbol{u}_i \bullet \boldsymbol{u}_i$, que es positivo. Entonces $\alpha_i = 0$, para todo $i = 1, \ldots, n$.

Ejemplo 6.3.1. El conjunto

$$\mathscr{B}' = \left\{ \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}$$

es un conjunto ortogonal, pero no es ortonormal. Como $\|\boldsymbol{u}_1\| = \sqrt{2}, \|\boldsymbol{u}_2\| = \sqrt{3}, \|\boldsymbol{u}_3\| = \sqrt{6}$, el conjunto $\mathscr{B} = \{\frac{1}{\sqrt{2}}\boldsymbol{u}_1, \frac{1}{\sqrt{3}}\boldsymbol{u}_2, \frac{1}{\sqrt{6}}\boldsymbol{u}_3\}$ es ortonormal.

Expansión de Fourier

Si $\mathscr{B} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ es una base ortonormal de un espacio vectorial euclídeo, entonces cada $\boldsymbol{x} \in \mathscr{V}$ se puede expresar como

$$\boldsymbol{x} = (\boldsymbol{x} \bullet \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{x} \bullet \boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{x} \bullet \boldsymbol{u}_n)\boldsymbol{u}_n.$$

Esta expresión se denomina ***expansión de Fourier*** del vector $\boldsymbol{x}$. Los escalares $x_i = \boldsymbol{x} \bullet \boldsymbol{u}_i$ son las coordenadas de $\boldsymbol{x}$ respecto de la base $\mathscr{B}$, y se denominan ***coeficientes de Fourier***.

Desde el punto de vista geométrico, la expansión de Fourier contiene la proyección ortogonal de $\boldsymbol{x}$ sobre el espacio generado por el conjunto de los $\boldsymbol{u}_i$.

Figura 6.1: Joseph Fourier (1768-1830)

6.4. Matrices ortogonales y unitarias

En esta sección examinamos las matrices cuadradas cuyas columnas (o filas) son ortonormales.

Matrices unitarias y ortogonales

- Una ***matriz unitaria*** es una matriz *compleja* $U_{n\times n}$ cuyas columnas (o filas) constituyen una base ortonormal de $\mathbb{C}^n$.
- Una ***matriz ortogonal*** es una matriz *real* $Q_{n\times n}$ cuyas columnas (o filas) constituyen una base ortonormal de $\mathbb{R}^n$.

Ejemplo 6.4.1. 1. La matriz

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

es una matriz unitaria, pues sus columnas $\boldsymbol{u}_1 = U_{*1}, \boldsymbol{u}_2 = U_{*2}$ verifican que

$$\boldsymbol{u}_1^* \boldsymbol{u}_1 = 1, \boldsymbol{u}_1^* \boldsymbol{u}_2 = 0, \boldsymbol{u}_2^* \boldsymbol{u}_2 = 1.$$

2. La matriz

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

es una matriz ortogonal, pues sus columnas $\boldsymbol{q}_1 = Q_{*1}, \boldsymbol{q}_2 = Q_{*2}$ verifican que

$$\boldsymbol{q}_1^t \boldsymbol{q}_1 = 1, \boldsymbol{q}_1^t \boldsymbol{q}_2 = 0, \boldsymbol{q}_2^t \boldsymbol{q}_2 = 1.$$

Las matrices unitarias y ortogonales tienen unas propiedades interesantes, una de las cuales es que son fáciles de invertir. Para ello, observemos que las columnas de $U_{n\times n} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_n \end{pmatrix}$ forman un conjunto ortonormal si y solamente si

$$[U^*U]_{ij} = \boldsymbol{u}_i^* \boldsymbol{u}_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j, \end{cases} \Leftrightarrow U^*U = I \Leftrightarrow U^{-1} = U^*.$$

Nótese que $U^*U = I \Leftrightarrow UU^* = I$, es decir, las columnas de U son ortonormales si y solamente si las filas de U son ortonormales.

Otra importante propiedad es que la multiplicación por una matriz unitaria no cambia la longitud de un vector. En efecto, si U es una matriz unitaria, entonces

$$\|U\boldsymbol{x}\|_2^2 = \boldsymbol{x}^* U^* U \boldsymbol{x} = \boldsymbol{x}^* \boldsymbol{x} = \|\boldsymbol{x}\|_2^2, \text{ para todo } \boldsymbol{x} \in \mathbb{C}^n. \tag{6.4.1}$$

Recíprocamente, si U es una matriz que verifica (6.4.1), entonces es unitaria. Para ello, consideremos en primer lugar $\boldsymbol{x} = \boldsymbol{e}_i$. Recordemos que $U\boldsymbol{e}_i = U_{*i} = \boldsymbol{u}_i$, de donde

$$\boldsymbol{e}_i^* U^* U \boldsymbol{e}_i = \boldsymbol{u}_i^* \boldsymbol{u}_i = 1.$$

Tenemos así el carácter unitario de las columnas de U. Veamos la ortogonalidad. Para ello, sea $\boldsymbol{x} = \boldsymbol{e}_j + \boldsymbol{e}_k$, con $j \neq k$. Entonces

$$(\boldsymbol{e}_j + \boldsymbol{e}_k)^* U^* U (\boldsymbol{e}_j + \boldsymbol{e}_k) = (\boldsymbol{e}_j + \boldsymbol{e}_k)^* (\boldsymbol{e}_j + \boldsymbol{e}_k).$$

El lado izquierdo de la igualdad es $\boldsymbol{e}_j^* \boldsymbol{e}_j + \boldsymbol{e}_j^* U^* U \boldsymbol{e}_k + \boldsymbol{e}_k^* U^* U \boldsymbol{e}_j + \boldsymbol{e}_k^* \boldsymbol{e}_k = 2 + \boldsymbol{u}_j^* \boldsymbol{u}_k + \boldsymbol{u}_k^* \boldsymbol{u}_j$. El lado derecho es igual a $2 + \boldsymbol{e}_j^* \boldsymbol{e}_k + \boldsymbol{e}_k^* \boldsymbol{e}_j = 0$. Entonces

$$\boldsymbol{u}_j^* \boldsymbol{u}_k + \boldsymbol{u}_k^* \boldsymbol{u}_j = 0,$$

es decir, $2\,\mathrm{Re}(\boldsymbol{u}_j^* \boldsymbol{u}_k) = 0$. Si ahora ponemos $\boldsymbol{x} = \boldsymbol{e}_j + i\boldsymbol{e}_k$, se sigue que $2\,\mathrm{Im}(\boldsymbol{u}_j^* \boldsymbol{u}_k) = 0$.

Caracterización de las matrices unitarias y ortogonales

- Las siguientes condiciones son equivalentes a que la matriz compleja $U_{n\times n}$ es unitaria.
  - U tiene columnas ortonormales.
  - U tiene filas ortonormales.
  - $U^{-1} = U^*$, o bien que $U^*U = I_n$.
  - $\|U\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.
- Las siguientes condiciones son equivalentes a que la matriz real $Q_{n\times n}$ es ortogonal.
  - Q tiene columnas ortonormales.
  - Q tiene filas ortonormales.
  - $Q^{-1} = Q^t$, o bien que $Q^tQ = I_n$.
  - $\|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

6.5. Procedimiento de Gram-Schmidt

Los espacios $\mathbb{R}^n$ y $\mathbb{C}^n$ tienen unas bases ortonormales muy sencillas, como son las bases estándares. Sin embargo, nos planteamos una pregunta: dado un espacio de dimensión finita, ¿tiene una base ortonormal? Pensemos en una variedad lineal, dada por un sistema de generadores, y de la que queremos calcular una base ortonormal para, posteriormente, realizar proyecciones. La respuesta a esta pregunta la encontramos con el procedimiento de Gram-Schmidt.

Sea $\mathscr{S} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s\}$ un conjunto de vectores linealmente independientes en $\mathbb{K}^m$. Vamos a construir un conjunto ortonormal de vectores $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_s\}$ tal que para cada k se verifica $\langle \boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\rangle = \langle \boldsymbol{q}_1, \ldots, \boldsymbol{q}_k\rangle$.

El proceso es por inducción sobre s. Para $s = 1$, tomamos simplemente $\boldsymbol{q}_1 = \boldsymbol{v}_1 / \|\boldsymbol{v}_1\|$. En general, escribimos

$$\begin{aligned}
\boldsymbol{q}'_1 &= \boldsymbol{v}_1,\\
\boldsymbol{q}'_2 &= \boldsymbol{v}_2 - \lambda_{12}\boldsymbol{q}'_1\\
&\vdots\\
\boldsymbol{q}'_s &= \boldsymbol{v}_s - \lambda_{1s}\boldsymbol{q}'_1 - \cdots - \lambda_{s-1,s}\boldsymbol{q}'_{s-1},
\end{aligned}$$

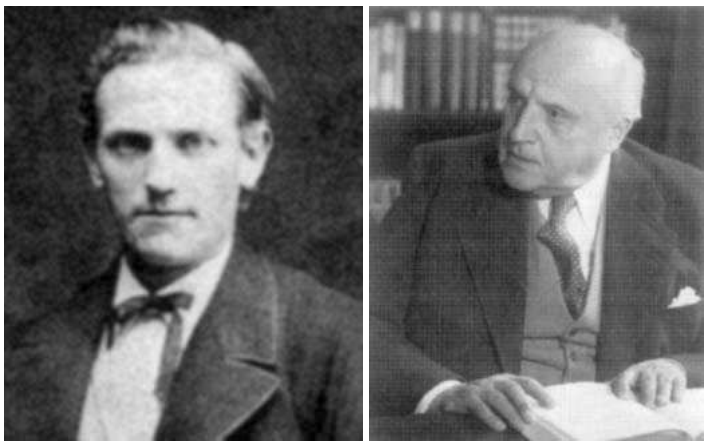

Figura 6.2: Jorgen P. Gram (1850-1916), Erhard Schmidt (1876-1959)

para ciertos escalares λ_{ij}. Buscamos esta forma en los vectores $\boldsymbol{q}'_i$ para que generen la misma variedad lineal que los $\boldsymbol{v}_i$. Queremos imponer la condición de ortogonalidad en el conjunto $\{\boldsymbol{q}'_1,\ldots,\boldsymbol{q}'_s\}$. Para ello tiene que ocurrir que

$$\boldsymbol{q}'_1 \bullet \boldsymbol{q}'_2 = 0 = (\boldsymbol{v}_2 \bullet \boldsymbol{q}'_1) - \lambda_{12} \left\|\boldsymbol{q}'_1\right\|^2,$$

de donde podemos calcular λ_{12} y entonces tenemos el vector $\boldsymbol{q}'_2$. Es claro que $\langle \boldsymbol{q}'_1, \boldsymbol{q}'_2\rangle = \langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle$. En general, si tenemos construidos $\boldsymbol{q}'_1,\ldots,\boldsymbol{q}'_{k-1}$, con

$$\langle \boldsymbol{v}_1,\ldots,\boldsymbol{v}_{k-1}\rangle = \langle \boldsymbol{q}'_1,\ldots,\boldsymbol{q}'_{k-1}\rangle,$$

entonces para obtener $\boldsymbol{q}'_k$ imponemos las condiciones

$$\begin{array}{rcccl}
0 & = & \boldsymbol{q}'_k \bullet \boldsymbol{q}'_1 & = & (\boldsymbol{v}_k \bullet \boldsymbol{q}'_1) - \lambda_{1k} \left\|\boldsymbol{q}'_1\right\|^2 \\
0 & = & \boldsymbol{q}'_k \bullet \boldsymbol{q}'_2 & = & (\boldsymbol{v}_k \bullet \boldsymbol{q}'_2) - \lambda_{2k} \left\|\boldsymbol{q}'_2\right\|^2 \\
 & & \vdots & & \\
0 & = & \boldsymbol{q}'_k \bullet \boldsymbol{q}'_{k-1} & = & (\boldsymbol{v}_k \bullet \boldsymbol{q}'_{k-1}) - \lambda_{k-1,k} \left\|\boldsymbol{q}'_{k-1}\right\|^2
\end{array}$$

y podemos calcular todos los λ_{jk}. Se tiene además que $\langle \boldsymbol{v}_1,\ldots,\boldsymbol{v}_k\rangle = \langle \boldsymbol{q}'_1,\ldots,\boldsymbol{q}'_k\rangle$. Si ahora ponemos $\boldsymbol{q}_i = \frac{1}{\|\boldsymbol{q}'_i\|}\boldsymbol{q}'_i$ conseguimos el conjunto ortonormal. Observemos que la normalización de cada $\boldsymbol{q}'_i$ se puede realizar en cada paso.

Procedimiento de ortogonalización de Gram-Schmidt

Si $\mathscr{S} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_s\}$ es un conjunto de vectores linealmente independiente, entonces la sucesión de Gram-Schmidt definida por

$$\boldsymbol{q}'_1 = \boldsymbol{v}_1, \boldsymbol{q}'_k = \boldsymbol{v}_k - \sum_{i=1}^{k-1} \frac{\boldsymbol{v}_k \bullet \boldsymbol{q}'_i}{\left\|\boldsymbol{q}'_i\right\|^2} \boldsymbol{q}'_i$$

es una base ortogonal de $\mathscr{S}$.
El coste del proceso es del orden de sm^2 flops (sumas y productos).

Si escribimos $\boldsymbol{q}_i = \frac{1}{\|\boldsymbol{q}'_i\|} \boldsymbol{q}'_i, i = 1, 2, \ldots, s$, entonces $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_s\}$ es una base ortonormal de $\mathscr{S}$.

Además, para cada $i = 1, 2, \ldots, s$, se verifica que

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i \rangle = \langle \boldsymbol{q}'_1, \boldsymbol{q}'_2, \ldots, \boldsymbol{q}'_i \rangle = \langle \boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_i \rangle.$$

Ejemplo 6.5.1. Sean

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}.$$

Vamos a calcular una base ortonormal de la variedad lineal $\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle$. El

proceso es:

$$\begin{aligned}
&\boldsymbol{q}_1' = \boldsymbol{v}_1\\
&\boldsymbol{q}_2' = \boldsymbol{v}_2 - \lambda_{12}\boldsymbol{q}_1', \lambda_{12} = \frac{\boldsymbol{v}_2 \bullet \boldsymbol{q}_1'}{\|\boldsymbol{q}_1'\|^2} = 1,\\
&\boldsymbol{q}_2' = \boldsymbol{v}_2 - \boldsymbol{q}_1' = \begin{pmatrix} 0\\ 2\\ 0\\ 0 \end{pmatrix},\\
&\boldsymbol{q}_3' = \boldsymbol{v}_3 - \lambda_{13}\boldsymbol{q}_1' - \lambda_{23}\boldsymbol{q}_2', \lambda_{13} = \frac{\boldsymbol{v}_3 \bullet \boldsymbol{q}_1'}{\|\boldsymbol{q}_1'\|^2} = 2, \lambda_{23} = \frac{\boldsymbol{v}_3 \bullet \boldsymbol{q}_2'}{\|\boldsymbol{q}_2'\|^2} = \frac{1}{2},\\
&\boldsymbol{q}_3' = \boldsymbol{v}_3 - 2\boldsymbol{q}_1' - \tfrac{1}{2}\boldsymbol{q}_2' = \begin{pmatrix} 1\\ 0\\ 1\\ 1 \end{pmatrix}\\
&\boldsymbol{q}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1\\ 0\\ 0\\ -1 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0\\ 1\\ 0\\ 0 \end{pmatrix}, \boldsymbol{q}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1\\ 0\\ 1\\ 1 \end{pmatrix}
\end{aligned}$$

Una consecuencia de lo anterior es que si $\mathscr{S} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s\}$ es un conjunto de vectores ortogonal (ortonormal), entonces puede ampliarse a una base ortogonal (ortonormal) del espacio completo.

En efecto, como los vectores son ortogonales entre sí, forman un conjunto linealmente independiente, por lo que existen $\boldsymbol{v}_{s+1}, \ldots, \boldsymbol{v}_n$ que amplían a una base del espacio. Si aplicamos Gram-Schmidt a este conjunto, deja inalterados los s primeros vectores y obtenemos una base ortogonal del espacio que amplía al conjunto inicial.

Ejemplo 6.5.2. En $\mathbb{R}^4$, consideremos los vectores

$$\boldsymbol{q}_1 = \frac{1}{3}\begin{pmatrix} -2\\ 1\\ 0\\ 2 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0\\ 0\\ 1\\ 0 \end{pmatrix}.$$

Es claro que $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$ es un conjunto ortonormal, y vamos a calcular una base ortonormal de $\mathbb{R}^4$ que lo contenga. En primer lugar, ampliamos a una base de $\mathbb{R}^4$ mediante el uso de la forma escalonada reducida por filas:

$$\begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{pmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 3/2\\ 0 & 1 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0 & 1\\ 0 & 0 & 0 & 1 & 0 & -1/2 \end{bmatrix}.$$

Entonces podemos hacer la ampliación con los vectores $\{e_1, e_2\}$. Ahora aplicamos el procedimiento de Gram-Schmidt al conjunto $\{q_1, q_2, e_1, e_2\}$. Comenzamos directamente con

$$\begin{aligned} q_3' &= e_1 - \lambda_{13} q_1 - \lambda_{23} q_2, \\ \lambda_{13} &= e_1 \bullet q_1 = -\frac{2}{3}, \lambda_{23} = e_1 \bullet q_2 = 0, \\ q_3' &= \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} -2/3 \\ 1/3 \\ 0 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 5/9 \\ 2/9 \\ 0 \\ 4/9 \end{bmatrix}. \end{aligned}$$

Ahora calculamos el siguiente vector:

$$\begin{aligned} q_4' &= e_2 - \lambda_{14} q_1 - \lambda_{24} q_2 - \lambda_{34} q_3', \\ \lambda_{14} &= e_2 \bullet q_1 = \frac{1}{3}, \lambda_{24} = e_2 \bullet q_2 = 0, \lambda_{34} = \frac{e_2 \bullet q_3'}{q_3' \bullet q_3'} = \frac{2}{5}, \\ q_4' &= \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -2/3 \\ 1/3 \\ 0 \\ 2/3 \end{bmatrix} - \frac{2}{5} \begin{bmatrix} 5/9 \\ 2/9 \\ 0 \\ 4/9 \end{bmatrix} = \begin{bmatrix} 0 \\ 4/5 \\ 0 \\ -2/5 \end{bmatrix}. \end{aligned}$$

Ya solamente queda normalizar los vectores q_3' y q_4':

$$q_3 = \frac{1}{\|q_3'\|} q_3' = \begin{bmatrix} 1/3\sqrt{5} \\ 2/15\sqrt{5} \\ 0 \\ \frac{4}{15}\sqrt{5} \end{bmatrix}, q_4 = \frac{1}{\|q_4'\|} q_4' = \begin{bmatrix} 0 \\ 2/5\sqrt{5} \\ 0 \\ -1/5\sqrt{5} \end{bmatrix}.$$

Por tanto, $\{q_1, q_2, q_3, v_4\}$ es una base ortonormal de $\mathbb{R}^4$ que amplía el conjunto inicial. Hay que hacer notar que esta ampliación no es única, y es posible encontrar infinitas bases con estas características.

Más adelante veremos un procedimiento más eficiente para calcular una ampliación de esta clase.

6.6. Factorización QR

El proceso de Gram-Schmidt se puede ver también en la forma de factorización de matrices. Sea $A_{m\times s} = \begin{pmatrix} v_1 & v_2 & \dots & v_s \end{pmatrix}$ una matriz con columnas

independientes. Cuando se aplica Gram-Schmidt a las columnas de A, estamos calculando una base ortonormal $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_s\}$ de $\mathrm{Col}(A)$, donde

$$\begin{aligned}
\boldsymbol{q}'_1 &= \boldsymbol{v}_1, \\
\boldsymbol{q}'_2 &= \boldsymbol{v}_2 - \lambda_{12}\boldsymbol{q}'_1, \\
&\vdots \\
\boldsymbol{q}'_s &= \boldsymbol{v}_s - \lambda_{1s}\boldsymbol{q}'_1 - \cdots - \lambda_{s-1,s}\boldsymbol{q}'_{s-1}.
\end{aligned}$$

Escribamos los vectores $\boldsymbol{v}_i$ en función de los $\boldsymbol{q}_j$. Entonces

$$\begin{array}{rcll}
\boldsymbol{v}_1 &=& \boldsymbol{q}'_1 & = r_{11}\boldsymbol{q}_1, \\
\boldsymbol{v}_2 &=& \lambda_{12}\boldsymbol{q}'_1 + \boldsymbol{q}'_2 & = r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2, \\
&\vdots & & \\
\boldsymbol{v}_s &=& \lambda_{1s}\boldsymbol{q}'_1 + \cdots + \lambda_{s-1,s}\boldsymbol{q}'_{s-1} + \boldsymbol{q}'_s & = r_{1s}\boldsymbol{q}_1 + \cdots + r_{s-1,s}\boldsymbol{q}_{s-1} + r_{ss}\boldsymbol{q}_s,
\end{array}$$

que en forma matricial podemos expresarlo como

$$\begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_s \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \ldots & r_{1s} \\ 0 & r_{22} & \ldots & r_{2s} \\ \vdots & & & \\ 0 & 0 & \ldots & r_{ss} \end{pmatrix}.$$

Observemos que todos los r_{ii} son positivos, pues son normas de vectores. Tenemos así que $A_{m\times s} = \hat{Q}_{m\times s}\hat{R}_{s\times s}$, donde las columnas de Q forman una base ortonormal de $\mathrm{Col}(A)$, y R es una matriz triangular superior con elementos no nulos en la diagonal, esto es, no singular.

A esta descomposición la llamaremos ***factorización QR reducida o rectangular***, porque la matriz $\hat{Q}$ es rectangular y $\hat{R}$ es cuadrada.

Se puede conseguir otra factorización con Q unitaria y R rectangular. Consiste en añadir a R filas de ceros hasta hacerla $m \times s$, y en añadir a Q $m-s$ columnas ortogonales a las anteriores para formar una matriz unitaria. En este caso se la llama ***factorización QR completa***.

En la sección 6.7 explicamos un método para calcular directamente una factorización QR completa de una matriz $A_{m\times n}, m \geq n$, esto es, una expresión de la forma $A = Q_{m\times m}R_{m\times n}$, con Q ortogonal (unitaria) y R de la misma dimensión que A y triangular superior. Para obtener una factorización QR reducida de A a partir de una completa, basta eliminar las $m-n$ últimas columnas de Q y las $m-n$ filas finales de R, que son nulas.

$$\begin{pmatrix} \boxed{\hat{Q}_{m\times m}} & \boxed{m-n} \end{pmatrix} \begin{pmatrix} \boxed{\hat{R}_{n\times n}} \\ \boxed{\mathbf{0}_{n\times(m-n)}} \end{pmatrix}.$$

Ejemplo 6.6.1. Partimos del ejemplo (6.5.1), donde

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix},$$

y habíamos calculado

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{q}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

El cambio de base viene dado por los coeficientes λ_{ij} del proceso, y podemos despejar los vectores $\boldsymbol{v}_i$ en función de los vectores $\boldsymbol{q}_j$.

$$\begin{array}{lllll} \boldsymbol{v}_1 = \|\boldsymbol{q}_1'\| \boldsymbol{q}_1 & & & = \sqrt{2}\boldsymbol{q}_1, & \\ \boldsymbol{v}_2 = \lambda_{12} \|\boldsymbol{q}_1'\| \boldsymbol{q}_1 & + \|\boldsymbol{q}_2'\| \boldsymbol{q}_2 & & = \sqrt{2}\boldsymbol{q}_1 & +2\boldsymbol{q}_2 \\ \boldsymbol{v}_3 = \lambda_{13} \|\boldsymbol{q}_1'\| \boldsymbol{q}_1 & +\lambda_{23} \|\boldsymbol{q}_2'\| \boldsymbol{q}_2 & + \|\boldsymbol{q}_3'\| \boldsymbol{q}_3 & = 2\sqrt{2}\boldsymbol{q}_1 & +\boldsymbol{q}_2 \quad +\sqrt{3}\boldsymbol{q}_3. \end{array}$$

La descomposición QR reducida queda de la forma

$$A = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & 2 & 1 \\ 0 & 0 & \sqrt{3} \end{pmatrix}.$$

Para obtener la factorización QR completa, debemos ampliar el conjunto $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$ a una base ortonormal. Como las variedades son iguales, ampliamos $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ a una base. Para ello, consideramos una forma escalonada por filas de la matriz

$$\begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 & \boldsymbol{e}_4 \end{pmatrix},$$

donde $\boldsymbol{e}_i, i = 1,2,3,4$ son los vectores de la base estándar. Nos queda

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -1/2 & -1/2 & -1 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -2 & 1 \end{pmatrix}$$

lo que indica que podemos tomar para ampliar $\boldsymbol{e}_1$. Sea $\boldsymbol{v}_4 = \boldsymbol{e}_1$. Hay que calcular

$$\boldsymbol{q}_4' = \boldsymbol{v}_4 - \lambda_{14}\boldsymbol{q}_1' - \lambda_{24}\boldsymbol{q}_2' - \lambda_{34}\boldsymbol{q}_3'$$

que sea ortogonal a $\boldsymbol{q}_1', \boldsymbol{q}_2', \boldsymbol{q}_3'$. Efectuando los cálculos llegamos a

$$\lambda_{14} = 1/2, \lambda_{24} = 0, \lambda_{34} = 1/3, \boldsymbol{q}_4' = \begin{pmatrix} 1/6 \\ 0 \\ -1/3 \\ 1/6 \end{pmatrix}, \boldsymbol{q}_4 = \sqrt{6}\boldsymbol{q}_4'.$$

Entonces

$$A = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \boldsymbol{q}_4 \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & 2 & 1 \\ 0 & 0 & \sqrt{3} \\ 0 & 0 & 0 \end{pmatrix}.$$

En el ejemplo anterior vemos que el cálculo de los coeficientes r_{ij} puede resultar arduo por la mezcla de los valores λ_{ij} y las normas de los vectores $\boldsymbol{q}_i'$. Se puede simplificar un poco dicho cálculo con la siguiente observación. La expresión $A = \hat{Q}\hat{R}$ significa que expresamos las columnas de la matriz A como combinación lineal de las columnas de Q. Como forman un conjunto ortonormal, tenemos que

$$\begin{array}{rcllll} \boldsymbol{v}_1 & = & (\boldsymbol{v}_1 \bullet \boldsymbol{q}_1)\boldsymbol{q}_1 & +(\boldsymbol{v}_1 \bullet \boldsymbol{q}_2)\boldsymbol{q}_2 & +\ldots & +(\boldsymbol{v}_1 \bullet \boldsymbol{q}_s)\boldsymbol{q}_s, \\ \boldsymbol{v}_2 & = & (\boldsymbol{v}_2 \bullet \boldsymbol{q}_1)\boldsymbol{q}_1 & +(\boldsymbol{v}_2 \bullet \boldsymbol{q}_2)\boldsymbol{q}_2 & +\ldots & +(\boldsymbol{v}_2 \bullet \boldsymbol{q}_s)\boldsymbol{q}_s, \\ & \vdots & & & & \\ \boldsymbol{v}_s & = & (\boldsymbol{v}_s \bullet \boldsymbol{q}_1)\boldsymbol{q}_1 & +(\boldsymbol{v}_s \bullet \boldsymbol{q}_2)\boldsymbol{q}_2 & +\ldots & +(\boldsymbol{v}_s \bullet \boldsymbol{q}_s)\boldsymbol{q}_s. \end{array}$$

Esto permite escribir la expresión

$$A = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_s \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1 \bullet \boldsymbol{q}_1 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_1 & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_1 \\ \boldsymbol{v}_1 \bullet \boldsymbol{q}_2 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_2 & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_1 \bullet \boldsymbol{q}_s & \boldsymbol{v}_2 \bullet \boldsymbol{q}_s & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_s \end{pmatrix}.$$

Recordemos que en el proceso de Gram-Schmidt, el vector $\boldsymbol{q}_j$ es ortogonal a $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{j-1}$, y por tanto también lo es a $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1}$. Esto significa que los elementos de la parte inferior de la matriz con productos escalares son todos nulos, y escribimos

$$A = \hat{Q}_{m\times s} \begin{pmatrix} \boldsymbol{v}_1 \bullet \boldsymbol{q}_1 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_1 & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_1 \\ 0 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_2 & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_2 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \boldsymbol{v}_s \bullet \boldsymbol{q}_s \end{pmatrix}.$$

Ejemplo 6.6.2. En el ejemplo anterior teníamos

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix},$$

y habíamos calculado

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{q}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Entonces

$$\hat{R} = \begin{pmatrix} \boldsymbol{v}_1 \bullet \boldsymbol{q}_1 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_1 & \boldsymbol{v}_3 \bullet \boldsymbol{q}_1 \\ 0 & \boldsymbol{v}_2 \bullet \boldsymbol{q}_2 & \boldsymbol{v}_3 \bullet \boldsymbol{q}_2 \\ 0 & 0 & \boldsymbol{v}_3 \bullet \boldsymbol{q}_3 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{2} & 2/\sqrt{2} & 4/\sqrt{2} \\ 0 & 2 & 1 \\ 0 & 0 & 3/\sqrt{3} \end{pmatrix}.$$

Nota 6.6.3. El algoritmo de Gram-Schmidt presenta problemas de estabilidad numérica, a causa de los errores de redondeo. Vamos a aplicar el procedimiento al conjunto de vectores

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 10^{-3} \\ 10^{-3} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 10^{-3} \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 10^{-3} \end{pmatrix},$$

con aritmética de 3 dígitos en coma flotante.

$$\begin{aligned}
&\boldsymbol{q}_1' = \boldsymbol{v}_1, \|\boldsymbol{q}_1'\| = 1, \\
&\boldsymbol{q}_2' = \boldsymbol{v}_2 - \lambda_{12}\boldsymbol{q}_1', \lambda_{12} = \tfrac{\boldsymbol{v}_2 \bullet \boldsymbol{q}_1'}{\|\boldsymbol{q}_1'\|^2} = 1, \\
&\boldsymbol{q}_2' = \begin{pmatrix} 0 \\ 0 \\ -10^{-3} \end{pmatrix}, \|\boldsymbol{q}_2'\| = 10^{-3}, \\
&\boldsymbol{q}_3' = \boldsymbol{v}_3 - \lambda_{13}\boldsymbol{q}_1' - \lambda_{23}\boldsymbol{q}_2', \lambda_{13} = 1, \lambda_{23} = -1, \\
&\boldsymbol{q}_3' = \begin{pmatrix} 0 \\ -10^{-3} \\ -10^{-3} \end{pmatrix}, \|\boldsymbol{q}_3'\| = 1{,}41 \times 10^{-3}.
\end{aligned}$$

Entonces

$$\boldsymbol{q}_1 = \begin{pmatrix} 1 \\ 10^{-3} \\ 10^{-3} \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{q}_3 = \begin{pmatrix} 0 \\ -0{,}709 \\ -0{,}709 \end{pmatrix},$$

lo que no es muy satisfactorio, pues $\boldsymbol{q}_2$ y $\boldsymbol{q}_3$ no son ortogonales con esta precisión. Por tanto, vamos a estudiar un método alternativo que nos permita realizar este cálculo con estabilidad numérica.

6.7. * Transformaciones de Householder

Matrices de Householder

Sea $\boldsymbol{v}_{n\times 1}$ un vector no nulo. La matriz

$$H(\boldsymbol{v}) = I_n - 2\frac{\boldsymbol{v}\cdot\boldsymbol{v}^*}{\boldsymbol{v}^*\boldsymbol{v}}$$

de orden n se denomina ***matriz o transformación de Householder*** del vector $\boldsymbol{v}$. También recibe el nombre de reflexión elemental o de Householder.

Figura 6.3: Alston S. Householder (1904-1993)

Observemos que $H(\boldsymbol{v}) = H(\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|})$. cuando $\boldsymbol{v}$ es unitario, tenemos que $H(\boldsymbol{v}) = I - 2\boldsymbol{v}\boldsymbol{v}^*$.

Propiedades de las matrices de Householder

- Si H es una matriz de Householder, entonces es unitaria, hermitiana, e involutiva ($H^2 = I$). Esto es,

$$H = H^* = H^{-1}.$$

- Si $\boldsymbol{x}_{n\times 1}$ es un vector cuya primera componente $x_1 \neq 0$, y si

$$\boldsymbol{u} = \boldsymbol{x} \pm \mu \|\boldsymbol{x}\| \boldsymbol{e}_1, \text{ donde } \mu = \begin{cases} 1 & \text{si } x_1 \text{ es real,} \\ x_1/|x_1| & \text{si } x_1 \text{ no es real,} \end{cases}$$

  se usa para construir la matriz de Householder $H(\boldsymbol{u})$, entonces

$$H(\boldsymbol{u})\boldsymbol{x} = \mp \mu \|\boldsymbol{x}\| \boldsymbol{e}_1.$$

Para evitar cancelaciones cuando se usa aritmética en coma flotante en matrices reales, tomaremos

$$\boldsymbol{u} = \boldsymbol{x} + \text{signo}(x_1) \|\boldsymbol{x}\| \boldsymbol{e}_1.$$

PRUEBA: En primer lugar, $H^* = I_n - 2\boldsymbol{v}\boldsymbol{v}^* = H$, y

$$H^2 = I - 2\boldsymbol{v}\boldsymbol{v}^* - 2\boldsymbol{v}\boldsymbol{v}^* + 4\boldsymbol{v}(\boldsymbol{v}^*\boldsymbol{v})\boldsymbol{v}^* = I.$$

Por último,

$$H(\boldsymbol{u})\boldsymbol{x} = \boldsymbol{x} - 2\frac{\boldsymbol{u}\boldsymbol{u}^*\boldsymbol{x}}{\boldsymbol{u}^*\boldsymbol{u}} = \boldsymbol{x} - 2\frac{\boldsymbol{u}^*\boldsymbol{x}}{\boldsymbol{u}^*\boldsymbol{u}}\boldsymbol{u},$$

y basta probar que $2\boldsymbol{u}^*\boldsymbol{x} = \boldsymbol{u}^*\boldsymbol{u}$, o lo que es lo mismo, $\boldsymbol{u}^*(2\boldsymbol{x} - \boldsymbol{u}) = 0$. Por un lado tenemos que $2\boldsymbol{x} - \boldsymbol{u} = \boldsymbol{x} \mp \mu \|\boldsymbol{x}\| \boldsymbol{e}_1$, y $\boldsymbol{u}^* = \boldsymbol{x}^* \pm \overline{\mu} \|\boldsymbol{x}\| \boldsymbol{e}_1^t$. Entonces

$$\begin{aligned} \boldsymbol{u}^*(2\boldsymbol{x} - \boldsymbol{u}) &= (\boldsymbol{x}^* \pm \overline{\mu} \|\boldsymbol{x}\| \boldsymbol{e}_1^t)(\boldsymbol{x} \mp \mu \|\boldsymbol{x}\| \boldsymbol{e}_1) \\ &= \boldsymbol{x}^*\boldsymbol{x} \mp \mu \|\boldsymbol{x}\| \boldsymbol{x}^*\boldsymbol{e}_1 \pm \overline{\mu} \|\boldsymbol{x}\| \boldsymbol{e}_1^t\boldsymbol{x} - \mu\overline{\mu} \|\boldsymbol{x}\|^2 \cdot 1 \\ &= \|\boldsymbol{x}\|^2 \mp \|\boldsymbol{x}\| \mu\overline{x_1} \pm \|\boldsymbol{x}\| \overline{\mu} x_1 - |\mu|^2 \|\boldsymbol{x}\|^2. \end{aligned}$$

Observemos que $|\mu| = 1$, por lo que

$$\begin{aligned} \boldsymbol{u}^*(2\boldsymbol{x} - \boldsymbol{u}) &= \mp \|\boldsymbol{x}\| \mu\overline{x_1} \pm \|\boldsymbol{x}\| \overline{\mu} x_1 \\ &= \|\boldsymbol{x}\| (\mp \mu\overline{x_1} \pm \overline{\mu} x_1) = 0, \end{aligned}$$

pues

$$-\mu\overline{x_1}+\overline{\mu}x_1=\begin{cases}\text{si } x_1\in\mathbb{R}, -x_1+x_1=0,\\ \text{si } x_1\in\mathbb{C}-\mathbb{R}, -\frac{x_1}{|x_1|}\overline{x_1}+\frac{\overline{x_1}}{|x_1|}x_1=-|x_1|+|x_1|=0.\end{cases}$$

Por tanto, $H(\boldsymbol{u})\boldsymbol{x}=\boldsymbol{x}-\boldsymbol{u}=\mp\mu\|\boldsymbol{x}\|\,\boldsymbol{e}_1$. □

Ejemplo 6.7.1. Dado un vector $\boldsymbol{x}\in\mathbb{C}^n, \|\boldsymbol{x}\|=1$, una forma eficiente de construir una base ortonormal de $\mathbb{C}^n$ que contenga a $\boldsymbol{x}$ es usar las transformaciones de Householder. Consiste en construir una matriz unitaria que tenga a $\boldsymbol{x}$ como primera columna. Sea $\boldsymbol{v}=\boldsymbol{x}\pm\mu\boldsymbol{e}_1$. Entonces $H(\boldsymbol{v})\boldsymbol{x}=\mp\mu\boldsymbol{e}_1$, de donde $\boldsymbol{x}=\mp\mu H(\boldsymbol{v})\boldsymbol{e}_1=[\mp\mu H(\boldsymbol{v})]_{*1}$, la primera columna de $H(\boldsymbol{v})$. Como $|\mp\mu|=1$, la matriz $U=\mp\mu H(\boldsymbol{v})$ es una matriz unitaria con $U_{*1}=\boldsymbol{x}$, por lo que las columnas de U proporcionan la base ortonormal pedida.

Por ejemplo, sea

$$\boldsymbol{x}=\frac{1}{3}\begin{pmatrix}-2\\1\\0\\2\end{pmatrix}$$

y tomemos

$$\boldsymbol{v}=\boldsymbol{x}-\boldsymbol{e}_1=\frac{1}{3}\begin{pmatrix}-5\\1\\0\\2\end{pmatrix}.$$

Entonces

$$H(\boldsymbol{v})=I-2\frac{\boldsymbol{v}\boldsymbol{v}^t}{\boldsymbol{v}^t\boldsymbol{v}}=\begin{pmatrix}-2/3 & 1/3 & 0 & 2/3\\ 1/3 & 14/15 & 0 & -2/15\\ 0 & 0 & 1 & 0\\ 2/3 & -2/15 & 0 & 11/15\end{pmatrix}.$$

6.8. * QR mediante transformaciones de Householder

Para lograr la factorización QR vamos a obtener la matriz Q como producto de transformaciones de Householder. Partimos de una matriz

$$A=\begin{pmatrix}A_{*1} & | & A_{*2} & | & \dots & | & A_{*n}\end{pmatrix}$$

de orden $m\times n$. Sea $\boldsymbol{a}=A_{*1}$ la primera columna de la matriz A. Calculamos la transformación de Householder

$$H_1=I-2\frac{\boldsymbol{v}\boldsymbol{v}^t}{\boldsymbol{v}^t\boldsymbol{v}},\ \text{donde } \boldsymbol{v}=A_{*1}+\text{signo}(a_{11})\,\|A_{*1}\|\,\boldsymbol{e}_1.$$

Entonces

$$H_1 A_{*1} = \mp \|A_{*1}\| \, \boldsymbol{e}_1 = \begin{pmatrix} r_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Aplicamos H_1 a la matriz A y obtenemos

$$\begin{aligned} H_1 A &= \begin{pmatrix} H_1 A_{*1} & | & H_1 A_{*2} & | & \dots & | & H_1 A_{*n} \end{pmatrix} \\ &= \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ 0 & * & \dots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \dots & * \end{pmatrix} \\ &= \begin{pmatrix} t_{11} & \boldsymbol{t}_1^t \\ \mathbf{0} & A_2 \end{pmatrix}, \end{aligned}$$

donde A_2 es de orden $(m-1) \times (n-1)$. Aplicamos el mismo procedimiento a A_2 para construir una transformación de Householder $\hat{H}_2$ que anule todas las entradas por debajo de la posición $(1,1)$ de A_2. Si escribimos $H_2 = \begin{pmatrix} 1 & 0 \\ 0 & \hat{H}_2 \end{pmatrix}$ entonces

$$H_2 H_1 A = \begin{pmatrix} t_{11} & \boldsymbol{t}_1^t \\ \mathbf{0} & \hat{H}_2 A_2 \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ 0 & t_{22} & t_{23} & \dots & t_{2n} \\ 0 & 0 & * & \dots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & * & \dots & * \end{pmatrix}.$$

El resultado tras $k-1$ pasos es $H_{k-1} \dots H_2 H_1 A = \begin{pmatrix} T_{k-1} & U_{k-1} \\ \mathbf{0} & A_k \end{pmatrix}$. En el paso k-ésimo construimos una transformación de Householder $\hat{H}_k$ para hacer ceros por debajo de la posición $(1,1)$ de la matriz A_k, y definimos $H_k = \begin{pmatrix} I_{k-1} & \mathbf{0} \\ \mathbf{0} & \hat{H}_k \end{pmatrix}$. En una de las iteraciones, habremos llegado al número total de filas o de co-

lumnas, y tendremos una de las formas

$$H_n \dots H_2 H_1 A = \begin{pmatrix} * & * & \dots & * \\ 0 & * & \dots & * \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & * \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \text{ cuando } m > n$$

$$H_{m-1} \dots H_2 H_1 A = \begin{pmatrix} * & * & \dots & * & * & \dots & * \\ 0 & * & \dots & * & * & \dots & * \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & * & * & \dots & * \end{pmatrix} \text{ cuando } m < n.$$

Si $m = n$ la forma final es una matriz triangular superior. Como el producto de matrices unitarias (ortogonales) es unitaria (ortogonal), tenemos $SA = R$, con S unitaria (ortogonal). Entonces $A = S^* R = QR$, con Q unitaria (ortogonal).

En las aplicaciones en que se necesita esta descomposición, no es necesario obtener Q de manera explícita, sino que basta con la secuencia de las H_i. Por ejemplo, para resolver el sistema $A\boldsymbol{x} = \boldsymbol{b}$, los pasos son los siguientes:

1. Calcula la factorización QR de A.
2. Calcula $\boldsymbol{y} = Q^* \boldsymbol{b}$.
3. Resuelve $R\boldsymbol{x} = \boldsymbol{y}$.

Lo que necesitamos entonces es el resultado de $Q^* \boldsymbol{b}$, que podemos ir realizando a medida que se van obteniendo las matrices de Householder.

Ejemplo 6.8.1. Vamos a aplicar las transformaciones de Householder para calcular la descomposición QR de la matriz

$$A = \begin{pmatrix} 4 & -3 & 4 \\ 2 & -14 & -3 \\ -2 & 14 & 0 \\ 1 & -7 & 15 \end{pmatrix}.$$

La secuencia es

$$\boldsymbol{u}_1 = A_{*1} + \|A_{*1}\| \boldsymbol{e}_1 = \begin{pmatrix} 9 \\ 2 \\ -2 \\ 1 \end{pmatrix}, H_1 = I - 2\frac{\boldsymbol{u}_1 \boldsymbol{u}_1^t}{\boldsymbol{u}_1^t \boldsymbol{u}_1}.$$

Para calcular $H_1A = \begin{pmatrix} H_1A_{*1} & H_1A_{*2} & H_1A_{*3} \end{pmatrix}$ no es necesario calcular explícitamente H_1. Observemos que

$$H_1A_{*j} = A_{*j} - 2\left(\frac{\boldsymbol{u}_1^tA_{*j}}{\boldsymbol{u}_1^t\boldsymbol{u}_1}\right)\boldsymbol{u}_1$$

por lo que basta calcular $\boldsymbol{u}_1^tA_{*j}$, $j = 1,2,3$. Nos queda

$$H_1A = \begin{pmatrix} -5 & 15 & -5 \\ 0 & -10 & -5 \\ 0 & 10 & 2 \\ 0 & -5 & 14 \end{pmatrix}, H_1 = \begin{bmatrix} -4/5 & -2/5 & 2/5 & -1/5 \\ -2/5 & \frac{41}{45} & \frac{4}{45} & -\frac{2}{45} \\ 2/5 & \frac{4}{45} & \frac{41}{45} & \frac{2}{45} \\ -1/5 & -\frac{2}{45} & \frac{2}{45} & \frac{44}{45} \end{bmatrix}.$$

Sea ahora

$$A_2 = \begin{pmatrix} -10 & -5 \\ 10 & 2 \\ -5 & 14 \end{pmatrix} \text{ y } \boldsymbol{u}_2 = [A_2]_{*1} - \|[A_2]_{*1}\|\,\boldsymbol{e}_1 = \begin{pmatrix} -25 \\ 10 \\ -5 \end{pmatrix}.$$

Si $\hat{H}_2 = I - 2\frac{\boldsymbol{u}_2\boldsymbol{u}_2^t}{\boldsymbol{u}_2^t\boldsymbol{u}_2}$ y $H_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{H}_2 \end{pmatrix}$, entonces

$$\hat{H}_2A_2 = \begin{pmatrix} 15 & 0 \\ 0 & 0 \\ 0 & 15 \end{pmatrix} \text{ y } H_2H_1A = \begin{pmatrix} -5 & 15 & -5 \\ 0 & 15 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 15 \end{pmatrix}, H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -2/3 & 2/3 & -1/3 \\ 0 & 2/3 & \frac{11}{15} & 2/15 \\ 0 & -1/3 & 2/15 & \frac{14}{15} \end{bmatrix}.$$

Tomamos a continuación

$$A_3 = \begin{pmatrix} 0 \\ 15 \end{pmatrix} \text{ y } \boldsymbol{u}_3 = [A_3]_{*1} + \|[A_3]_{*1}\|\,\boldsymbol{e}_1 = \begin{pmatrix} 15 \\ 15 \end{pmatrix}.$$

Si $\hat{H}_3 = I - 2\frac{\boldsymbol{u}_3\boldsymbol{u}_3^t}{\boldsymbol{u}_3^t\boldsymbol{u}_3}$ y $H_3 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \hat{H}_3 \end{pmatrix}$, entonces

$$\hat{H}_3A_3 = \begin{pmatrix} -15 \\ 0 \end{pmatrix} \text{ y } H_3H_2H_1A = \begin{pmatrix} -5 & 15 & -5 \\ 0 & 15 & 0 \\ 0 & 0 & -15 \\ 0 & 0 & 0 \end{pmatrix}, H_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

En este caso,

$$Q^t = H_3 H_2 H_1 = \begin{pmatrix} -4/5 & -2/5 & 2/5 & -1/5 \\ -3/5 & -8/15 & 8/15 & -4/15 \\ 0 & 1/3 & -2/15 & -14/15 \\ 0 & -2/3 & -11/15 & -2/15 \end{pmatrix}.$$

6.9. * Estabilidad y coste de la ortogonalización

Un algoritmo se considera ***numéricamente estable*** si, bajo aritmética en coma flotante, siempre devuelve una respuesta que es solución exacta de un problema *cercano*. La reducción de Householder es un algoritmo estable para producir la factorización QR de $A_{n\times n}$.

Ejemplo 6.9.1. En un ejemplo anterior veíamos los problemas de estabilidad numérica que presentaba Gram-Schmidt para calcular la factorización QR. Partimos de los vectores

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 10^{-3} \\ 10^{-3} \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 10^{-3} \\ 0 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 10^{-3} \end{pmatrix},$$

con aritmética de 3 dígitos en coma flotante. Sea $A = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix}$. Entonces

$$\boldsymbol{u}_1 = A_{*1} + \text{signo}(1)\,\|A_{*1}\|\,\boldsymbol{e}_1 = \begin{pmatrix} 2{,}00 \\ 0{,}00100 \\ 0{,}00100 \end{pmatrix},$$

$$H_1 = H(\boldsymbol{u}_1) = \begin{pmatrix} -1{,}0 & -0{,}00100 & -0{,}00100 \\ -0{,}00100 & 1{,}0 & -0{,}500\cdot 10^{-6} \\ -0{,}00100 & -0{,}500\cdot 10^{-6} & 1{,}0 \end{pmatrix}.$$

Entonces

$$H_1 A = \begin{bmatrix} -1{,}0 & -1{,}0 & -1{,}0 \\ -0{,}500\cdot 10^{-9} & 0{,}0 & -0{,}00100 \\ -0{,}500\cdot 10^{-9} & -0{,}00100 & 0{,}0 \end{bmatrix}.$$

Observemos que en la primera columna hay valores que deberían ser nulos, pero tienen un valor muy pequeño con respecto a las otras entradas de la matriz. Sea ahora

$$A_2 = \begin{pmatrix} 0{,}0 & -0{,}00100 \\ -0{,}00100 & 0{,}0 \end{pmatrix} \text{ y } \boldsymbol{u}_2 = [A_2]_{*1} + \text{signo}(0)\,\|[A_2]_{*1}\|\,\boldsymbol{e}_1 = \begin{pmatrix} 0{,}00100 \\ -0{,}00100 \end{pmatrix}.$$

Ahora tenemos que

$$\hat{H}_2 = H(\boldsymbol{u}_2) = \begin{pmatrix} 0 & 1{,}00 \\ 1{,}00 & 0 \end{pmatrix} \text{ y } H_2 H_1 A = \begin{bmatrix} -1{,}0 & -1{,}0 & -1{,}0 \\ -0{,}500 \cdot 10^{-9} & -0{,}00100 & 0{,}0 \\ -0{,}500 \cdot 10^{-9} & 0{,}0 & -0{,}00100 \end{bmatrix}.$$

Entonces

$$Q = H_1 H_2 = \begin{pmatrix} -1{,}0 & -0{,}00100 & -0{,}00100 \\ -0{,}00100 & -0{,}500 \cdot 10^{-6} & 1{,}0 \\ -0{,}00100 & 1{,}0 & -0{,}500 \cdot 10^{-6} \end{pmatrix} \text{ y}$$

$$Q^t Q = \begin{bmatrix} 1{,}0 & 0{,}500 \cdot 10^{-9} & 0{,}500 \cdot 10^{-9} \\ 0{,}500 \cdot 10^{-9} & 1{,}0 & 0{,}0 \\ 0{,}500 \cdot 10^{-9} & 0{,}0 & 1{,}0 \end{bmatrix},$$

que es un resultado mucho mejor que el obtuvimos con Gram-Schmidt.

La eliminación gaussiana no es un algoritmo estable porque surgen problemas debido al crecimiento de magnitud de los números que aparecen en el proceso. Sin embargo, si se usa pivoteo *completo* sobre una matriz $A_{n\times n}$ bien escalada para la que máx$|a_{ij}| = 1$, entonces los coeficientes que aparecen tiene un crecimiento muy lento respecto a n, por lo que se puede garantizar la estabilidad del algoritmo. Por tanto, la eliminación gaussiana con pivoteo completo es estable, pero con pivoteo parcial no. Por fortuna, en el trabajo práctico es raro encontrar matrices en el que el pivoteo parcial falle en el control del crecimiento de los coeficientes, por lo que pivoteo parcial se considera, en general, como un algoritmo "prácticamente" estable.

Los algoritmos que se basan en Gram-Schmidt son más complicados. En primer lugar, el algoritmo de Gram-Schmidt difiere del de Householder en que no se aplica una sucesión de transformaciones elementales ortogonales. En segundo lugar, como algoritmo para producir la factorización QR, puede devolver un factor Q que esté lejos de ser ortogonal, y el argumento intuitivo de estabilidad numérica usado anteriormente no es válido. Existe una versión modificada de Gram-Schmidt que sigue siendo no estable para la factorización QR general, pero se puede demostrar que es estable para el tratamiento del problema de mínimos cuadrados.

Sumario de estabilidad numérica

- La eliminación gaussiana con pivoteo parcial y escalado es teóricamente no estable, pero es prácticamente estable, es decir, estable para la mayoría de problemas.
- El pivoteo completo hace a la eliminación gaussiana estable sin condiciones.
- Para la factorización QR, el procedimiento de Gram-Schmidt (clásico o modificado) no es estable. Sin embargo, el procedimiento modificado es un algoritmo estable para resolver mínimos cuadrados.
- La reducción de Householder es estable sin condiciones para el cálculo de la factorización QR.

Coste comparado de la factorización QR

El número aproximado de flops que se requieren para reducir una matriz $n \times n$ a una matriz triangular superior es como sigue:

- Eliminación gaussiana (escalado y pivoteo parcial) $\approx \frac{2}{3}n^3$.
- Procedimiento de Gram-Schmidt (clásico y modificado) $\approx 2n^3$.
- Reducción de Householder $\approx \frac{4}{3}n^3$.

No es sorprendente que los métodos estables sin condiciones sean más costosos. Ninguna técnica de triangulación se puede considerar óptima, y cada una tiene un lugar en el día a día. Por ejemplo, para resolver sistemas lineales en donde la matriz no presenta alguna estructura, la probabilidad de que la eliminación gaussiana con pivoteo parcial y escalado falle no es lo bastante alta para justificar el empleo de Householder, o incluso pivoteo completo. Para mínimos cuadrados se usa Householder o Gram-Schmidt modificado. Para ortogonalizar Col(A), donde A es una matriz sin una estructura determinada y densa, se usa la reducción de Householder.

6.10. Descomposición ortogonal

Complemento ortogonal

Para un subconjunto $\mathcal{M}$ de un espacio vectorial euclídeo $\mathcal{V}$, el ***complemento ortogonal*** $\mathcal{M}^\perp$ de $\mathcal{M}$ es el conjunto de todos los vectores de $\mathcal{V}$ que son ortogonales a todos los vectores de $\mathcal{M}$. Esto es,

$$\mathcal{M}^\perp = \{\boldsymbol{x} \in \mathcal{V} \mid \boldsymbol{m} \bullet \boldsymbol{x} = 0 \text{ para todo } \boldsymbol{m} \in \mathcal{M}\}.$$

Vamos a probar en primer lugar que $\mathcal{M}^\perp$ es un subespacio vectorial de $\mathcal{V}$. Sean $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{M}^\perp$. Entonces

$$(\boldsymbol{v}_1 + \boldsymbol{v}_2) \bullet \boldsymbol{m} = \boldsymbol{v}_1 \bullet \boldsymbol{m} + \boldsymbol{v}_2 \bullet \boldsymbol{m} = 0 \text{ para todo } \boldsymbol{m} \in \mathcal{M},$$

y

$$(\alpha \boldsymbol{v}_1) \bullet \boldsymbol{m} = \alpha(\boldsymbol{v}_1 \bullet \boldsymbol{m}) = 0 \text{ para todo } \boldsymbol{m} \in \mathcal{M}.$$

Esto es independiente de la estructura de $\mathcal{M}$. Sin embargo, el caso que nos interesa especialmente es cuando $\mathcal{M}$ es un subespacio de $\mathcal{V}$. En tal caso, existe una base finita de $\mathcal{M}$ formada por los vectores $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_r$. La definición impone, en principio, un conjunto infinito de condiciones para caracterizar a los elementos de $\mathcal{M}^\perp$. Sin embargo, vamos a ver que

$$\mathcal{M}^\perp = \{\boldsymbol{x} \in \mathcal{V} \mid \boldsymbol{m}_i \bullet \boldsymbol{x} = 0 \text{ para todo } i = 1, \ldots, r\}.$$

Es claro que si $\boldsymbol{x} \in \mathcal{M}^\perp$, entonces está en el conjunto de la derecha. Si ahora $\boldsymbol{m}_i \bullet \boldsymbol{x} = 0$ para todo $i = 1, \ldots, r$, consideremos un vector $\boldsymbol{m} \in \mathcal{M}$. Entonces $\boldsymbol{m}$ se puede expresar como combinación lineal de $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_r$, esto es, $\boldsymbol{m} = \sum_{i=1}^r \alpha_i \boldsymbol{m}_i$, y

$$\boldsymbol{m} \bullet \boldsymbol{x} = \left(\sum_{i=1}^r \alpha_i \boldsymbol{m}_i \right) \bullet \boldsymbol{x} = \sum_{i=1}^r \alpha_i (\boldsymbol{m}_i \bullet \boldsymbol{x}) = 0.$$

Tenemos así el resultado.

Introducimos aquí una notación. Decimos que un espacio L es ***suma directa*** de L_1 y L_2, notado por $L = L_1 \oplus L_2$ si

$$L = L_1 + L_2 \text{ y } \{\mathbf{0}\} = L_1 \cap L_2.$$

En tal caso,

$$\dim L = \dim L_1 + \dim L_2 - \dim(L_1 \cap L_2) = \dim L_1 + \dim L_2.$$

Complemento ortogonal de un subespacio

Si $\mathscr{M}$ es un subespacio de un espacio euclídeo de dimensión finita $\mathscr{V}$, entonces

$$\mathscr{V} = \mathscr{M} \oplus \mathscr{M}^{\perp}.$$

Además, si $\mathscr{N}$ es un subespacio tal que $\mathscr{V} = \mathscr{M} \oplus \mathscr{N}$ y $\mathscr{N} \perp \mathscr{M}$, entonces

$$\mathscr{N} = \mathscr{M}^{\perp}.$$

PRUEBA: Si $\boldsymbol{v} \in \mathscr{M} \cap \mathscr{M}^{\perp}$, entonces $\boldsymbol{v}$ es un vector ortogonal a sí mismo, es decir, $\boldsymbol{v} \bullet \boldsymbol{v} = 0$, de donde $\boldsymbol{v} = \boldsymbol{0}$. Sea $\mathscr{B}_{\mathscr{M}} = \{\boldsymbol{m}_1, \ldots, \boldsymbol{m}_r\}$ una base ortonormal de $\mathscr{M}$. La ampliamos a una base ortonormal de $\mathscr{V}$, mediante Gram-Schmidt, con la forma $\mathscr{B} = \{\boldsymbol{m}_1, \ldots, \boldsymbol{m}_r, \boldsymbol{m}_{r+1}, \ldots, \boldsymbol{m}_n\}$. Tenemos entonces que $\boldsymbol{m}_i \bullet \boldsymbol{m}_j = 0$ para todo $i = 1, \ldots, r$, $j = r+1, \ldots, n$.

Vamos a probar que $\mathscr{M}^{\perp} = \langle \boldsymbol{m}_{r+1}, \ldots, \boldsymbol{m}_n \rangle$. Si $\boldsymbol{v} \in \mathscr{M}^{\perp}$ entonces $\boldsymbol{m}_i \bullet \boldsymbol{v} = 0$ para todo $i = 1, \ldots, r$. Como $\mathscr{B}$ es una base de $\mathscr{V}$, el vector $\boldsymbol{v}$ se puede expresar como combinación lineal de los vectores de dicha base:

$$\boldsymbol{v} = \alpha_1 \boldsymbol{m}_1 + \cdots + \alpha_r \boldsymbol{m}_r + \alpha_{r+1} \boldsymbol{m}_{r+1} + \cdots + \alpha_n \boldsymbol{m}_n.$$

Entonces, para todo $i = 1, \ldots, r$,

$$0 = \boldsymbol{m}_i \bullet \boldsymbol{v} = \overline{\alpha_i}, \text{ de donde } \alpha_i = 0,$$

y esto implica que $\boldsymbol{v} \in \langle \boldsymbol{m}_{r+1}, \ldots, \boldsymbol{m}_n \rangle$. Recíprocamente, si $\boldsymbol{v} \in \langle \boldsymbol{m}_{r+1}, \ldots, \boldsymbol{m}_n \rangle$, entonces

$$\boldsymbol{v} = \alpha_{r+1} \boldsymbol{m}_{r+1} + \cdots + \alpha_n \boldsymbol{m}_n,$$

y para cada $i = 1, \ldots, r$ se verifica

$$\boldsymbol{m}_i \bullet \boldsymbol{v} = \boldsymbol{m}_i \left(\sum_{j=r+1}^{n} \alpha_j \boldsymbol{m}_j \right) = \sum_{j=r+1}^{n} \overline{\alpha_j} \boldsymbol{m}_i \bullet \boldsymbol{m}_j = 0.$$

Entonces $\boldsymbol{v} \in \mathscr{M}^{\perp}$.

Para la segunda parte del enunciado, observemos que si $\mathscr{N} \perp \mathscr{M}$, entonces $\mathscr{N} \subset \mathscr{M}^{\perp}$. Por otro lado, de $\dim \mathscr{N} = \dim \mathscr{V} - \dim \mathscr{M} = \dim \mathscr{M}^{\perp}$, deducimos que las dimensiones coinciden. Entonces, son iguales. □

Ejemplo 6.10.1. En el ejemplo (6.5.1) habíamos obtenido una factorización QR reducida de un conjunto de vectores, y habíamos detallado un método para encontrar una factorización QR completa mediante la ampliación de una base. Vamos a aprovechar el complemento ortogonal para calcular dicha factorización de una manera algo más sencilla.

Partimos entonces de

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix},$$

y sabemos que los vectores

$$\boldsymbol{q}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{q}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

constituyen una base ortonormal del espacio $\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle$. Para el cálculo de una factorización QR completa, precisamos ampliar el conjunto $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$ a una base ortonormal de $\mathbb{R}^4$. Por ello, obtenemos una base del espacio $\langle \boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3 \rangle^{\perp}$, que se corresponde a la resolución del sistema lineal homogéneo

$$\begin{cases} \frac{1}{\sqrt{2}}x_1 & & & -\frac{1}{\sqrt{2}}x_4 & = & 0, \\ & x_2 & & & = & 0, \\ \frac{1}{\sqrt{3}}x_1 & & +\frac{1}{\sqrt{3}}x_3 & +\frac{1}{\sqrt{3}}x_4 & = & 0. \end{cases}$$

Como siempre,

$$\begin{bmatrix} 1/2\sqrt{2} & 0 & 0 & -1/2\sqrt{2} \\ 0 & 1 & 0 & 0 \\ 1/3\sqrt{3} & 0 & 1/3\sqrt{3} & 1/3\sqrt{3} \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \Rightarrow \begin{cases} x_1 & = & x_4, \\ x_2 & = & 0, \\ x_3 & = & -2x_4, \\ x_4 & = & x_4. \end{cases}$$

Entonces un vector ortogonal a $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ es

$$\boldsymbol{q}_4' = \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \end{bmatrix}.$$

Antes de continuar, observemos que el cálculo del espacio ortogonal $\langle \boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3 \rangle^\perp$ es equivalente al cálculo del espacio $\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle^\perp$, por la construcción de los vectores $\boldsymbol{q}_i$. Por ello, podíamos haber resuelto el sistema

$$\left\{\begin{array}{rrrrcl} x_1 & & & -x_4 & = & 0, \\ x_1 & +2x_2 & & -x_4 & = & 0, \\ 3x_1 & +x_2 & +x_3 & -x_4 & = & 0. \end{array}\right.$$

Por completar,

$$\left[\begin{array}{cccc} 1 & 0 & 0 & -1 \\ 1 & 2 & 0 & -1 \\ 3 & 1 & 1 & -1 \end{array}\right] \overset{\text{rref}}{\longrightarrow} \left[\begin{array}{cccc} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right],$$

y obtenemos la misma solución.

Ahora debemos aplicar el procedimiento de Gram-Schmidt para obtener una base ortonormal de $\langle \boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3 \rangle^\perp$. Como está generado por un único vector $\boldsymbol{q}_4'$, basta normalizarlo:

$$\boldsymbol{q}_4 = \frac{1}{\|\boldsymbol{q}_4'\|}\boldsymbol{q}_4' = \begin{bmatrix} 1/6\sqrt{6} \\ 0 \\ -1/3\sqrt{6} \\ 1/6\sqrt{6} \end{bmatrix}.$$

Por tanto, una factorización QR completa de la matriz $A = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix}$ es

$$A = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \boldsymbol{q}_4 \end{pmatrix} \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & 2 & 1 \\ 0 & 0 & \sqrt{3} \\ 0 & 0 & 0 \end{pmatrix}.$$

Ejemplo 6.10.2. Vamos a aplicar el método anterior para la ampliación a una base ortonormal del ejemplo (6.5.2). Partíamos de los vectores

$$\boldsymbol{q}_1 = \frac{1}{3}\begin{pmatrix} -2 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \boldsymbol{q}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

que ya forman un conjunto ortonormal. Calculamos una base de su espacio ortogonal $\langle \boldsymbol{q}_1, \boldsymbol{q}_2 \rangle^\perp$:

$$\left\{\begin{array}{rrrrcl} \frac{-2}{3}x_1 & +\frac{1}{3}x_2 & & +\frac{2}{3}x_4 & = & 0, \\ & & x_3 & & = & 0. \end{array}\right. \Rightarrow \left\{\begin{array}{rrrrcl} x_1 & -\frac{1}{2}x_2 & & -x_4 & = & 0, \\ & & x_3 & & = & 0. \end{array}\right.$$

Entonces podemos escribir

$$\left\{\begin{array}{lcll} x_1 & = & \frac{1}{2}x_2 & +x_4, \\ x_2 & = & x_2, & \\ x_3 & = & 0, & \\ x_4 & = & & x_4. \end{array}\right. \quad \text{y obtenemos } \langle \boldsymbol{q}_1, \boldsymbol{q}_2\rangle^{\perp} = \langle \boldsymbol{w}_3 = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{w}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\rangle.$$

Ahora aplicamos Gram-Schmidt al conjunto $\{\boldsymbol{w}_3, \boldsymbol{w}_4\}$:

$$\begin{aligned} \boldsymbol{q}_3' &= \boldsymbol{w}_3, \\ \boldsymbol{q}_4' &= \boldsymbol{w}_4 - \lambda \boldsymbol{q}_3', \end{aligned}$$

$$\lambda = \frac{\boldsymbol{w}_4 \bullet \boldsymbol{q}_3'}{\boldsymbol{q}_3' \bullet \boldsymbol{q}_3'} = \frac{2}{5},$$

$$\boldsymbol{q}_4' = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{5}\begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} 4/5 \\ -2/5 \\ 0 \\ 1 \end{bmatrix}.$$

Tras normalizar,

$$\boldsymbol{q}_3 = \frac{1}{\|\boldsymbol{q}_3'\|}\boldsymbol{q}_3' = \begin{bmatrix} 1/5\sqrt{5} \\ 2/5\sqrt{5} \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{q}_4 = \frac{1}{\|\boldsymbol{q}_4'\|}\boldsymbol{q}_4' = \begin{bmatrix} \frac{4}{15}\sqrt{5} \\ -2/15\sqrt{5} \\ 0 \\ 1/3\sqrt{5} \end{bmatrix}.$$

Propiedades del complemento ortogonal

Si $\mathcal{M}_1, \mathcal{M}_2$ son subespacios de un espacio vectorial euclídeo de dimensión n, entonces

- $\mathcal{M}_1^{\perp\perp} = \mathcal{M}_1$.
- $(\mathcal{M}_1 + \mathcal{M}_2)^{\perp} = \mathcal{M}_1^{\perp} \cap \mathcal{M}_2^{\perp}$.
- $(\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp} = \mathcal{M}_1^{\perp} + \mathcal{M}_2^{\perp}$.

PRUEBA:

- Sea $\boldsymbol{v} \in \mathscr{M}_1^{\perp^\perp}$. Como $\mathscr{V} = \mathscr{M}_1 \oplus \mathscr{M}_1^\perp$, entonces $\boldsymbol{v} = \boldsymbol{m} + \boldsymbol{n}$, con $\boldsymbol{m} \in \mathscr{M}_1$ y $\boldsymbol{n} \in \mathscr{M}_1^\perp$. De aquí,

$$0 = \boldsymbol{v} \bullet \boldsymbol{n} = \boldsymbol{m} \bullet \boldsymbol{n} + \boldsymbol{n} \bullet \boldsymbol{n} = \boldsymbol{n} \bullet \boldsymbol{n} \Rightarrow \boldsymbol{n} = \boldsymbol{0},$$

  y tenemos que $\boldsymbol{v} \in \mathscr{M}_1$.

- Observemos que

$$\begin{aligned} \boldsymbol{v} \in (\mathscr{M}_1 + \mathscr{M}_2)^\perp \Leftrightarrow & \quad \boldsymbol{v} \perp \mathscr{M}_1 + \mathscr{M}_2 \\ \Leftrightarrow & \quad \boldsymbol{v} \perp \mathscr{M}_1 \text{ y } \boldsymbol{v} \perp \mathscr{M}_2 \\ \Leftrightarrow & \quad \boldsymbol{v} \in (\mathscr{M}_1^\perp \cap \mathscr{M}_2^\perp). \end{aligned}$$

- Aplicamos lo anterior para obtener

$$(\mathscr{M}_1^\perp + \mathscr{M}_2^\perp)^\perp = \mathscr{M}_1^{\perp^\perp} \cap \mathscr{M}_2^{\perp^\perp} = \mathscr{M}_1 \cap \mathscr{M}_2.$$

□

Ejemplo 6.10.3. Este teorema proporciona un método para el cálculo de la intersección de dos variedades lineales. Consideremos los subespacios de $\mathbb{R}^4$ dados por

$$U = \langle \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \boldsymbol{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \rangle, V = \langle \boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} \rangle.$$

Calculamos en primer lugar $U^\perp$ y $V^\perp$. El conjunto $U^\perp$ es el espacio de soluciones del sistema lineal homogéneo

$$\begin{cases} x_1 + 2x_2 + x_3 + 2x_4 &= 0, \\ x_2 + x_4 &= 0, \end{cases}$$

por lo que calculamos la forma escalonada reducida por filas de la matriz de coeficientes:

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{pmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \begin{cases} x_1 &= -x_3, \\ x_2 &= -x_4, \\ x_3 &= x_3, \\ x_4 &= x_4, \end{cases},$$

$$U^\perp = \langle \boldsymbol{u}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{u}_4 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \rangle.$$

Análogamente para $V^\perp$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \begin{cases} x_1 &= -x_3 & -2x_4, \\ x_2 &= & x_4, \\ x_3 &= x_3, & \\ x_4 &= & x_4, \end{cases},$$

$$V^\perp = \langle \boldsymbol{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_4 = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \rangle.$$

Observemos que lo que se ha hecho es calcular el espacio nulo del espacio de filas, tal como se indica en el teorema.

El subespacio vectorial $U^\perp + V^\perp$ está generado por los vectores $\{\boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{v}_3, \boldsymbol{v}_4\}$. Ahora tenemos en cuenta la relación $U \cap V = (U^\perp + V^\perp)^\perp$, por lo que calculamos el espacio ortogonal a este conjunto de vectores:

$$\begin{pmatrix} \boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{v}_3, \boldsymbol{v}_4 \end{pmatrix}^t = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ -2 & 1 & 0 & 1 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Entonces

$$U \cap V = (U^\perp + V^\perp)^\perp = \langle \boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \rangle.$$

Teorema de descomposición ortogonal

Si $A_{m\times n}$ es una matriz real,

$$\text{Col}(A)^\perp = \text{null}(A^t) \text{ y } \text{null}(A)^\perp = \text{Col}(A^t).$$

PRUEBA: Tenemos las siguientes equivalencias:

$$\begin{aligned} \boldsymbol{x} \in \text{Col}(A)^\perp &\Leftrightarrow \boldsymbol{x} \bullet A\boldsymbol{y} = 0 \text{ para todo } \boldsymbol{y} \in \mathbb{R}^n \\ &\Leftrightarrow \boldsymbol{y}^t A^t \boldsymbol{x} = 0 \text{ para todo } \boldsymbol{y} \in \mathbb{R}^n, \text{ en particular los } \boldsymbol{e}_i \\ &\Leftrightarrow A^t \boldsymbol{x} = \boldsymbol{0} \\ &\Leftrightarrow \boldsymbol{x} \in \text{null}(A^t). \end{aligned}$$

Con esto hemos probado la primera parte. Si se la aplicamos a A^t, obtenemos $\text{Col}(A^t)^\perp = \text{null}(A)$, y tomamos el complemento ortogonal a cada lado para obtener la segunda. □

Igualdad de espacios nulos

Para dos matrices A y B de la misma forma,

$$\text{null}(A) = \text{null}(B) \text{ si y solamente si } A \overset{\text{f}}{\sim} B.$$

PRUEBA: Podemos escribir que $\boldsymbol{v} \in \text{null}(A) \Leftrightarrow A\boldsymbol{v} = \mathbf{0} \Leftrightarrow \boldsymbol{v} \in \text{Col}(A^t)^\perp$. Entonces $\text{Col}(A^t) = \text{Col}(B^t)$, que es lo mismo que decir $A \overset{\text{f}}{\sim} B$. □

Capítulo 7

Autovalores y autovectores

7.1. Propiedades elementales

El objetivo en este tema es calcular una base respecto de la cual la aplicación lineal representada por una matriz A sea lo más sencilla posible. Para el estudio de sistemas dinámicos discretos nos permitirá averiguar el comportamiento de las potencias de una matriz. En estadística, averiguaremos la estructura de las matrices de covarianza.

Autovalores y autovectores

Para una matriz A de orden $n \times n$, los escalares λ y los vectores $\boldsymbol{x}_{n\times 1} \neq \mathbf{0}$ que satisfacen

$$A\boldsymbol{x} = \lambda \boldsymbol{x}$$

se denominan ***autovalores*** y ***autovectores*** de A, respectivamente. El conjunto de autovalores *distintos*, notado por $\sigma(A)$, se denomina *espectro* de A. Dado λ autovalor, el conjunto de autovectores asociados es

$$\{\boldsymbol{x} \neq \mathbf{0} \mid \boldsymbol{x} \in \text{null}(\lambda I - A)\} = \{\boldsymbol{x} \neq \mathbf{0} \mid \boldsymbol{x} \in \text{null}(A - \lambda I)\}.$$

Los autovalores también reciben el nombre de raíces características o latentes, y algo análogo para los autovectores. Si $\boldsymbol{v}$ es un autovector, con λ su autovalor asociado, nos referiremos a ellos como el autopar $(\lambda, \boldsymbol{v})$. Al espacio $\text{null}(\lambda I - A)$ lo notaremos por $V_1(\lambda)$, supuesta dada la matriz A. Contiene los autovectores asociados al autovalor λ y el vector nulo; es un subespacio vectorial.

Polinomio característico y ecuación característica

- El ***polinomio característico*** de $A_{n\times n}$ es $p(\lambda) = \det(\lambda I - A)$. El grado de $p(\lambda)$ es n, y su término líder es λ^n.
- La ***ecuación característica*** de A es $p(\lambda) = 0$.

Ejemplo 7.1.1. El polinomio característico de la matriz

$$A = \begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix}$$

es

$$\begin{aligned} \det(\lambda I - A) &= \begin{pmatrix} \lambda - 1 & 4 & 4 \\ -8 & \lambda + 11 & 8 \\ 8 & -8 & \lambda - 5 \end{pmatrix} \\ &= (\lambda - 1)(\lambda + 3)^2. \end{aligned}$$

Autovalores como raíces

- Los autovalores de A son las soluciones de la ecuación característica, esto es, las raíces del polinomio característico.
- En su conjunto, A tiene n autovalores, pero algunos pueden ser complejos (aunque A tenga entradas reales), y algunos autovalores pueden estar repetidos.
- Si A es una matriz real, entonces sus autovalores complejos no reales vienen en pares conjugados, es decir, si $\lambda \in \sigma(A)$ entonces $\overline{\lambda} \in \sigma(A)$, con la misma multiplicidad.
- El polinomio característico es invariante para matrices semejantes.

PRUEBA: La primera condición se tiene por la equivalencia entre el carácter nulo del determinante de una matriz y la existencia de soluciones no triviales de su espacio nulo. Las condiciones dos y tres son genéricas de polinomios.

Para la última afirmación, sea B semejante a la matriz A; entonces existe P no singular tal que $B = P^{-1}AP$, y

$$\begin{aligned}\det(\lambda I - B) &= \det(\lambda I - P^{-1}AP) = \det(P^{-1}(\lambda I - A)P)\\ &= \det(P^{-1})\det(\lambda I - A)\det(P) = \det(\lambda I - A).\end{aligned}$$

□

Dado un autovalor λ, todos los elementos no nulos de $\text{null}(A-\lambda I) = \text{null}(\lambda I - A)$ son autovectores asociados. Por ello, para calcularlos, debemos resolver un sistema lineal homogéneo.

Ejemplo 7.1.2. Consideremos la matriz

$$A = \begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix}, \text{ y } \det(\lambda I - A) = \det\begin{pmatrix} \lambda - 7 & 4 \\ -5 & \lambda + 2 \end{pmatrix} = (\lambda - 2)(\lambda - 3).$$

Los autovalores son $\lambda_1 = 2, \lambda_2 = 3$. Calculemos los espacios de autovectores asociados.

Para $\lambda_1 = 2$,

$$A - \lambda_1 I = \begin{pmatrix} 5 & -4 \\ 5 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & -4/5 \\ 0 & 0 \end{pmatrix} \to \begin{cases} x_1 = 4/5 x_2 \\ x_2 \text{ libre} \end{cases} \to \text{null}(A - \lambda_1 I) = \langle \boldsymbol{v}_1 \rangle,$$

donde

$$\boldsymbol{v}_1 = \begin{pmatrix} 4/5 \\ 1 \end{pmatrix}.$$

Para $\lambda_2 = 3$,

$$A - \lambda_2 I = \begin{pmatrix} 4 & -4 \\ 5 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \to \begin{cases} x_1 = x_2 \\ x_2 \text{ libre} \end{cases} \to \text{null}(A - \lambda_2 I) = \langle \boldsymbol{v}_2 \rangle,$$

donde

$$\boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Ejemplo 7.1.3. Consideremos la matriz

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \text{ y } \det(A - \lambda I) = \det\begin{pmatrix} 1 - \lambda & -1 \\ 1 & 1 - \lambda \end{pmatrix} = \lambda^2 - 2\lambda + 2$$
$$= (\lambda - (1 + i))(\lambda - (1 - i)).$$

Los autovalores son $\lambda_1 = 1 + i, \lambda_2 = 1 - i$. Calculemos los espacios de autovectores asociados.
Para $\lambda_1 = 1 + i$,

$$A - \lambda_1 I = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \to \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix} \to \begin{cases} x_1 = i x_2 \\ x_2 \text{ libre} \end{cases} \to \text{null}(A - \lambda_1 I) = \langle \boldsymbol{v}_1 \rangle,$$

donde

$$\boldsymbol{v}_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Para $\lambda_2 = 1 - i$,

$$A - \lambda_2 I = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \to \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix} \to \begin{cases} x_1 = -ix_2 \\ x_2 \text{ libre} \end{cases} \to \text{null}(A - \lambda_2 I) = \langle \boldsymbol{v}_2 \rangle,$$

donde

$$\boldsymbol{v}_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Como vemos, el cálculo de autovalores conduce a la resolución de una ecuación polinómica, lo que puede ser una tarea de difícil solución. No existe una fórmula exacta para determinar, en general, las raíces de un polinomio de grado n, por lo que se usan métodos iterados para obtener valores aproximados.

Nota 7.1.4.

- Si una matriz A es triangular (superior o inferior), sus autovalores son sus entradas diagonales.
- Los autovalores y autovectores se modifican al efectuar transformaciones elementales en una matriz (de fila o columna).
- Sea A una matriz con autovalores $\lambda_1, \ldots, \lambda_n$.
  - Si $k > 0$ es un número natural, entonces la matriz A^k tiene como autovalores $\lambda_1^k, \ldots, \lambda_n^k$, con los mismos autovectores asociados.
  - Si A es no singular, entonces A^{-1} tiene como autovalores $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$, con los mismos autovectores asociados.

Multiplicidad algebraica y geométrica

Si λ es autovalor de A, llamamos

- ***multiplicidad algebraica*** de λ a la multiplicidad de λ como raíz del polinomio característico de A.
- ***multiplicidad geométrica*** de λ a la dimensión del espacio $V_1(\lambda) = \text{null}(A - \lambda I)$.

En general, si λ_i es autovalor de una matriz $A_{n\times n}$, escribiremos m_i para su multiplicidad algebraica y q_i para su multiplicidad geométrica. Es inmediato que $n = \sum_{i=1}^{r} m_i$, donde r es el número de autovalores distintos.

A lo largo de este tema, identificaremos una matriz A con la aplicación lineal inducida f sobre el espacio vectorial $V = \mathbb{R}^n$ o $\mathbb{C}^n$.

Desigualdad entre la multiplicidad algebraica y geométrica

Sea λ_0 autovalor de A, y llamemos q_0 a su multiplicidad geométrica, y m_0 a su multiplicidad algebraica.

1. $q_0 \leq m_0$.
2. Si $\text{null}(\lambda_0 I - A) = \text{null}(\lambda_0 I - A)^2$ entonces $m_0 = q_0$.

PRUEBA:

1. Consideremos una base $\mathscr{B}_0$ de $V_1(\lambda_0)$, que tiene q_0 vectores, y la prolongamos a una base $\mathscr{B}$ de V. Entonces la matriz de la aplicación lineal respecto a la nueva base $\mathscr{B}$ es de la forma

$$A' = \begin{pmatrix} D_0 & M \\ 0 & Q \end{pmatrix}, \tag{7.1.1}$$

donde D_0 es una matriz diagonal de orden q_0 con entradas iguales a λ_0. El polinomio característico de f es igual entonces a $(\lambda-\lambda_0)^{q_0}\det(\lambda I - Q)$, por lo que la multiplicidad algebraica de λ_0 es mayor o igual que q_0.

2. Supongamos que el polinomio característico de la matriz Q en la expresión (7.1.1) tiene el autovalor λ_0, y $q_0 < m_0$. Sea $\boldsymbol{a}'$ un autovector de Q asociado a λ_0. Entonces el vector

$$\boldsymbol{a} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{a}' \end{pmatrix}, \text{ coordenadas respecto a la base } \mathscr{B},$$

es independiente de los q_0 primeros vectores de la base $\mathscr{B}$, que generan a $\text{null}(\lambda_0 I - A)$ (sus primeras q_0 componentes son nulas). Escribamos

$$\boldsymbol{a} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{q_0+1} \\ \vdots \\ a_n \end{pmatrix}$$

en coordenadas respecto de la base $\mathscr{B}$. Entonces

$$(\lambda_0 I - Q)\begin{pmatrix} a_{q_0+1} \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Observemos que $\text{null}(\lambda_0 I - A') = \text{null}(\lambda_0 I - A')^2$. Por un lado, si $\boldsymbol{z} \in \text{null}(\lambda_0 I - A')$, entonces $(\lambda_0 I - A')\boldsymbol{z} = \mathbf{0}$, y $(\lambda_0 I - A')(\lambda_0 I - A')\boldsymbol{z} = \mathbf{0}$, de donde $\boldsymbol{z} \in \text{null}(\lambda_0 I - A')^2$. Por otro, existe P no singular tal que $A' = PAP^{-1}$, y si $(\lambda_0 I - A')^2 \boldsymbol{z} = \mathbf{0}$, entonces $P(\lambda_0 I - A)^2 P^{-1}\boldsymbol{z} = \mathbf{0}$, es decir, $P^{-1}\boldsymbol{z} \in \text{null}(\lambda_0 I - A)^2 = \text{null}(\lambda_0 I - A)$. Entonces $(\lambda_0 I - A)(P^{-1}\boldsymbol{z}) = \mathbf{0}$, y multiplicando a la izquierda por P llegamos a $(\lambda_0 I - A')\boldsymbol{z} = \mathbf{0}$.

Tenemos que

$$(\lambda_0 I - A')\boldsymbol{a} = \begin{pmatrix} \lambda_0 I - D_0 & -M \\ 0 & \lambda_0 I - Q \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{q_0+1} \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_{q_0} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Como únicamente aparecen componentes en las q_0 primeras coordenadas, el vector $(\lambda_0 I - A')\boldsymbol{a}$ pertenece a $\text{null}(\lambda_0 I - A')$, esto es, $\boldsymbol{a} \in \text{null}(\lambda_0 I - A')^2 = \text{null}(\lambda_0 I - A')$, lo que es contradictorio con la elección de $\boldsymbol{a}$. Entonces todos los factores $(\lambda - \lambda_0)$ del polinomio característico de A' están en $\det(\lambda I - M)$, y $m_0 = q_0$.

□

¿Qué ocurre con los autovectores asociados a autovalores distintos? La respuesta es que forman un conjunto linealmente independiente, esto es, si $\mu_1, \ldots, \mu_s$ son autovalores de una matriz A, distintos dos a dos, y $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s$ son autovectores asociados respectivos, entonces $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_s\}$ es un conjunto linealmente independiente.

La prueba es por inducción sobre s. Para $s = 1$ es trivial. Supongamos que $s > 1$ y el resultado es válido para conjuntos de $s-1$ autovectores asociados a autovalores distintos. Consideremos una combinación lineal $\sum_{i=1}^{s} \alpha_i \boldsymbol{v}_i = \mathbf{0}$. Si aplicamos A, nos queda $\sum_{i=1}^{s} \alpha_i \mu_i \boldsymbol{v}_i = \mathbf{0}$. Si multiplicamos la primera suma por μ_s y le restamos la segunda obtenemos $\sum_{i=1}^{s-1} \alpha_i (\mu_s - \mu_i)\boldsymbol{v}_i = \mathbf{0}$. Por hipótesis de inducción, los vectores $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{s-1}$ son linealmente independientes, por lo que $\alpha_i(\mu_s - \mu_i) = 0, i = 1, \ldots, s-1$. Como los autovalores son distintos, nos queda $\alpha_1 = \cdots = \alpha_{s-1} = 0$. Volvemos a la primera ecuación, y obtenemos $\alpha_s = 0$.

7.2. Matrices diagonalizables

Un problema fundamental en Álgebra Lineal es el siguiente: dado un una aplicación lineal sobre un espacio vectorial de dimensión finita, calcular una

base del espacio respecto de la cual la matriz de la aplicación sea lo más sencilla posible. Sabemos que las diferentes representaciones de una aplicación respecto a las bases del espacio están relacionadas por la semejanza de matrices. La cuestión, desde el punto de vista matricial, es dada una matriz A, encontrar una matriz no singular P tal que $P^{-1}AP$ sea lo más sencilla posible.

Hacemos bien en pensar que la forma más sencilla es una matriz diagonal. Las que se pueden transformar reciben un nombre.

Matriz diagonalizable

Una matriz cuadrada $A_{n\times n}$ se dice ***diagonalizable*** si A es semejante a una matriz diagonal.

Ejemplo 7.2.1.

- La matriz

$$A=\begin{pmatrix} 7 & -4 \\ 5 & -2 \end{pmatrix}$$

es diagonalizable, por que para

$$P=\begin{pmatrix} 4/5 & 1 \\ 1 & 1 \end{pmatrix} \text{ se verifica } P^{-1}AP=\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}.$$

- La matriz

$$A=\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

no se puede transformar en una matriz diagonal. Observemos en primer lugar que $A^2=0$. Si existiera P no singular tal que $P^{-1}AP=D$, con D diagonal, entonces

$$D^2=P^{-1}APP^{-1}AP=P^{-1}A^2P=0,$$

de donde $D=0$, y llegaríamos a que $A=0$.

Por tanto, si no todas las matrices se pueden transformar en una diagonal mediante transformaciones de semejanza, ¿qué caracteriza a las que sí se puede? Una respuesta se puede derivar fácilmente mediante el examen de la ecuación

$$P^{-1}A_{n\times n}P=\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

que implica

$$A\begin{pmatrix} P_{*1} & \dots & P_{*n} \end{pmatrix} = \begin{pmatrix} P_{*1} & \dots & P_{*n} \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix},$$

o de manera equivalente,

$$\begin{pmatrix} AP_{*1} & \dots & AP_{*n} \end{pmatrix} = \begin{pmatrix} \lambda_1 P_{*1} & \dots & \lambda_n P_{*n} \end{pmatrix}.$$

En consecuencia, $AP_{*j} = \lambda_j P_{*j}$ para cada $j = 1, \dots, n$, lo que significa que (λ_j, P_{*j}) es un par autovalor-autovector de A. En otras palabras, $P^{-1}AP = D$ implica que P debe ser una matriz cuyas columnas constituyen un conjunto de autovectores linealmente independientes, y D es una matriz diagonal cuyas entradas son los autovalores correspondientes. El recíproco es inmediato, es decir, si existe un conjunto linealmente independiente de n autovectores que usamos para construir una matriz no singular P, y D es la matriz diagonal cuyas entradas son los autovalores correspondientes, entonces $P^{-1}AP = D$.

Tenemos entonces que $A_{n\times n}$ es una matriz diagonalizable si y solamente si existen n autovectores independientes, es decir $\sum_{i=1}^{r} q_i = n$, donde q_i es la multiplicidad geométrica del autovalor λ_i. Como

$$\sum_{i=1}^{r} q_i = n = \sum_{i=1}^{r} m_i, \text{ y } q_i \leq m_i,$$

el carácter diagonalizable de A es equivalente a que para cada autovalor λ_i, $i = 1, \dots, r$, la multiplicidad algebraica m_i coincide con la multiplicidad geométrica q_i.

Lo escribimos a modo de sumario.

Diagonalización

- $A_{n\times n}$ es diagonalizable si y solamente si A tiene un conjunto de n autovectores linealmente independientes. Además, $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$ si y solamente si las columnas de P son una base del espacio formada por autovectores, y los λ_j son los autovalores asociados.
- Para cada autovalor, la multiplicidad algebraica coincide con la multiplicidad geométrica.

Ejemplo 7.2.2. Consideremos la matriz

$$A=\begin{pmatrix} 1 & -4 & -4 \\ 8 & -11 & -8 \\ -8 & 8 & 5 \end{pmatrix},$$

de polinomio característico $(\lambda-1)(\lambda+3)^2$. Entonces $\lambda_1=1, m_1=1, \lambda_2=-3, m_2=2$. Vamos a calcular los espacios $V_1(\lambda_1), V_1(\lambda_2)$.

$$\lambda_1 I-A=\begin{pmatrix} 0 & 4 & 4 \\ -8 & 12 & 8 \\ 8 & -8 & -4 \end{pmatrix}\to\begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}\to\begin{cases} x_1=-1/2x_3 \\ x_2=-x_3 \\ x_3 \text{ libre} \end{cases}$$

$$\to \text{null}(\lambda_1 I-A)=\langle \boldsymbol{v}_{11}\rangle, \text{ donde } \boldsymbol{v}_{11}=\begin{pmatrix} -1/2 \\ -1 \\ 1 \end{pmatrix}.$$

Para $V_1(\lambda_2)$ tenemos

$$\lambda_2 I-A=\begin{pmatrix} -4 & 4 & 4 \\ -8 & 8 & 8 \\ 8 & -8 & -8 \end{pmatrix}\to\begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\to\begin{cases} x_1=x_2+x_3 \\ x_2, x_3 \text{ libres} \end{cases}$$

$$\to \text{null}(\lambda_2 I-A=\langle \boldsymbol{v}_{21}, \boldsymbol{v}_{22}\rangle, \text{ donde } \boldsymbol{v}_{21}=\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{v}_{22}=\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Entonces $\dim V_1(\lambda_1)=m_1, \dim V_1(\lambda_2)=m_2$, por lo que A es diagonalizable, y

$$\begin{pmatrix} 1 & & \\ & -3 & \\ & & -3 \end{pmatrix}=P^{-1}AP, \text{ donde } P=\begin{pmatrix} \boldsymbol{v}_{11} & \boldsymbol{v}_{21} & \boldsymbol{v}_{22} \end{pmatrix}=\begin{pmatrix} -1/2 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

7.3. Lema de Schur

Lema de Schur

Sea A una matriz cuadrada con coeficientes complejos. Entonces existe una matriz unitaria U tal que $U^{-1}AU=U^*AU$ es triangular superior.

PRUEBA: La prueba es por inducción sobre la dimensión de V. Para $n=1$ es trivial. Sea $\boldsymbol{w}_1$ un autovector asociado a un autovalor λ_1 de A (sobre $\mathbb{C}$ tenemos garantía de su existencia), y lo normalizamos a $\boldsymbol{v}_1$. Ampliamos a una base

Figura 7.1: I. Schur (1875-1941)

de $\mathbb{C}^n$, y mediante Gram-Schmidt o la factorización QR obtenemos una base ortonormal que tiene a $\boldsymbol{v}_1$ como primer vector (otra forma de obtener esta base ortonormal es mediante una matriz de Householder, tal como se hizo en el ejemplo 6.7.1). Sea U_1 la matriz del cambio de base, que es unitaria. Entonces

$$U_1^{-1}AU_1 = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}.$$

Por hipótesis de inducción, existe V_2 unitaria de dimensión $n-1$ tal que $V_2^{-1}A_1V_2$ es triangular superior. Sea

$$U_2 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & V_2 & \\ 0 & & & \end{pmatrix}.$$

Entonces U_2 es unitaria, y $U_2^{-1}(U_1^{-1}AU_1)U_2$ es triangular superior. Para $U = U_1U_2$ tenemos el resultado. □

Nota 7.3.1. Si los autovalores de la matriz A están en $\mathbb{R}$, entonces se sigue de la prueba que podemos construir U *ortogonal* tal que U^tAU es triangular superior. Basta observar en la prueba que las matrices unitarias empleadas son ortogonales.

Ejemplo 7.3.2. Consideremos la matriz

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

con autovalores $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 3$. Calculamos el espacio de autovectores para λ_1:

$$\lambda_1 I - A = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ y un autovector es } \boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Observemos que $\boldsymbol{v}_1$ ya es unitario. Es inmediato ampliar a una base ortonormal, con lo que obtenemos

$$U_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Entonces

$$U_1^{-1} A U_1 = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

y llamamos

$$A_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

que sabemos que tiene como autovalores λ_2, λ_3. Calculamos el espacio de autovectores de λ_2 en la matriz A_2:

$$\lambda_2 I - A_2 = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \text{ y un autovector es } \boldsymbol{w}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

El vector $\boldsymbol{w}_2$ no está normalizado, por lo que construimos $\boldsymbol{v}_2 = \frac{1}{\|\boldsymbol{w}_2\|}\boldsymbol{w}_2 = \frac{1}{\sqrt{2}}\boldsymbol{w}_2$. Ahora debemos encontrar una base ortonormal de $\mathbb{R}^2$ que contenga a $\boldsymbol{v}_2$. Vamos a hacerlo de dos formas:

1. Mediante Gram-Schmidt. En primer lugar, calculamos el complemento ortogonal de forma análoga al ejemplo 6.5.2.

$$\langle \boldsymbol{v}_2 \rangle^\perp \equiv \left\{ -\tfrac{1}{\sqrt{2}}x_1 + \tfrac{1}{\sqrt{2}}x_2 = 0. \right. \Rightarrow \left\{ \begin{array}{rcl} x_1 & = & x_2, \\ x_2 & = & x_2. \end{array} \right.$$

Entonces

$$\langle \boldsymbol{v}_1 \rangle^{\perp} = \langle \boldsymbol{q}_2' = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle,$$

y basta normalizar

$$\boldsymbol{q}_2 = \frac{1}{\|\boldsymbol{q}_2'\|} \boldsymbol{q}_2' = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

para que $\{\boldsymbol{v}_1, \boldsymbol{q}_2\}$ sea base ortonormal de $\mathbb{R}^2$.

2. Mediante Householder. En el caso de partir de un único vector, mediante el cálculo de una matriz de Householder podemos encontrar la base. Sea $\boldsymbol{w} = \boldsymbol{v}_2 - \boldsymbol{e}_1$. Entonces

$$H(\boldsymbol{w}) = I_2 - \frac{2}{\boldsymbol{w}^t \boldsymbol{w}} \boldsymbol{w}\boldsymbol{w}^t = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}.$$

La primera columna es $\boldsymbol{v}_2$, y la segunda el vector que amplía a una base ortonormal de $\mathbb{R}^2$.

En cualquier caso, hemos construido la matriz ortogonal

$$V_2 = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Entonces si

$$U_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

nos queda que

$$U_2^{-1} U_1^{-1} A U_1 U_2 = \begin{pmatrix} 2 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

La matriz unitaria buscada es

$$U_1 U_2 = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}.$$

El teorema de triangulación de Schur asegura que toda matriz cuadrada A es semejante mediante una transformación unitaria a una matriz triangular superior, esto es, $U^* A U = T$. Pero incluso siendo A real, las matrices U y T serán complejas si A tiene autovalores complejos conjugados. Sin embargo, las matrices se pueden encontrar reales si permitimos bloques 2×2 en la diagonal. Se

puede probar que si $A \in \mathbb{R}^{n\times n}$, existe una matriz ortogonal $P \in \mathbb{R}^{n\times n}$ y matrices reales B_{ij} tales que

$$P^t AP = \begin{pmatrix} B_{11} & B_{12} & \dots & B_{1k} \\ 0 & B_{22} & \dots & B_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_{kk} \end{pmatrix}, \text{ donde } B_{jj} \text{ es } 1\times 1 \text{ o } 2\times 2.$$

Si $B_{jj} = [\lambda_j]$, entonces λ_j es autovalor (real) de A, y si B_{jj} es 2×2, entonces los autovalores de B_{jj} son complejos conjugados del espectro de A.

7.4. Teoremas espectrales

Matrices normales

Una matriz $A \in \mathbb{C}^{n\times n}$ es semejante a través de una matriz unitaria a una matriz diagonal si y solamente si $A^*A = AA^*$, es decir, si A es una matriz normal.

PRUEBA: Supongamos, en primer lugar, que A es una matriz normal. Por el lema de Schur, existe una matriz unitaria U tal que $U^*AU = T$, con T triangular superior. Entonces $T^* = U^*A^*U$, y

$$TT^* = U^*AUU^*A^*U = U^*AA^*U = U^*A^*AU = T^*T.$$

Como T es triangular superior, el elemento $(1,1)$ de TT^* es de la forma $|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2$, pero el elemento $(1,1)$ de T^*T es $|t_{11}|^2$. Por tanto, todos los elementos t_{1j}, $j \geq 2$ son nulos, y podemos escribir

$$T = \begin{pmatrix} t_{11} & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & T_1 \end{pmatrix},$$

con T_1 triangular superior. Entonces

$$TT^* = \begin{pmatrix} |t_{11}|^2 & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & T_1T_1^* \end{pmatrix} = T^*T = \begin{pmatrix} |t_{11}|^2 & \mathbf{0}_{1\times(n-1)} \\ \mathbf{0}_{(n-1)\times 1} & T_1^*T_1 \end{pmatrix},$$

y, por inducción, llegamos a la conclusión de que T es diagonal.

Recíprocamente, si existe U unitaria tal que $U^*AU = D$, con D matriz diagonal, entonces

$$A^*A = UD^*U^*UDU^* = UD^*DU^* = UDD^*U^* = AA^*,$$

y A es una matriz normal. □

Un corolario de lo anterior es que los espacios de autovectores de autovalores distintos de una matriz normal son *ortogonales* entre sí. En efecto, sean λ_1 y λ_2 autovalores de una matriz normal A, con $\lambda_1 \neq \lambda_2$. Sea U una matriz unitaria tal que $U^* AU = D$, con D diagonal. Entonces las columnas de U son autovectores y forman una base ortonormal del espacio, por lo que los espacios de autovectores asociados a λ_1 y λ_2 son de la forma $V_1(\lambda_1) = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{q_1} \rangle$, $V_1(\lambda_2) = \langle \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{q_2} \rangle$, con $\boldsymbol{u}_i, \boldsymbol{v}_j$ columnas distintas de la matriz U, que son ortogonales entre sí. Entonces cada uno de los generadores de $V_1(\lambda_1)$ es ortogonal a cada uno de los generadores de $V_1(\lambda_2)$, y tenemos lo que queríamos.

Ortogonalidad de los espacios de autovectores

Sea A una matriz normal. Entonces los espacios de autovectores asociados a autovalores distintos son mutuamente ortogonales.

Muchos tipos de matrices son normales. Entre ellas tenemos a las simétricas reales y las hermitianas, las anti-simétricas reales y las anti-hermitianas, las ortogonales y las unitarias. Todas ellas comparten las propiedades anteriores, pero vamos a fijarnos un poco más en las simétricas reales y las hermitianas, porque sus autovalores tienen algunas propiedades especiales.

Sea A simétrica real o hermitiana, y $(\lambda, \boldsymbol{v})$ un par autovalor-autovector de A. Entonces $\boldsymbol{v}^* \boldsymbol{v} \neq 0$, y $A\boldsymbol{v} = \lambda \boldsymbol{v}$ implica $\boldsymbol{v}^* A^* = \overline{\lambda} \boldsymbol{v}^*$. Entonces

$$\boldsymbol{v}^* A\boldsymbol{v} = \lambda \boldsymbol{v}^* \boldsymbol{v}, \boldsymbol{v}^* A^* \boldsymbol{v} = \overline{\lambda} \boldsymbol{v}^* \boldsymbol{v},$$

y como $A^* = A$, podemos restar y queda $0 = (\lambda - \overline{\lambda}) \boldsymbol{v}^* \boldsymbol{v}$. Dado que $\boldsymbol{v}^* \boldsymbol{v} \neq 0$, se sigue que $\lambda = \overline{\lambda}$. Por tanto los autovalores de una matriz simétrica real o una hermitiana son ***reales***.

Matrices hermitianas y simétricas

- Las matrices simétricas reales y las hermitianas tienen autovalores reales.
- Una matriz A es real simétrica si y solamente si A es semejante de forma *ortogonal* a una matriz diagonal real D, es decir, $D = P^t AP$ con P ortogonal.

Ejemplo 7.4.1. Sea A la matriz

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix}.$$

Como A es simétrica, el teorema espectral de matrices simétricas nos dice que A es diagonalizable en $\mathbb{R}$. Los autovalores son $\lambda_1 = 2, \lambda_2 = 8$, con multiplicidades algebraicas respectivas $m_1 = 2, m_2 = 1$.

- Para λ_1, el espacio de autovectores es

$$V_1(\lambda_1) = \text{null}(A - \lambda_1 I) \Rightarrow \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Para resolver este sistema lineal homogéneo, calculamos la forma escalonada reducida por filas de la matriz de coeficientes:

$$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{cases} x_1 &= -x_2 & -x_3, \\ x_2 &= x_2, & \\ x_3 &= & x_3. \end{cases}$$

Entonces

$$V_1(\lambda_1) = \langle \boldsymbol{v}_{11} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{v}_{12} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \rangle.$$

Ahora aplicamos Gram-Schmidt a este conjunto para transformarlo en

un conjunto ortonormal.

$$\begin{aligned}
&\boldsymbol{q}'_{11} = \boldsymbol{v}_{11},\\
&\boldsymbol{q}'_{12} = \boldsymbol{v}_{12} - \lambda_{12}\boldsymbol{q}'_{11}, \lambda_{12} = \frac{\boldsymbol{v}_{12}\bullet\boldsymbol{q}'_{11}}{\boldsymbol{q}'_{11}\bullet\boldsymbol{q}'_{11}} = \frac{1}{2},\\
&\boldsymbol{q}'_{12} = \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix},\\
&\boldsymbol{q}_{11} = \frac{1}{\|\boldsymbol{q}'_{11}\|}\boldsymbol{q}'_{11} = \begin{bmatrix} -1/2\sqrt{2} \\ 1/2\sqrt{2} \\ 0 \end{bmatrix},\\
&\boldsymbol{q}_{12} = \frac{1}{\|\boldsymbol{q}'_{12}\|}\boldsymbol{q}'_{12} = \begin{bmatrix} -1/6\sqrt{6} \\ -1/6\sqrt{6} \\ 1/3\sqrt{6} \end{bmatrix}.
\end{aligned}$$

- Para λ_2, procedemos análogamente.

$$V_1(\lambda_2) = \text{null}(A - \lambda_2 I) \Rightarrow \begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Para resolver este sistema lineal homogéneo, calculamos la forma escalonada reducida por filas de la matriz de coeficientes:

$$\begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{cases} x_1 = x_3, \\ x_2 = x_3, \\ x_3 = x_3. \end{cases}$$

Entonces

$$V_1(\lambda_2) = \langle \boldsymbol{v}_{21} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \rangle.$$

Observemos, como era de esperar, que $V_1(\lambda_2)$ es un subespacio vectorial ortogonal a $V_1(\lambda_1)$, es decir, los vectores $\boldsymbol{v}_{1i}$ son ortogonales a los vectores $\boldsymbol{v}_{2j}$. De nuevo, aplicamos Gram-Schmidt a $V_1(\lambda_2)$, pero aquí solamente

tenemos que normalizar el vector $\boldsymbol{v}_{21}$:

$$\boldsymbol{q}_{21} = \frac{1}{\|\boldsymbol{v}_{21}\|}\boldsymbol{v}_{21} = \begin{bmatrix} 1/3\sqrt{3} \\ 1/3\sqrt{3} \\ 1/3\sqrt{3} \end{bmatrix}.$$

Por tanto, la matriz $P = \begin{pmatrix} \boldsymbol{q}_{11} & \boldsymbol{q}_{12} & \boldsymbol{q}_{21} \end{pmatrix}$ es una matriz ortogonal formada por autovectores, por lo que

$$P^t AP = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

Ejemplo 7.4.2. Consideremos la matriz simétrica

$$A = \begin{bmatrix} \frac{309}{187} & -\frac{67}{187} & \frac{12}{17} & -\frac{63}{187} \\ -\frac{67}{187} & \frac{210}{187} & \frac{21}{17} & \frac{30}{187} \\ \frac{12}{17} & \frac{21}{17} & 1/17 & \frac{3}{17} \\ -\frac{63}{187} & \frac{30}{187} & \frac{3}{17} & \frac{218}{187} \end{bmatrix}.$$

Sus autovectores son $\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = -1$, de multiplicidades algebraicas respectivas $m_1 = 2, m_2 = 1, m_3 = -1$. Los espacios de autovectores son:

- Para λ_1,

$$\lambda_1 I - A = \begin{bmatrix} \frac{65}{187} & \frac{67}{187} & -\frac{12}{17} & \frac{63}{187} \\ \frac{67}{187} & \frac{164}{187} & -\frac{21}{17} & -\frac{30}{187} \\ -\frac{12}{17} & -\frac{21}{17} & \frac{33}{17} & -\frac{3}{17} \\ \frac{63}{187} & -\frac{30}{187} & -\frac{3}{17} & \frac{156}{187} \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

  Entonces

$$V_1(\lambda_1) = \langle \boldsymbol{v}_{11} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{v}_{12} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \rangle.$$

- Para λ_2,

$$\lambda_2 I - A = \begin{bmatrix} -\frac{122}{187} & \frac{67}{187} & -\frac{12}{17} & \frac{63}{187} \\ \frac{67}{187} & -\frac{23}{187} & -\frac{21}{17} & -\frac{30}{187} \\ -\frac{12}{17} & -\frac{21}{17} & \frac{16}{17} & -\frac{3}{17} \\ \frac{63}{187} & -\frac{30}{187} & -\frac{3}{17} & -\frac{31}{187} \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & -1/3 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Entonces

$$V_1(\lambda_2) = \langle \boldsymbol{v}_{21} = \begin{bmatrix} 1/3 \\ -1/3 \\ 0 \\ 1 \end{bmatrix} \rangle.$$

- Para λ_3,

$$\lambda_3 I - A = \begin{bmatrix} -\frac{496}{187} & \frac{67}{187} & -\frac{12}{17} & \frac{63}{187} \\ \frac{67}{187} & -\frac{397}{187} & -\frac{21}{17} & -\frac{30}{187} \\ -\frac{12}{17} & -\frac{21}{17} & -\frac{18}{17} & -\frac{3}{17} \\ \frac{63}{187} & -\frac{30}{187} & -\frac{3}{17} & -\frac{405}{187} \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 11 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Entonces

$$V_1(\lambda_3) = \langle \boldsymbol{v}_{31} = \begin{bmatrix} 4 \\ 7 \\ -11 \\ 1 \end{bmatrix} \rangle.$$

Sabemos que los espacios de autovectores de autovalores diferentes son mutuamente ortogonales, pero los generadores de $V_1(\lambda_1)$ calculados no lo son en este caso. ¿Cómo podemos conseguirlo? Aplicamos Gram-Schmidt (QR) a $V_1(\lambda_1)$. Recordemos que con este procedimiento no nos salimos de la variedad lineal. Por tanto, los vectores que calculemos seguirán siendo autovectores. En este caso resulta

$$\begin{aligned} \boldsymbol{q}_1' &= \boldsymbol{v}_{11}, \\ \boldsymbol{q}_2' &= \boldsymbol{v}_{12} - \lambda_{12}\boldsymbol{q}_1', \\ \lambda_{12} &= \frac{\boldsymbol{v}_{12} \bullet \boldsymbol{q}_1'}{\boldsymbol{q}_1' \bullet \boldsymbol{q}_1'} = -\frac{1}{6}, \\ \boldsymbol{q}_2' &= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 7/6 \\ 1 \\ 1/6 \end{bmatrix}. \end{aligned}$$

Tanto $\boldsymbol{q}_1'$ como $\boldsymbol{q}_2'$ son autovectores asociados al autovalor λ_1. Para formar la

matriz de paso ortogonal, solamente tenemos que normalizar:

$$\boldsymbol{q}_1 = \tfrac{1}{\|\boldsymbol{q}_1'\|}\boldsymbol{q}_1' = \begin{bmatrix} -1/3\sqrt{6} \\ 1/6\sqrt{6} \\ 0 \\ 1/6\sqrt{6} \end{bmatrix}, \boldsymbol{q}_2 = \tfrac{1}{\|\boldsymbol{q}_2'\|}\boldsymbol{q}_2' = \begin{bmatrix} \frac{2}{51}\sqrt{102} \\ \frac{7}{102}\sqrt{102} \\ 1/17\sqrt{102} \\ \frac{1}{102}\sqrt{102} \end{bmatrix},$$

$$\boldsymbol{q}_3 = \tfrac{1}{\|\boldsymbol{v}_{21}\|}\boldsymbol{v}_{21} = \begin{bmatrix} 1/11\sqrt{11} \\ -1/11\sqrt{11} \\ 0 \\ 3/11\sqrt{11} \end{bmatrix}, \boldsymbol{q}_4 = \tfrac{1}{\|\boldsymbol{v}_{31}\|}\boldsymbol{v}_{31} = \begin{bmatrix} \frac{4}{187}\sqrt{187} \\ \frac{7}{187}\sqrt{187} \\ -1/17\sqrt{187} \\ \frac{1}{187}\sqrt{187} \end{bmatrix}.$$

Por tanto, la matriz ortogonal

$$Q = \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \boldsymbol{q}_4 \end{pmatrix}$$

verifica que $Q^t AQ = \operatorname{diag}(\lambda_1, \lambda_1, \lambda_2, \lambda_3)$.

Las matrices ortogonales y unitarias tienen autovalores complejos de módulo igual a 1, pues si $A\boldsymbol{v} = \lambda\boldsymbol{v}$, entonces $A\boldsymbol{v} \bullet A\boldsymbol{v} = \overline{\lambda}\lambda\boldsymbol{v} \bullet \boldsymbol{v} = \boldsymbol{v} \bullet \boldsymbol{v}$, de donde $\|\lambda\| = 1$.

Ejemplo 7.4.3. Consideremos la matriz unitaria

$$U = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/2 & 1/2 \\ -1/\sqrt{2} & 0 & -1/2 & 1/2 \\ 0 & 1/\sqrt{2} & 1/2 & 1/2 \\ 0 & -1/\sqrt{2} & 1/2 & 1/2 \end{pmatrix}$$

Tiene como autovalores a $\lambda_1 = 1$, con $m_1 = 2$, y dos autovalores complejos conjugados λ_2, λ_3. El espacio de autovectores asociado a λ_1 es $V_1(\lambda_1) = \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle$, donde

$$\boldsymbol{v}_1 = \begin{pmatrix} 0{,}862856 \\ -0{,}357407 \\ -0{,}252725 \\ 0{,}252725 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 0{,}143356 \\ -0{,}059380 \\ 0{,}655291 \\ 0{,}739267 \end{pmatrix}$$

Podemos calcular una base ortonormal de $V_1(\lambda_1)$, que nos proporciona

$$\boldsymbol{w}_1 = \begin{pmatrix} -0{,}862856 \\ 0{,}357407 \\ 0{,}252725 \\ -0{,}252725 \end{pmatrix}, \boldsymbol{w}_2 = \begin{pmatrix} 0 \\ 0 \\ -0{,}707107 \\ -0{,}707107 \end{pmatrix}.$$

Sea ahora A una matriz ortogonal de orden n, y vamos a probar que existe una matriz P ortogonal tal que P^tAP es una matriz diagonal por cajas de la forma

$$P^tAP = \begin{pmatrix} U_1 & & & \\ & U_2 & & \\ & & \ddots & \\ & & & U_r \end{pmatrix}, \text{ con } U_i = \begin{pmatrix} a_i & b_i \\ -b_i & a_i \end{pmatrix}, a_i^2 + b_i^2 = 1, 2r = n.$$

La idea consiste en agrupar los autovalores conjugados y los autovectores correspondientes. En primer lugar, si $(\lambda, \boldsymbol{v})$ es un par autovalor/autovector, con λ real, entonces $\lambda = \pm 1$. Basta tomar entonces $a = \pm 1, b = 0$. Consideremos entonces un autovalor $\lambda = a + bi, b \neq 0$, con $\boldsymbol{v}$ autovector unitario asociado. Entonces $A\boldsymbol{v} = \lambda\boldsymbol{v}$, y si conjugamos esta igualdad obtenemos $A\overline{\boldsymbol{v}} = \overline{\lambda}\overline{\boldsymbol{v}}$, donde $\overline{\boldsymbol{v}}$ es el vector que se obtiene al conjugar todas las componentes de $\boldsymbol{v}$. Expresemos $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2$, donde $\boldsymbol{v}_1, \boldsymbol{v}_2$ son ahora vectores de componentes reales. Entonces $\overline{\boldsymbol{v}} = \boldsymbol{v}_1 - i\boldsymbol{v}_2$, y

$$A\boldsymbol{v} = A\boldsymbol{v}_1 + iA\boldsymbol{v}_2 = (a + ib)(\boldsymbol{v}_1 + i\boldsymbol{v}_2) = (a\boldsymbol{v}_1 - b\boldsymbol{v}_2) + i(b\boldsymbol{v}_1 + a\boldsymbol{v}_2),$$

de donde

$$A\boldsymbol{v}_1 = a\boldsymbol{v}_1 - b\boldsymbol{v}_2, A\boldsymbol{v}_2 = b\boldsymbol{v}_1 + a\boldsymbol{v}_2. \tag{7.4.1}$$

Por otro lado, sabemos que $\boldsymbol{v}$ y $\overline{\boldsymbol{v}}$ son ortogonales con el producto escalar complejo, es decir $\overline{\boldsymbol{v}}^*\boldsymbol{v} = 0$. Entonces

$$0 = (\boldsymbol{v}_1 - i\boldsymbol{v}_2)^*(\boldsymbol{v}_1 + i\boldsymbol{v}_2) = (\boldsymbol{v}_1^t + i\boldsymbol{v}_2^t)(\boldsymbol{v}_1 + i\boldsymbol{v}_2) = (\boldsymbol{v}_1^t\boldsymbol{v}_1 - \boldsymbol{v}_2^t\boldsymbol{v}_2) + 2i\boldsymbol{v}_2^t\boldsymbol{v}_1.$$

Esto significa que $\boldsymbol{v}_2^t\boldsymbol{v}_1 = 0$, es decir, son ortogonales con el producto escalar real, y que $\|\boldsymbol{v}_1\| = \|\boldsymbol{v}_2\|$. Si normalizamos tanto $\boldsymbol{v}_1$ como $\boldsymbol{v}_2$, y llamamos $\boldsymbol{w}_i = \frac{1}{\|\boldsymbol{v}_i\|}\boldsymbol{v}_i, i = 1, 2$, la ecuación 7.4.1 queda como

$$A\boldsymbol{w}_1 = a\boldsymbol{w}_1 - b\boldsymbol{w}_2, A\boldsymbol{w}_2 = b\boldsymbol{w}_1 + a\boldsymbol{w}_2. \tag{7.4.2}$$

La idea es sustituir cada par de autovectores conjugados $(\boldsymbol{v}, \overline{\boldsymbol{v}})$ por $(\boldsymbol{w}_1, \boldsymbol{w}_2)$. Falta comprobar que estos vectores forman una base ortonormal del espacio. Ya hemos visto que son ortogonales entre sí. Por ello, consideremos dos autovectores $\boldsymbol{x}\boldsymbol{x}_1 + i\boldsymbol{x}_2, \boldsymbol{y} = \boldsymbol{y}_1 + i\boldsymbol{y}_2$ de A asociados a autovalores distintos. Sabemos que son ortogonales, por lo que

$$0 = \boldsymbol{y}^*\boldsymbol{x} = (\boldsymbol{y}_1^t\boldsymbol{x}_1 + \boldsymbol{y}_2^t\boldsymbol{x}_2) + i(\boldsymbol{y}_1^t\boldsymbol{x}_2 - \boldsymbol{y}_2^t\boldsymbol{x}_1).$$

Por otro lado, $\overline{\boldsymbol{x}}$ y $\boldsymbol{y}$ son también ortogonales, de donde

$$0 = \boldsymbol{y}^*\overline{\boldsymbol{x}} = (\boldsymbol{y}_1^t\boldsymbol{x}_1 - \boldsymbol{y}_2^t\boldsymbol{x}_2) - i(\boldsymbol{y}_1^t\boldsymbol{x}_2 + \boldsymbol{y}_2^t\boldsymbol{x}_1).$$

De las identidades $\boldsymbol{y}_1^t\boldsymbol{x}_1+\boldsymbol{y}_2^t\boldsymbol{x}_2=0, \boldsymbol{y}_1^t\boldsymbol{x}_1-\boldsymbol{y}_2^t\boldsymbol{x}_2=0$ se deduce que $\boldsymbol{y}_1^t\boldsymbol{x}_1=\boldsymbol{y}_2^t\boldsymbol{x}_2=0$. Análogamente, $\boldsymbol{y}_1^t\boldsymbol{x}_2=\boldsymbol{y}_2^t\boldsymbol{x}_1=0$. En resumen, los vectores $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y}_1,\boldsymbol{y}_2$ son ortogonales dos a dos. Por ello, al considerar los vectores normalizados procedentes de cada autovector, obtenemos una base ortonormal del espacio, y la matriz de la aplicación lineal respecto de esta nueva base es de la forma requerida.

Ejemplo 7.4.4. Consideremos la matriz

$$A=\begin{pmatrix} 1 & 0 & 0 & 0 & 0\\ 0 & 3/5 & 0 & -4/5 & 0\\ 0 & 0 & 3/5 & 0 & 4/5\\ 0 & 4/5 & 0 & 3/5 & 0\\ 0 & 0 & -4/5 & 0 & 3/5 \end{pmatrix}$$

Los autovalores de la matriz A son $\lambda_1=1, \lambda_2=3/5+4/5i, \lambda_3=3/5-4/5i$. Entonces

$$V_1(\lambda_1)=\begin{cases} x_2=0\\ x_3=0\\ x_4=0\\ x_5=0 \end{cases} =\langle \boldsymbol{w}_1=(1,0,0,0,0)^t\rangle$$

$$V_1(\lambda_2)=\begin{cases} x_1=0\\ x_2=ix_4\\ x_3=ix_5 \end{cases} =\phi(\boldsymbol{w}_2=(0,0,1,0,i)^t, \boldsymbol{w}_3=(0,i,1,1,i)^t\rangle.$$

$$V_1(\lambda_3)=\begin{cases} x_1=0\\ x_2=-ix_4\\ x_3=-ix_5 \end{cases} =\phi(\boldsymbol{w}_4=(0,0,1,0,-i)^t, \boldsymbol{w}_5=(0,-i,1,1,-i)^t\rangle.$$

Para el autovalor real λ_1 tomamos $\boldsymbol{v}_{11}=(1,0,0,0,0)^t$. Para el autovalor complejo λ_2, consideramos la parte real y la parte imaginaria de sus autovectores asociados. Sean

$$\boldsymbol{w}_2=\boldsymbol{w}_{21}+i\boldsymbol{w}_{22}, \boldsymbol{w}_{21}=(0,0,1,0,0)^t, \boldsymbol{w}_{22}=(0,0,0,0,1)^t,$$
$$\boldsymbol{w}_3=\boldsymbol{w}_{31}+i\boldsymbol{w}_{32}, \boldsymbol{w}_{31}=(0,0,1,1,0)^t, \boldsymbol{w}_{32}=(0,1,0,0,1)^t.$$

Vemos que los vectores procedentes de los diferentes espacios de autovectores son ortogonales entre sí. Debemos aplicar Gram-Schmidt a estos vectores para obtener unos ortonormales. Tras el proceso nos queda

$$\boldsymbol{v}_2=(0,0,1,0,0)^t, \boldsymbol{v}_3=(0,0,0,0,1)^t, \boldsymbol{v}_4=(0,0,0,1,0)^t, \boldsymbol{v}_5=(0,1,0,0,0)^t.$$

Entonces la matriz de paso $P = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 & \boldsymbol{v}_5 \end{pmatrix}$ es ortogonal, y

$$P^t AP = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 3/5 & 4/5 & 0 & 0 \\ 0 & -4/5 & 3/5 & 0 & 0 \\ 0 & 0 & 0 & 3/5 & -4/5 \\ 0 & 0 & 0 & 4/5 & 3/5 \end{pmatrix}.$$

7.5. Formas cuadráticas

Las formas cuadráticas juegan un papel importante en inferencia estadística. Nos aparecerán al considerar ciertas expresiones de variables aleatorias que siguen una distribución normal o al tratar con matrices de covarianza.

El tratamiento que seguiremos se restringe a matrices reales, aunque admite una generalización natural para el caso complejo.

7.5.1. Definición y propiedades elementales

Forma cuadrática

Sea $A_{n\times n}$ una matriz real. La función $f : \mathbb{R}^n \to \mathbb{R}$ definida por

$$f(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

se denomina **forma cuadrática**.

Ejemplo 7.5.1. Consideremos la matriz

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

Entonces la forma cuadrática que se obtiene a partir de A es

$$\begin{aligned} f(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x} &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + 2x_3 & -x_1 + x_2 + 2x_3 & 2x_2 + 3x_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= x_1^2 - x_1x_2 + 2x_1x_3 + x_2^2 + 4x_2x_3 + 3x_3^2. \end{aligned}$$

Observemos que siempre se tiene que $\boldsymbol{x}^t A\boldsymbol{x} = \boldsymbol{x}^t \left(\frac{1}{2}(A + A^t)\right)\boldsymbol{x}$, y que la matriz $B = \frac{1}{2}(A + A^t)$ es simétrica. Por este motivo, siempre se puede asociar una forma cuadrática con una matriz simétrica. Por ejemplo, en el caso anterior

$$f(\boldsymbol{x}) = x_1^2 - x_1x_2 + 2x_1x_3 + x_2^2 + 4x_2x_3 + 3x_3^2 = \boldsymbol{x}^t \begin{pmatrix} 1 & -1/2 & 1 \\ -1/2 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix} \boldsymbol{x}.$$

Tenemos entonces una correspondencia biyectiva entre las formas cuadráticas y las matrices simétricas. Nuestro objetivo es extraer propiedades de la forma cuadrática a partir de la matriz simétrica asociada.

Tipos de formas cuadráticas

Sea $f : \mathbb{R}^n \to \mathbb{R}$ una forma cuadrática.

- f es definida positiva si $f(\boldsymbol{x}) > 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$.
- f es semidefinida positiva si $f(\boldsymbol{x}) \geq 0$ para todo $\boldsymbol{x}$ y existe $\boldsymbol{u} \neq \boldsymbol{0}$ tal que $f(\boldsymbol{u}) = 0$.
- f es definida negativa si $f(\boldsymbol{x}) < 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$.
- f es semidefinida negativa si $f(\boldsymbol{x}) \leq 0$ para todo $\boldsymbol{x}$ y existe $\boldsymbol{u} \neq \boldsymbol{0}$ tal que $f(\boldsymbol{u}) = 0$.
- f es indefinida si existe $\boldsymbol{u}$ tal que $f(\boldsymbol{u}) > 0$ y existe $\boldsymbol{v}$ tal que $f(\boldsymbol{v}) < 0$.

Ejemplo 7.5.2. Es importante tener presente la dimensión en la que la forma cuadrática está definida.

1. La forma $f_1 : \mathbb{R}^3 \to \mathbb{R}$ dada por $f_1(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2$ es definida positiva.
2. La forma $f_2 : \mathbb{R}^4 \to \mathbb{R}$ dada por $f_2(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2$ es semidefinida positiva, pues $f_2(\boldsymbol{e}_4) = 0$.
3. La forma $g_1 = -f_1$ es definida negativa.
4. La forma $g_2 = -f_2$ es semidefinida negativa.
5. La forma $h : \mathbb{R}^2 \to \mathbb{R}$ definida como $h(\boldsymbol{x}) = x_1^2 - x_2^2$ es indefinida, pues $h(\boldsymbol{e}_1) > 0, h(\boldsymbol{e}_2) < 0$.

Una matriz simétrica A es definida positiva si la forma cuadrática asociada lo es. Por tanto, la condición es equivalente a que $\boldsymbol{x}^t A\boldsymbol{x} > 0$ para todo $\boldsymbol{x} \neq \mathbf{0}$.

Nota 7.5.3. Sea B una matriz de orden $p \times n$, con $p \leq n$ y rango$(B) = p$. Entonces la matriz BB^t es simétrica (inmediato) y definida positiva; en efecto, sea $\boldsymbol{v}_{n\times 1} \neq \mathbf{0}$. Como rango$(B^t) = p$, se tiene que null$(B^t) = \mathbf{0}$ y esto implica que $B^t\boldsymbol{v} \neq \mathbf{0}$. Entonces

$$\boldsymbol{v}^t BB^t\boldsymbol{v} = \boldsymbol{w}^t\boldsymbol{w} > 0, \text{ donde } \boldsymbol{w} = B^t\boldsymbol{v}.$$

Si A es una matriz de orden $m \times n, m \geq n$ y rango$(A) = n$, entonces, de forma análoga, $A^t A$ es simétrica y definida positiva. En ambos casos, si el rango no es máximo, obtenemos matrices semidefinidas positiva. Estas matrices aparecen al tratar la covarianza de variables aleatorias y en el ajuste de puntos por mínimos cuadrados.

Propiedades elementales de las matrices simétricas definidas positiva

Sea A una matriz simétrica definida positiva. Se verifica que

1. A es no singular.
2. A^{-1} es simétrica definida positiva.
3. Los elementos diagonales a_{ii} son todos positivos.
4. traza$(A) > 0$.
5. Si B es semidefinida positiva, entonces $A + B$ es definida positiva.

7.5.2. Caracterización por los autovalores

Sabemos que los autovalores de una matriz simétrica son reales. Esto permite dar otra caracterización del carácter definida positiva.

Caracterización por los autovalores

Sea $A_{n\times n}$ una matriz simétrica.

- A es definida positiva si y solamente si todos los autovalores de A son positivos.
- A es semidefinida positiva si y solamente si sus autovalores son no negativos y alguno es nulo.
- A es indefinida si y solamente si la matriz A tiene algún autovalor positivo y otro negativo.

PRUEBA: Por el teorema espectral para las matrices simétricas, existe P matriz ortogonal tal que $P^tAP = D$, con D matriz diagonal y los autovalores λ_i en las posiciones de la diagonal; entonces $A = PDP^t$. Sea λ autovalor de A y $\boldsymbol{v}$ autovector asociado. Si A es definida positiva, se tiene que $\boldsymbol{v}^tA\boldsymbol{v} > 0$, y en este caso

$$0 < \boldsymbol{v}^tA\boldsymbol{v} = \boldsymbol{v}^t\lambda\boldsymbol{v} = \lambda(\boldsymbol{v}^t\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2, \text{ de donde } \lambda > 0.$$

Recíprocamente, supongamos que los autovalores λ_i son positivos. Dado $\boldsymbol{u} \neq \boldsymbol{0}$, obtenemos

$$\boldsymbol{u}^tA\boldsymbol{u} = \boldsymbol{u}^tPDP^t\boldsymbol{u} = \boldsymbol{w}^tD\boldsymbol{w} = \sum_{i=1}^{n}\lambda_i w_i^2 > 0, \text{ donde } \boldsymbol{w} = P^t\boldsymbol{u}.$$

Observemos que $\boldsymbol{w} \neq \boldsymbol{0}$, pues P es una matriz no singular. Las restantes equivalencias se prueban de forma análoga. □

- Una matriz simétrica definida positiva (negativa) es no singular.
- Una matriz simétrica semidefinida positiva (negativa) es singular.
- Una matriz simétrica definida positiva tiene una única raíz cuadrada.

La prueba anterior muestra un método para diagonalizar una forma cuadrática a partir del teorema espectral.

Autovalores de BB^t y B^tB

Sea B una matriz de orden $m \times n$ y rango$(B) = r$. Entonces

1. Las matrices BB^t y B^tB tienen los mismos r autovalores no nulos $\lambda_j, j = 1, \ldots, r$.
2. Si $\boldsymbol{v}$ es un autovector unitario de B^tB asociado al autovalor λ entonces el vector $\boldsymbol{u} = \frac{1}{\sqrt{\lambda}} B\boldsymbol{v}$ es autovector unitario de BB^t asociado al autovalor λ.

PRUEBA: La matriz B^tB es cuadrada de orden n y BB^t es cuadrada de orden m. Ambas son simétricas y semidefinidas negativas, de rango igual al de B. Por tanto, tienen r autovalores no nulos. Sea λ un autovalor no nulo de B^tB, y $\boldsymbol{v}$ un autovector asociado. Entonces $B^tB\boldsymbol{v} = \lambda\boldsymbol{v}$ y $BB^tB\boldsymbol{v} = \lambda B\boldsymbol{v}$. Sea $\boldsymbol{w}_{m\times 1} = B\boldsymbol{v}$. Si $\boldsymbol{w} = \boldsymbol{0}$, entonces tendríamos que $\boldsymbol{0}B^t\boldsymbol{w} = \lambda\boldsymbol{v}$, lo que implicaría que $\lambda = 0$. Entonces $\boldsymbol{w} \neq \boldsymbol{0}$ y $\boldsymbol{w}$ es autovector asociado a λ de la matriz BB^t. La otra inclusión es similar.

Como $\boldsymbol{w} = B\boldsymbol{v}$ es autovector de BB^t, entonces cualquier vector no nulo proporcional también lo es, como $\boldsymbol{u} = \frac{1}{\sqrt{\lambda}} B\boldsymbol{v}$. Si $\boldsymbol{v}$ es unitario, entonces

$$\boldsymbol{u}^t\boldsymbol{u} = \frac{1}{\sqrt{\lambda}}\boldsymbol{v}^tB^t\frac{1}{\sqrt{\lambda}}B\boldsymbol{v} = \frac{1}{\lambda}\boldsymbol{v}^tB^tB\boldsymbol{v} = \frac{1}{\lambda}\lambda\boldsymbol{v}^t\boldsymbol{v} = 1.$$

□

7.5.3. Factorización de Cholesky

Nuestro objetivo es probar que las matrices simétricas definidas positiva tienen una factorización que será de gran utilidad cuando estudiemos mínimos cuadrados.

Sea A una matriz simétrica de orden n, definida positiva, y X una matriz de orden $n \times m, n \geq m$, y rango$(X) = m$. Entonces X^tAX es simétrica definida positiva.

En primer lugar, el carácter simétrico se deduce de $(X^tAX)^t = X^tA^tX = X^tAX$.

Supongamos que tenemos un sistema de la forma $X\boldsymbol{x} = \boldsymbol{0}$. El vector $\boldsymbol{x}$ tiene m componentes. Como rango$(X) = m =$ número de incógnitas, este sistema tiene solución única, que es necesariamente $\boldsymbol{x} = \boldsymbol{0}$. Expresado de otra forma, si

$\boldsymbol{v}$ es un vector no nulo, entonces $X\boldsymbol{v} \neq \mathbf{0}$. Así,

$$\boldsymbol{v}^t(X^tAX)\boldsymbol{v} = (X\boldsymbol{v})^tA(X\boldsymbol{v}) > 0.$$

Nos fijamos ahora en unas submatrices especiales de la matriz A. Una *submatriz principal* de orden k de A es una matriz formada por las filas y columnas $i_1, i_2, \ldots, i_k$. Gráficamente la podemos ver como una submatriz que se *apoya* en la diagonal principal de A. Por ejemplo, si

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

y escogemos $i_1 = 1, i_2 = 3, i_3 = 4$, entonces

$$B = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 5 & 2 \\ 0 & 2 & 4 \end{pmatrix}.$$

Submatrices principales

Si A es simétrica definida positiva, entonces toda submatriz principal de A es simétrica definida positiva.

PRUEBA: Sea B la submatriz principal de orden k de A formada por las filas (y columnas) $i_1, i_2, \ldots, i_k$. Basta aplicar el resultado anterior a una matriz X de orden $n \times k$ con el vector $\boldsymbol{e}_{i_j}$ en la columna i_j. □

En particular, cada elemento diagonal de A es positivo. Por ejemplo, para extraer la submatriz principal de A formada por las $n-1$ últimas filas y columnas de A tomaremos

$$X = \begin{pmatrix} 0 \\ I_{n-1} \end{pmatrix}.$$

Factorización de Cholesky

Sea $A_{n\times n}$ una matriz simétrica definida positiva. Entonces existe $R_{n\times n}$ triangular superior, con entradas diagonales positivas, tal que $A = R^tR$. A esta descomposición se la conoce como ***factorización de Cholesky*** de A.

PRUEBA: La prueba es por inducción sobre n, el orden de la matriz A. Para $n=1$, tenemos que $A=(a)$, $a>0$, y basta tomar $R=(\sqrt{a})$ y $U=(\sqrt{a})$. Supongamos el resultado cierto para $n-1$, es decir, si A' es una matriz simétrica definida positiva de orden $n-1$, existe R_1 triangular superior con entradas diagonales positivas tal que $A_1=R_1^*R_1$. Sea entonces $n>1$; como $a_{11}>0$, podemos usarlo como pivote en la eliminación gaussiana. En concreto, expresamos la matriz A como

$$A=\begin{pmatrix} a_{11} & \boldsymbol{w}^* \\ \boldsymbol{w} & A_1 \end{pmatrix}.$$

Entonces

$$L_1A=\begin{pmatrix} 1 & \mathbf{0}^t \\ -\frac{1}{a_{11}}\boldsymbol{w} & I \end{pmatrix}\begin{pmatrix} a_{11} & \boldsymbol{w}^* \\ \boldsymbol{w} & A_1 \end{pmatrix}=\begin{pmatrix} a_{11} & \boldsymbol{w}^* \\ \mathbf{0} & -\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^*+A_1 \end{pmatrix}.$$

Por el carácter simétrico de A, podemos hacer también ceros en la primera fila con una transformación similar:

$$L_1AL_1^*=\begin{pmatrix} a_{11} & \boldsymbol{w}^* \\ \mathbf{0} & -\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^*+A_1 \end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{w}^* \\ \mathbf{0} & I_{n-1} \end{pmatrix}=\begin{pmatrix} a_{11} & \mathbf{0}^t \\ \mathbf{0} & -\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^*+A_1 \end{pmatrix}.$$

Como L_1 es una matriz triangular, con entradas en la diagonal iguales a 1, es no singular. Por la proposición anterior, la matriz $L_1AL_1^*$ es simétrica definida positiva, y la submatriz inferior $-\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^*+A_1$ es simétrica definida positiva, de orden $n-1$. Por la hipótesis de inducción, existe una matriz R_1 triangular superior con entradas diagonales positivas tal que $-\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^*+A_1=R_1^*R_1$. Podemos escribir entonces

$$L_1AL_1^*=\begin{pmatrix} \sqrt{a_{11}} & \mathbf{0}^t \\ \mathbf{0} & R_1^* \end{pmatrix}\begin{pmatrix} \sqrt{a_{11}} & \mathbf{0}^t \\ \mathbf{0} & R_1 \end{pmatrix}=R_2^*R_2,$$

donde R_2 es una matriz triangular superior. Se deduce que $A=L_1^{-1}R_2^*R_2(L_1^*)^{-1}=R^*R$, con $R=R_2(L_1^*)^{-1}$, que es triangular superior, con entradas diagonales positivas.

□

Es fácil ver que si A tiene una factorización de Cholesky, entonces es simétrica definida positiva. Si fijamos los signos de las raíces cuadradas, entonces la matriz R es única.

La factorización de Cholesky es, en el fondo, una eliminación gaussiana, con una pequeña modificación con respecto a los pivotes. La matriz de salida R es la forma triangular que queda en la eliminación gaussiana, salvo un factor.

El algoritmo de cálculo se puede expresar como

$R = A$.
for $k = 1$ **to** n
  for $j = k+1$ **to** n
    $r[j, j:n] = r[j, j:n] -$
      $r[k, j:n]\, r[k, j] / r[k, k]$
  end for
$r[k, k:n] = r[k, k:n] / \sqrt{r[k, k]}$
end for

No es necesario verificar de partida si la matriz A es definida positiva. Si en el algoritmo obtenemos un pivote negativo, ya no puede ser definida positiva. Si llegamos a un pivote nulo, entonces es que la matriz de partida es singular.

Coste de la factorización de Cholesky

Sea A una matriz simétrica definida positiva, de orden $n \times n$. Entonces la factorización de Cholesky precisa del orden de

$$\frac{1}{3} n^3 \text{ flops},$$

esto es, del orden de la mitad de la eliminación gaussiana.

Ejemplo 7.5.4. Sea

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

Evidentemente es una matriz simétrica. El carácter definida positiva lo extraeremos del propio algoritmo. Procedemos como si calculásemos la eliminación gaussiana.

$$A=\begin{pmatrix} 4 & 1 & -1 & 0 \\ 1 & 3 & -1 & 0 \\ -1 & -1 & 5 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix} \xrightarrow[F_3+\frac{1}{4}F_1]{F_2-\frac{1}{4}F_1} \begin{pmatrix} 4 & 1 & -1 & 0 \\ 0 & 11/4 & -3/4 & 0 \\ 0 & -3/4 & 19/4 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

$$\xrightarrow{F_3+\frac{3}{11}F_2} \begin{pmatrix} 4 & 1 & -1 & 0 \\ 0 & 11/4 & -3/4 & 0 \\ 0 & 0 & 50/11 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

$$\xrightarrow{F_4-\frac{11}{25}F_3} \begin{pmatrix} 4 & 1 & -1 & 0 \\ 0 & 11/4 & -3/4 & 0 \\ 0 & 0 & 50/11 & 2 \\ 0 & 0 & 0 & 78/25 \end{pmatrix}.$$

Observemos que todos los pivotes son positivos. Ahora dividimos *cada fila* por la raíz cuadrada del pivote correspondiente. En nuestro caso, la primera fila hay que dividirla por 2, la segunda fila por $\frac{\sqrt{11}}{2}$, la tercera por $\sqrt{\frac{50}{11}}$, y la cuarta por $\sqrt{\frac{78}{25}}$.

$$\begin{pmatrix} 4 & 1 & -1 & 0 \\ 0 & 11/4 & -3/4 & 0 \\ 0 & 0 & 50/11 & 2 \\ 0 & 0 & 0 & 78/25 \end{pmatrix} \begin{matrix} E_2(1,\sqrt{\frac{1}{4}}) \\ E_2(2,\sqrt{\frac{4}{11}}) \\ E_2(3,\sqrt{\frac{11}{50}}) \\ E_2(4,\sqrt{\frac{25}{78}}) \end{matrix} \rightarrow R=\begin{pmatrix} 2 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2}\sqrt{11} & -\frac{3}{22}\sqrt{11} & 0 \\ 0 & 0 & \frac{5}{11}\sqrt{22} & \frac{1}{5}\sqrt{22} \\ 0 & 0 & 0 & \frac{1}{5}\sqrt{78} \end{pmatrix}.$$

Es fácil comprobar que $A = R^t R$.

Nota 7.5.5. Si A es simétrica definida positiva, la factorización de Cholesky de A nos permite escribir $A = R^t R$, con R triangular superior. Sabemos que es posible escoger los elementos diagonales de R mayores que cero. Vamos a ver que, en tal caso, la matriz R es única.

En efecto, supongamos que $A = R_1^t R_1 = R_2^t R_2$, con R_i, $i = 1,2$ matrices triangulares superiores con elementos diagonales positivos. Entonces $(R_2^t)^{-1} R_1^t = R_2 R_1^{-1}$. La parte derecha de esta igualdad es una matriz triangular inferior, con elementos diagonales positivos, y la parte izquierda es una matriz triangular superior con elementos diagonales positivos. Entonces, $(R_2^t)^{-1} R_1^t = R_2 R_1^{-1} = D = \text{diag}(d_1, \ldots, d_n)$ es una matriz diagonal con elementos positivos.

Por un lado, $D = R_2 R_1^{-1}$, y por otro, $D = (R_2^t)^{-1} R_1^t = (R_2^{-1})^t R_1^t = (R_1 R_2^{-1})^t = (D^{-1})^t$. Entonces, para cada $i = 1, \ldots, n$ se verifica que $d_i = \frac{1}{d_i}$, de donde $d_i = 1, i = 1, \ldots, n$. En definitiva, $D = I_n$ y $R_1 = R_2$.

7.5.4. Caracterización por los menores principales

Existe un criterio basado en determinantes de ciertas submatrices para caracterizar a una matriz simétrica definida positiva. Dada $A_{n\times n}$ una matriz, notaremos por A_k la submatriz de A formada por las primeras k filas y columnas de A. Por ejemplo, $A_1 = a_{11}$ y $A_n = A$.

Menores principales positivos

Sea $A_{n\times n}$ una matriz simétrica y denotemos por A_k la submatriz principal de orden k.

- A es definida positiva si y solamente se $\det(A_k) > 0$ para todo $k = 1,\dots,n$.
- A es semidefinida positiva si y solamente si $\det(A_k) \geq 0$ para todo $k = 1,\dots,n$ y $\det(A_j) = 0$ para algún j.

PRUEBA: Haremos uso de la demostración usada en la factorización de Cholesky. En primer lugar, si A es una matriz cualquiera y L es una matriz asociada a una trasformación elemental de tipo I, entonces la matriz $B = LAL^t$ verifica que $\det(A_j) = \det(B_j)$, $j = 1,2,\dots,n$, donde B_j es la submatriz principal de orden j de B. Es inmediato por la invariancia del determinante por dichas trasformaciones.

Supongamos que A es simétrica definida positiva. Sabemos que todas sus submatrices principales son definidas positiva. Probemos el resultado por inducción sobre n. Para $n = 1$, la matriz A se reduce al escalar a_{11}, que es positivo. Supongamos que el resultado es cierto para $n-1$. Tal como hacíamos en la factorización de Cholesky, escribamos

$$A = \begin{pmatrix} a_{11} & \boldsymbol{w}^t \\ \boldsymbol{w} & A_1 \end{pmatrix}.$$

y consideremos la matriz

$$L_1 = \begin{pmatrix} 1 & \mathbf{0}^t \\ -\frac{1}{a_{11}}\boldsymbol{w} & I \end{pmatrix}.$$

Esta matriz es el producto de matrices de tipo I con coeficientes en la primera columna. Por tanto, la matriz $B = L_1 A L_1^t$ verifica que $\det(B_k) = \det(A_k)$, como

hemos comentado al principio. El efecto sobre la matriz A es obtener ceros en la primera columna y en la primera fila; en concreto,

$$L_1 A L_1^t = \begin{pmatrix} a_{11} & \mathbf{0}^t \\ \mathbf{0} & -\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^t + A_1 \end{pmatrix}.$$

Como L_1 es una matriz triangular, con entradas en la diagonal iguales a 1, es no singular, por lo que B es simétrica definida positiva y la submatriz principal $A' = -\frac{1}{a_{11}}\boldsymbol{w}\boldsymbol{w}^t + A_1$ de orden $n-1$ es simétrica definida positiva. Por hipótesis de inducción, para cada $k = 1, \ldots, n-1$, los determinantes $\det(A'_k)$ son positivos. Dado que $\det(B_j) = a_{11}\det(A'_{j-1}$, tenemos el resultado.

Recíprocamente, supongamos que $\det(A_j) > 0$ para todo $j = 1, 2, \ldots, n$. Esto implica que en el proceso de la factorización de Cholesky siempre vamos a encontrar que el elemento diagonal del paso j-ésimo es mayor que cero, es decir, es posible calcular R matriz triangular superior tal que $A = R^t R$ y los elementos diagonales $r_{ii} > 0$, y sabemos entonces que A es definida positiva. □

Ejemplo 7.5.6. Consideremos la matriz simétrica

$$A = \begin{bmatrix} 5 & 3 & 5 \\ 3 & 4 & 4 \\ 5 & 4 & 4 \end{bmatrix}.$$

Entonces

$$\det(A_1) = 5, \det(A_2) = \det\begin{pmatrix} 5 & 3 \\ 3 & 4 \end{pmatrix} = 11, \det(A_3) = \det(A) = -16,$$

por lo que A no es definida positiva.

Ejemplo 7.5.7. Para cada $n \geq 1$, se define la *matriz de Hilbert* H_n como

$$H_n = \left(\frac{1}{i+j-1}\right), \text{ con } i, j = 1, \ldots, n.$$

Por ejemplo,

$$H_2 = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}, H_3 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

Por la definición, H_n es una matriz simétrica, y la submatriz de H_n formada por las k primeras filas y columnas es H_k. Nuestro objetivo es probar que la matriz

H_n es definida positiva. Para ello, basta ver que para cada $m \geq 1$, el determinante de H_m es positivo, pues la caracterización por los menores principales implica el resultado.

Consideramos un problema más general, pero que permite resolver lo anterior de manera sencilla. Sean $a_1, \ldots, a_n, b_1, \ldots, b_n$ números cualesquiera, con $a_i + b_j \neq 0$, y definimos la matriz de orden n dada por $A = \left(\frac{1}{a_i+b_j}\right)$. Entonces

$$\det(A) = \prod_{\substack{j,k=1 \\ j>k}}^{n} (a_j - a_k)(b_j - b_k) \frac{1}{\prod_{j,k=1}^{n}(a_j + b_k)}.$$

La matriz de Hilbert corresponde al caso $a_i = i, b_j = j-1$, por lo que para $j > k$ se verifica que $a_j - a_k = j - k > 0, b_j - b_k = j - 1 - k + 1 > 0$, y $\det(H_m) > 0$ para todo m.

La prueba del resultado para la matriz A es por inducción sobre la dimensión. Para $n = 1$ es inmediato. Supongamos el resultado cierto para una matriz de dicho tipo con dimensión $n-1$. Efectuamos operaciones elementales que conservan el valor del determinante. En primer lugar, restamos a cada una de las filas $1, 2, \ldots, n-1$ la última. Observemos que

$$\frac{1}{a_i + b_j} - \frac{1}{a_n + b_j} = \frac{a_n - a_i}{(a_i + b_j)(a_n + b_j)},$$

por lo que podemos extraer de las columnas los factores respectivos

$$\frac{1}{a_n + b_1}, \frac{1}{a_n + b_2}, \ldots, \frac{1}{a_n + b_{n-1}}, \frac{1}{a_n + b_n},$$

y de las filas los factores respectivos

$$a_n - a_1, a_n - a_2, \ldots, a_n - a_{n-1}, 1.$$

El determinante que queda tiene la forma

$$\det \begin{pmatrix} \frac{1}{a_1+b_1} & \frac{1}{a_1+b_2} & \cdots & \frac{1}{a_1+b_n} \\ \frac{1}{a_2+b_1} & \frac{1}{a_2+b_2} & \cdots & \frac{1}{a_2+b_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{n-1}+b_1} & \frac{1}{a_{n-1}+b_2} & \cdots & \frac{1}{a_{n-1}+b_n} \\ 1 & 1 & \ldots & 1 \end{pmatrix}.$$

Ahora restamos a cada una de las columnas $1, 2, \ldots, n-1$ la última columna, y sacamos de las columnas y las filas los factores respectivos

$$b_n - b_1, b_n - b_2, \ldots, b_n - b_{n-1}, 1,$$

y

$$\frac{1}{a_1+b_n}, \frac{1}{a_2+b_n}, \ldots, \frac{1}{a_{n-1}+b_n}, 1.$$

Se desarrolla el determinante por la última fila, y se obtiene el determinante de una matriz de la misma forma que A, pero de orden $n-1$. Aplicamos entonces la hipótesis de inducción, y obtenemos el resultado.

7.5.5. * Raíz cuadrada

Raíz cuadrada de una matriz hermitiana definida positiva

Sea A una matriz hermitiana definida positiva. Entonces existe una única matriz S hermitiana definida positiva tal que $S^2 = A$.

PRUEBA: Sea A una matriz hermitiana definida positiva de orden n. Por el teorema espectral, existe P unitaria tal que $P^*AP = D$, con D matriz diagonal, y elementos $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$. Sea $D^{1/2}$ la matriz diagonal con entradas $\sqrt{\lambda_i}, i = 1,2,\ldots,n$. Entonces $D^{1/2}D^{1/2} = D$ y la matriz $S = PD^{1/2}P^*$ verifica que

$$S^2 = PD^{1/2}P^*PD^{1/2}P^* = PDP^* = A, \text{ y además es hermitiana.}$$

de donde tenemos la existencia. Veamos ahora la unicidad. Sea R una matriz hermitiana definida positiva tal que $R^2 = A$. Entonces los autovalores de R son $\sqrt{\lambda_i}, i = 1,2,\ldots,n$. Existe Q matriz unitaria tal que $Q^*RQ = D^{1/2}$, y $A = R^2 = QDQ^*$. Queremos probar que $R = S$. Para ello, observemos que

$$S = PD^{1/2}P^*, R = QD^{1/2}Q^*, R^2 = QDQ^* = A = PDP^* = S^2.$$

Si probamos que la igualdad $QDQ^* = PDP^*$ implica la igualdad $QD^{1/2}Q^* = PD^{1/2}Q^*$ tendremos el resultado. Partimos entonces de la igualdad $QDQ^* = PDP^*$, lo que implica que $P^*QDQ^*P = D$. Sea $M = P^*Q$, que es una matriz unitaria. Entonces $MDM^* = D$, o bien $MD = DM$. Debemos probar que $MD^{1/2}M^* = D^{1/2}$, con lo que tendremos el resultado.

Sea $M = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_n \end{pmatrix}$. Como $D = \begin{pmatrix} \lambda_1\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_2 & \ldots & \lambda_n\boldsymbol{e}_n \end{pmatrix}$, se verifica que

$$MD = M\begin{pmatrix} \lambda_1\boldsymbol{e}_1 & \lambda_2\boldsymbol{e}_2 & \ldots & \lambda_n\boldsymbol{e}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{u}_1 & \lambda_2\boldsymbol{u}_2 & \ldots & \lambda_n\boldsymbol{u}_n \end{pmatrix},$$
$$DM = \begin{pmatrix} D\boldsymbol{u}_1 & D\boldsymbol{u}_2 & \ldots & D\boldsymbol{u}_n \end{pmatrix},$$

por lo que $D\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, i = 1,2,\ldots,n$. Por tanto, las columnas de M son autovectores de la matriz D. Sean $\mu_1, \mu_2, \ldots, \mu_s$ los autovalores distintos de D, con $\mu_1 > \mu_2 > \cdots > \mu_s > 0$. Entonces, para ciertos valores $r_1, \ldots, r_s$ se tiene que

$$\begin{aligned} \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{r_1} \in & \quad V_1(\mu_1) = \langle \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{r_1} \rangle, \\ \boldsymbol{u}_{r_1+1}, \ldots, \boldsymbol{u}_{r_1+r_2} \in & \quad V_1(\mu_2) = \langle \boldsymbol{e}_{r_1+1}, \ldots, \boldsymbol{e}_{r_1+r_2} \rangle, \\ & \vdots \\ \boldsymbol{u}_{r_1+\cdots+r_s+1}, \ldots, \boldsymbol{u}_n \in & \quad V_1(\mu_s) = \langle \boldsymbol{e}_{r_1+\cdots+r_s+1}, \ldots, \boldsymbol{e}_n \rangle. \end{aligned}$$

Esto implica que M se puede particionar en la forma

$$M = \begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_s \end{pmatrix}, \text{ con cada } M_i \text{ unitaria.}$$

Entonces

$$\begin{aligned} MD^{1/2}M^* &= \begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_s \end{pmatrix} \begin{pmatrix} \sqrt{\mu_1} I & & & \\ & \sqrt{\mu_2} I & & \\ & & \ddots & \\ & & & \sqrt{\mu_s} I \end{pmatrix} \begin{pmatrix} M_1^* & & & \\ & M_2^* & & \\ & & \ddots & \\ & & & M_s^* \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\mu_1} M_1 & & & \\ & \sqrt{\mu_2} M_2 & & \\ & & \ddots & \\ & & & \sqrt{\mu_s} M_s \end{pmatrix} \begin{pmatrix} M_1^* & & & \\ & M_2^* & & \\ & & \ddots & \\ & & & M_s^* \end{pmatrix} = D^{1/2}, \end{aligned}$$

como queríamos demostrar. □

7.5.6. * Distancia de Mahalanobis

Supongamos que queremos calcular la distancia entre dos vectores $\boldsymbol{x}$ y $\boldsymbol{\mu}$ en $\mathbb{R}^m$, donde $\boldsymbol{x}$ es una observación de una distribución con vector de media $\boldsymbol{\mu}$ y matriz de covarianza Ω. Si queremos tener en cuenta el efecto de la covarianza, la distancia euclídea no sería apropiada, a menos que $\Omega = I_m$. Por ejemplo, si $m = 2$ y $\Omega = \text{diag}(0,5,2)$ entonces un valor grande de $(x_1 - \mu_1)^2$ sería más sorprendente que ese mismo valor en $(x_2 - \mu_2)^2$, porque la varianza de la primera componente de $\boldsymbol{x}$ es menor que la varianza de la segunda componente. Por ello, parece razonable definir una distancia que ponga más peso en $(x_1 - \mu_1)^2$ que en $(x_2 - \mu_2)^2$. Así, definimos la función distancia como

$$d_\Omega(\boldsymbol{x}, \boldsymbol{\mu}) = ((\boldsymbol{x} - \boldsymbol{\mu})^t \Omega^{-1} (\boldsymbol{x} - \boldsymbol{\mu}))^{1/2}.$$

Como Ω es definida positiva, la función anterior define una distancia, pues Ω^{-1} también es definida positiva. Esto es equivalente a probar que la aplicación $\|\boldsymbol{u}\|_\Omega = \boldsymbol{u}^t\Omega^{-1}\boldsymbol{u}$ es una norma. Esta distancia se usa en análisis discriminante, y la encontramos en el exponente de la función de densidad normal multivariante: si $\boldsymbol{x} \sim N_m(\boldsymbol{\mu}, \Omega)$ entonces su función de densidad es

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{m/2}\det(\Omega)^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\Omega^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

para todo $\boldsymbol{x} \in \mathbb{R}^m$.

Esta distancia recibe el nombre de ***distancia de Mahalanobis***. Está relacionada con la euclídea a través de una transformación lineal. Supongamos que tenemos varios vectores $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ en $\mathbb{R}^3$, que son observaciones de distribuciones, que comparten la misma matriz de covarianza Ω. Si estamos interesados en cómo estos vectores difieren de los demás, entonces un dibujo en $\mathbb{R}^3$ puede ayudar. Sin embargo, por lo que hemos visto antes, si Ω no es la matriz identidad, la distancia euclídea no es adecuada, por lo que es difícil comparar e interpretar las diferencias observadas entre los r puntos. Vamos entonces a efectuar una trasformación lineal para que la distancia euclídea nos valga, y que es válido para vectores en $\mathbb{R}^n$. Como Ω es definida positiva, existe, por la factorización de Cholesky, una matriz no singular B tal que $\Omega = BB^t$ (B es triangular inferior). Tomemos las nuevas variables $\boldsymbol{u}_i = B^{-1}\boldsymbol{x}_i$. Entonces

$$\begin{aligned} d_\Omega(\boldsymbol{x}_i, \boldsymbol{x}_j) &= ((\boldsymbol{x}_i - \boldsymbol{x}_j)^t\Omega^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)) \\ &= ((\boldsymbol{x}_i - \boldsymbol{x}_j)^t(B^{-1})^tB^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)) \\ &= ((B^{-1}\boldsymbol{x}_i - B^{-1}\boldsymbol{x}_j)^t(B^{-1}\boldsymbol{x}_i - B^{-1}\boldsymbol{x}_j)) \\ &= ((\boldsymbol{u}_i - \boldsymbol{u}_j)^t(\boldsymbol{u}_i - \boldsymbol{u}_j)) = d(\boldsymbol{u}_i, \boldsymbol{u}_j). \end{aligned}$$

Además, la varianza de las $\boldsymbol{u}_i$ es igual a

$$\begin{aligned} \operatorname{var}(\boldsymbol{u}_i) &= \operatorname{var}(B^{-1}\boldsymbol{x}_i) = B^{-1}\operatorname{var}\boldsymbol{x}_i(B^{-1})^t \\ &= B^{-1}\Omega(B^{-1})^t = B^{-1}BB^t(B^{-1})^t = I_n \end{aligned}$$

Así, la transformación $\boldsymbol{u}_i = B^{-1}\boldsymbol{x}_i$ produce vectores en los que la distancia euclídea es una medida adecuada de la distancia entre puntos.

7.6. Descomposición en valores singulares

Descomposición en valores singulares (SVD)

Sea A una matriz compleja (real) de orden $m \times n$. Entonces A se puede factorizar como $A = U\Sigma V^*$, donde U es una matriz unitaria (ortogonal) de orden m, V es una matriz unitaria (ortogonal) de orden n y Σ es una matriz de orden $m \times n$ de la forma $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$, $p = \text{mín}\{m, n\}$, y $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$.

PRUEBA: Consideremos la matriz A^*A, de orden n. Es hermitiana (simétrica) semidefinida positiva, y sus autovalores son reales mayores o iguales que cero. Los ordenamos en forma decreciente, y supongamos que $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r > 0$ son los positivos, donde $r = \text{rango}(A)$. Sea $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ una base ortonormal de autovectores de A^*A, y llamemos

$$\sigma_i = \sqrt{\lambda_i}, i = 1, \ldots, p = \text{mín}\{m, n\}, \boldsymbol{u}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i \in \mathbb{K}^m, i = 1, \ldots, r.$$

Observemos que los vectores $\boldsymbol{u}_i$ están bien definidos, pues $r \leq p$. Tenemos que

$$\boldsymbol{u}_i \bullet \boldsymbol{u}_i = \frac{1}{\lambda_i} \boldsymbol{v}_i^t A^* A\boldsymbol{v}_i = 1, \text{ para } i = 1, \ldots, r,$$

y si $i \neq j$, entonces

$$\boldsymbol{u}_i \bullet \boldsymbol{u}_j = \frac{1}{\sigma_i \sigma_j} \boldsymbol{v}_j^* A^* A\boldsymbol{v}_i = \frac{\lambda_i}{\sigma_i \sigma_j} (\boldsymbol{v}_i \bullet \boldsymbol{v}_j) = 0.$$

Por tanto, los vectores $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ son unitarios y ortogonales entre sí. Completamos, mediante Gram-Schmidt o QR, a una base ortonormal del espacio

$$\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r, \boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m\}.$$

Sean

$$U = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m \end{pmatrix}, V = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{pmatrix}.$$

Vamos a probar que $U^*AV = \Sigma$. Por lo anterior,

$$\begin{aligned} U^*AV &= U^*A\begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{pmatrix} = U^*\begin{pmatrix} A\boldsymbol{v}_1 & A\boldsymbol{v}_2 & \ldots & A\boldsymbol{v}_n \end{pmatrix} \\ &= U^*\begin{pmatrix} \sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 & \ldots & \sigma_r\boldsymbol{u}_r & A\boldsymbol{v}_{r+1} & \ldots & A\boldsymbol{v}_n \end{pmatrix}. \end{aligned}$$

Observemos las columnas $A\boldsymbol{v}_j$, $j = r+1,\ldots,n$. Sabemos, por la ordenación de los autovalores, que $A^*A\boldsymbol{v}_j = \mathbf{0}$ para $j = r+1,\ldots,n$. Como $\text{null}(A^*A) = \text{null}(A)$, resulta que $A\boldsymbol{v}_j = \mathbf{0}$, para todo $j = r+1,\ldots n$. Entonces

$$\begin{aligned} U^*AV &= U^*\begin{pmatrix} \sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 & \ldots & \sigma_r\boldsymbol{u}_r & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{u}_1^* \\ \boldsymbol{u}_2^* \\ \vdots \\ \boldsymbol{u}_m^* \end{pmatrix}\begin{pmatrix} \sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 & \ldots & \sigma_r\boldsymbol{u}_r & \mathbf{0} & \ldots & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix} = \Sigma_{m\times n}, \end{aligned}$$

y tenemos el resultado. □

Los números $\sigma_1,\ldots,\sigma_p$ se denominan ***valores singulares*** de la matriz A.

Ejemplo 7.6.1. Vamos a calcular la descomposición en valores singulares de la matriz

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

1. Los autovalores de la matriz $A^tA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ son $\lambda_1 = 3, \lambda_2 = 1$.
2. Determinamos los no nulos, y ordenamos. En este caso, $r = 2$.
3. Los espacios de autovectores de A^tA son
   - Para λ_1,

$$\lambda_1 I - A^tA = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

$$\boldsymbol{w}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{v}_1 = \frac{1}{\|\boldsymbol{w}_1\|}\boldsymbol{w}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}.$$

- Para λ_2,

$$\lambda_2 I - A^t A = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$
$$\boldsymbol{w}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \boldsymbol{v}_2 = \frac{1}{\|\boldsymbol{w}_2\|}\boldsymbol{w}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}.$$

4. Una base ortonormal de autovectores de $A^t A$ es

$$\boldsymbol{v}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix},$$

   con $\boldsymbol{v}_i$ asociado a λ_i.

5. La matriz de valores singulares es

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \sigma_1 = \sqrt{3}, \sigma_2 = 1.$$

6. Definimos

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \frac{\sqrt{3}}{3} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} \sqrt{6}/3 \\ \sqrt{6}/6 \\ \sqrt{6}/6 \end{pmatrix},$$
$$\boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \begin{pmatrix} 0 \\ \sqrt{2}/2 \\ -\sqrt{2}/2 \end{pmatrix}.$$

7. Completamos $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ a una base ortonormal de $\mathbb{R}^3$. Para ello, calculamos una base del espacio ortogonal al subespacio $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle$, y se construye una ortonormal con Gram-Schmidt.

   El espacio ortogonal a $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle$ está definido por el sistema lineal homogéneo

$$\begin{cases} \frac{\sqrt{6}}{3}x_1 + \frac{\sqrt{6}}{6}x_2 + \frac{\sqrt{6}}{6}x_3 = 0, \\ \frac{\sqrt{2}}{2}x_2 - \frac{\sqrt{2}}{2}x_3 = 0. \end{cases}$$

   Calculamos la forma escalonada reducida por filas de la matriz de coeficientes:

$$\begin{bmatrix} \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

Entonces

$$\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle^{\perp} = \langle \boldsymbol{w}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \rangle.$$

Ahora aplicamos Gram-Schmidt a esta base. En este caso, solamente hay que normalizar el vector:

$$\boldsymbol{u}_3 = \frac{1}{\|\boldsymbol{w}_3\|_2} \boldsymbol{w}_3 = \begin{bmatrix} -1/3\sqrt{3} \\ 1/3\sqrt{3} \\ 1/3\sqrt{3} \end{bmatrix}.$$

8. Las matrices calculadas son

$$U = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{pmatrix} = \begin{pmatrix} \sqrt{6}/3 & 0 & -\sqrt{3}/3 \\ \sqrt{6}/6 & \sqrt{2}/2 & \sqrt{3}/3 \\ \sqrt{6}/6 & -\sqrt{2}/2 & \sqrt{3}/3 \end{pmatrix},$$
$$V = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{pmatrix} = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

y $A = U\Sigma V^t$.

Ejemplo 7.6.2.

$$A = \begin{pmatrix} 2 & 1 & -2 \end{pmatrix}.$$

1. Autovalores de A^tA: $\lambda_1 = 9, m_1 = 1, \lambda_2 = 0, m_2 = 2$.

2. Valores singulares: $\sigma_1 = 3$. Rango $r = 1$.

3. Autovectores de A^tA. Para λ_1,

$$\lambda_1 I - A^tA = \begin{bmatrix} 5 & -2 & 4 \\ -2 & 8 & 2 \\ 4 & 2 & 5 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{w}_1 = \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix}, \boldsymbol{v}_1 = \frac{1}{\|\boldsymbol{w}_1\|} \boldsymbol{w}_1 = \frac{2}{3} \boldsymbol{w}_1 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Para λ_2,

$$\lambda_2 I - A^t A = \begin{bmatrix} -4 & -2 & 4 \\ -2 & -1 & 2 \\ 4 & 2 & -4 \end{bmatrix} \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 1/2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{w}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{w}_3 = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}.$$

Observemos que los vectores que conforman $V_1(\lambda_2)$ no son ortogonales entre sí. Aplicamos Gram-Schmidt o factorización QR para conseguir una base ortogonal de dicho espacio.

$$\boldsymbol{v}_2' = \boldsymbol{w}_2,$$
$$\boldsymbol{v}_3' = \boldsymbol{w}_3 - \lambda_{12}\boldsymbol{v}_1',$$

$$\lambda_{12} = \frac{\boldsymbol{w}_3 \bullet \boldsymbol{v}_1'}{\boldsymbol{v}_1' \bullet \boldsymbol{v}_1'} = -\frac{1}{4},$$

$$\boldsymbol{v}_3' = \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} + \frac{1}{4}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}.$$

Y ahora normalizamos:

$$\boldsymbol{v}_2 = \frac{1}{\|\boldsymbol{v}_2'\|}\boldsymbol{v}_2' = \begin{bmatrix} 1/2\sqrt{2} \\ 0 \\ 1/2\sqrt{2} \end{bmatrix}, \boldsymbol{v}_3 = \frac{1}{\|\boldsymbol{v}_3'\|}\boldsymbol{v}_3' = \begin{bmatrix} -1/6\sqrt{2} \\ 2/3\sqrt{2} \\ 1/6\sqrt{2} \end{bmatrix}.$$

Una base ortonormal de autovectores de $A^t A$ es $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$.

4.
$$V = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix} = \begin{bmatrix} -2/3 & 1/2\sqrt{2} & -1/6\sqrt{2} \\ -1/3 & 0 & 2/3\sqrt{2} \\ 2/3 & 1/2\sqrt{2} & 1/6\sqrt{2} \end{bmatrix}.$$

5. Matriz de valores singulares:
$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \end{pmatrix}.$$

6. $\boldsymbol{u}_1 = \frac{1}{3}A\boldsymbol{v}_1 = (-1), U = (\boldsymbol{u}_1)$.

7. $A = U\Sigma V^t$.

Propiedades matriciales de la SVD

Sea $A_{m\times n} = U\Sigma V^*$ una descomposición en valores singulares de A, con $\sigma_r \neq 0, \sigma_{r+1} = 0$. Entonces

- rango$(A) = r$.
- Col$(A) = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle$, null$(A) = \langle \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n \rangle$, y estos conjunto son bases ortonormales de cada espacio.
- Col$(A^*) = \langle \boldsymbol{v}_1, \ldots, \boldsymbol{v}_r \rangle$, null$(A^*) = \langle \boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m \rangle$, y estos conjunto son bases ortonormales de cada espacio.

PRUEBA: La multiplicación por matrices no singulares no altera el rango. Como $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ es una base, entonces

$$\text{Col}(A) = \langle A\boldsymbol{v}_1, \ldots, A\boldsymbol{v}_n \rangle = \langle \sigma_1 \boldsymbol{u}_1, \ldots \sigma_n \boldsymbol{u}_n \rangle = \langle \boldsymbol{u}_1, \ldots, \boldsymbol{u}_r \rangle.$$

Por otra parte, como null(A) = null(ΣV^*) y null$(\Sigma) = \langle \boldsymbol{e}_{r+1}, \ldots, \boldsymbol{e}_n \rangle$, nos queda que

$$\begin{aligned} \text{null}(\Sigma V^*) &= \{\boldsymbol{x} \mid \Sigma V^* \boldsymbol{x} = \boldsymbol{0}\} = \{\boldsymbol{x} \mid V^* \boldsymbol{x} \in \text{null}(\Sigma)\} \\ &= \{\boldsymbol{x} \mid V^* \boldsymbol{x} \in \langle \boldsymbol{e}_{r+1}, \ldots, \boldsymbol{e}_n \rangle\} = \langle V\boldsymbol{e}_{r+1}, \ldots, V\boldsymbol{e}_n \rangle \\ &= \langle \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n \rangle. \end{aligned}$$

Dado que $A^* = V\Sigma^* U^* = V\Sigma^t U^*$, es una descomposición en valores singulares de A^*, y basta aplicar lo visto para la imagen y el espacio nulo. □

Nota 7.6.3.

- Sea $m \geq n$. La descomposición $A = U\Sigma V^*$, con U, V unitarias y Σ diagonal del mismo orden que A se puede poner también como $A = \hat{U}\hat{\Sigma}V^*$, con $\hat{U}$ de columnas ortonormales, de la misma dimensión que A, $\hat{\Sigma}$ diagonal y V^* unitaria. Basta quitar las últimas columnas de U y las últimas filas de Σ. Esta factorización recibe el nombre de SVD *reducida*. Podemos hacer algo análogo si $m < n$ eliminando las últimas columnas de Σ y las correspondientes filas de V^*. Como curiosidad, MATLAB hace esta operación cuando $m \geq n$, pero no cuando $m < n$. Esto se debe a que lo más frecuente corresponde a la primera opción, pues es lo habitual en las matrices de mínimos cuadrados.

- Si rango$(A) = r$, consideremos la SVD completa $A = U\Sigma V^*$. Tenemos otra forma de escribir la descomposición anterior. Sea $U = (U_1|U_2)$ y $V = (V_1|V_2)$, con U_1 de orden $m \times r$, y V_1 de orden $n \times r$. Entonces

$$U_1^* U_1 = I_r, Q_1^* Q_1 = I_r, \text{ y } A = U_1 \tilde{\Sigma} V_1^*,$$

  donde $\tilde{\Sigma}$ es la matriz diagonal con los valores singulares *no nulos* de A. Nos referiremos a esta expresión como SVD *corta* de A, y la notaremos $A = \tilde{U}\tilde{\Sigma}\tilde{V}^*$.

- Si $r =$ rango(A), entonces tenemos que $\sigma_r \neq 0, \sigma_{r+1} = 0$, y se verifica que

$$\begin{aligned} A &= \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_r \end{pmatrix} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_r \end{pmatrix}^* \\ &= \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^*. \end{aligned}$$

  Esta expresión recibe el nombre de SVD *compacta*, que no es más que otra forma de escribir la SVD corta.

El método que hemos empleado para encontrar la SVD de una matriz se podría usar como algoritmo para calcularla. Sin embargo, existen procedimientos más sofisticados para el cálculo práctico, que no expondremos aquí ([**?**]).

Nota 7.6.4. Cuando se habla del rango de una matriz, se suele usar lo que se conoce como ***rango numérico***, que es igual a $\tilde{r}$ para $\sigma_{\tilde{r}} > \epsilon \geq \sigma_{\tilde{r}+1}$, donde ϵ es un valor que se establece como límite de valores nulos. Vemos, en consecuencia, que tres objetos básicos de una matriz, como el espacio imagen, el espacio nulo y el rango se calculan, en realidad, a través de la descomposición en valores singulares.

7.7. * Descomposición de Jordan

La factorización que consideramos ahora de una matriz cuadrada A intenta llevar la matriz a una forma lo más parecida a una diagonal.

Un bloque de Jordan de orden k es una matriz cuadrada con k filas y columnas que tiene todos los elementos de la diagonal idénticos, la linea por encima de la diagonal está formada por unos y los restantes elementos son cero. En forma simbólica, $B = (b_{ij})$ es un bloque de Jordan de orden k si

$$b_{ij} = \begin{cases} \lambda & \text{si} \quad i = j \\ 1 & \text{si} \quad i+1 = j \\ 0 & \text{en el resto} \end{cases}, B = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}.$$

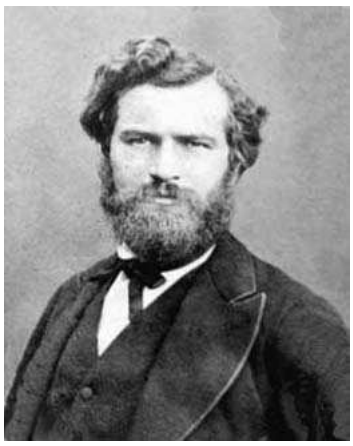

Figura 7.2: M.E. Camille Jordan (1838-1922)

Un segmento de Jordan $J(\lambda_1)$ es una matriz diagonal por bloques

$$J(\lambda_1) = \begin{pmatrix} J_1(\lambda_1) & 0 & \dots & 0 \\ 0 & J_2(\lambda_1) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{t_1}(\lambda_1) \end{pmatrix},$$

donde cada caja $J_k(\lambda_1)$ es un bloque de Jordan. Una matriz de Jordan es una matriz diagonal por bloques de manera que cada bloque es un segmento de Jordan, esto es, una matriz J es de Jordan si

$$J = \begin{pmatrix} J(\lambda_1) & 0 & \dots & 0 \\ 0 & J(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J(\lambda_r) \end{pmatrix},$$

donde cada $J(\lambda_i)$ es un segmento de Jordan.

Ejemplo 7.7.1. Un bloque de Jordan de orden 1 es un número. Un bloque de Jordan de orden 2 es de la forma

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

y uno de orden 3

$$\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

Una matriz de Jordan es, por ejemplo,

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

Forma canónica de Jordan

Para cada matriz $A \in \mathbb{C}^{n\times n}$, con autovalores distintos $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_s\}$, existe una matriz no singular P tal que

$$P^{-1}AP = J = \begin{pmatrix} J(\lambda_1) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & J(\lambda_2) & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & J(\lambda_s) \end{pmatrix},$$

donde J tiene un *segmento de Jordan* $J(\lambda_j)$ por cada autovalor $\lambda_j \in \sigma(A)$.

La matriz J anterior se denomina ***forma canónica de Jordan*** de A. La estructura de esta forma es única en el sentido de que el número de segmentos de Jordan en J, así como el número y tamaño de los bloques de Jordan en cada segmento está unívocamente determinado por la matriz A. Además, cada matriz semejante a A tiene la misma estructura de Jordan, es decir, $A, B \in \mathbb{C}^{n\times n}$ son semejantes si y solamente si A y B tienen la misma estructura de Jordan. La matriz P no es única.

La búsqueda de la forma canónica de Jordan es el cálculo de J y una matriz no singular P tales que $J = P^{-1}AP$, que es lo mismo que pedir $PJ = AP$. Esto significa que buscamos una base de vectores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, formando una cadena encabezada por un autovector. Para cada i, se tiene que verificar

$$A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \text{ o bien } A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i + \boldsymbol{x}_{i-1}.$$

Los vectores $\boldsymbol{x}_i$ forman las columnas de la matriz de paso P, y cada cadena forma un bloque de Jordan. Lo que vamos a probar es cómo se pueden construir estas cadenas para una matriz dada $A_{n\times n}$.

Procedemos por inducción. Para $n = 1$, la matriz ya está en forma canónica de Jordan. Supongamos entonces que la construcción se tiene para matrices de orden menor que n.

Paso 1. Supongamos que A es singular, lo que significa que tiene el autovalor $\lambda = 0$. El espacio Col(A) tiene dimensión $r < n$, y consideramos una base de Col(A) formada por los vectores $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r$ tales que

$$A\boldsymbol{w}_i = \lambda_i \boldsymbol{w}_i \text{ o bien } A\boldsymbol{w}_i = \lambda_i \boldsymbol{w}_i + \boldsymbol{w}_{i-1}. \tag{7.7.1}$$

Paso 2. Supongamos que $L = \text{null}(A) \cap \text{Col}(A)$ tiene dimensión p. Todo vector de null(A) es autovector asociado al autovalor $\lambda = 0$. Entonces tiene que haber p cadenas en el paso anterior que empiezan en estos autovectores. Nos interesan los vectores $\boldsymbol{w}_i$ que van al final de las cadenas. Cada uno de estos p vectores está en Col(A), por lo que existen $\boldsymbol{y}_i$ tales que $\boldsymbol{w}_i = A\boldsymbol{y}_i$, $i = 1, \ldots, p$. Por claridad, lo escribiremos como $A\boldsymbol{y}_i = 0\boldsymbol{y}_i + \boldsymbol{w}_i$.

Paso 3. El espacio nulo null(A) tiene dimensión $n - r$. Entonces, aparte de los p vectores independientes que podemos encontrar en su intersección con Col(A), hay $n-r-p$ vectores adicionales $\boldsymbol{z}_i$ que están fuera de esta intersección.

Ponemos juntos estos tres pasos para dar el teorema de Jordan:

Los r vectores $\boldsymbol{w}_i$, los p vectores $\boldsymbol{y}_i$, y los $n - r - p$ vectores $\boldsymbol{z}_i$ forman cadenas de Jordan, y son linealmente independientes.

Si renumeramos los vectores como $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, cada $\boldsymbol{y}_i$ debe ir inmediatamente del vector $\boldsymbol{w}_i$ del que procede. Ellos completan una cadena para $\lambda_i = 0$. Los vectores $\boldsymbol{z}_i$ van al final del todo, cada uno en su propia cadena, y dan lugar a cajas de orden 1.

Los bloques con autovalores no nulos se completan en el primer paso, los bloques con autovalor cero aumenta su tamaño en una fila y columna en el paso 2, y el paso 3 contribuye con bloques de orden 1 del tipo $J_i = [0]$.

Lo que tenemos que probar en primer lugar es que el conjunto $\{\boldsymbol{w}_i, \boldsymbol{y}_j, \boldsymbol{z}_k\}$ es linealmente independiente. Escribamos entonces

$$\sum \alpha_i \boldsymbol{w}_i + \sum \beta_i \boldsymbol{y}_i + \sum \gamma_i \boldsymbol{z}_i = \boldsymbol{0}. \tag{7.7.2}$$

Multiplicamos por la matriz A, y recordamos las ecuaciones 7.7.1 para los vectores $\boldsymbol{w}_i$, así como la relación $A\boldsymbol{z}_i = \boldsymbol{0}$. Entonces

$$\sum \alpha_i \begin{bmatrix} \lambda_i \boldsymbol{w}_i \\ \text{o bien} \\ \lambda \boldsymbol{w}_i + \boldsymbol{w}_{i-1} \end{bmatrix} + \sum \gamma_i A\boldsymbol{y}_i = \boldsymbol{0}. \tag{7.7.3}$$

Los vectores $A\boldsymbol{y}_i$ son los vectores especiales $\boldsymbol{w}_i$ del final de las cadenas, correspondientes a $\lambda_i = 0$, por lo que no pueden aparecer en la primera suma (están

multiplicados por cero en $\lambda_i \boldsymbol{w}_i$). Como 7.7.3 es una combinación lineal de vectores $\boldsymbol{w}_i$, cada coeficiente es nulo, en concreto $\beta_i = 0$. Si volvemos a 7.7.2, nos queda que

$$\sum \alpha_i \boldsymbol{w}_i = -\sum \gamma_i \boldsymbol{z}_i.$$

el lado izquierdo está en Col(A), los vectores $\boldsymbol{z}_i$ los escogimos fuera de dicho espacio. Entonces $\gamma_i = 0$, y por la independencia de los $\boldsymbol{w}_i$ tenemos también que $\alpha_i = 0$.

Si la matriz A no fuera singular, aplicamos los tres pasos a $A' = A - \lambda I$, con λ autovalor. El método anterior calcula P y J' con $P^{-1}A'P = J'$, con J' forma canónica de Jordan de A'. Entonces

$$P^{-1}AP = P^{-1}A'P + \lambda P^{-1}P = J' + \lambda I = J,$$

que es la forma canónica de A.

Esto completa la prueba de que toda matriz A es semejante a una matriz de Jordan J. Excepto por la ordenación de los bloques, es semejante a una única J.

Ejemplo 7.7.2. Consideremos la matriz

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

de autovalores $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = 2$, con multiplicidades algebraicas respectivas $m_1 = 1, m_2 = 1, m_3 = 2$. Con respecto a los dos primeros autovalores, bastará encontrar un generador del espacio de autovectores asociado a cada uno, por lo que el proceso de inducción lo haremos para λ_3. Sea $T_1 = A - \lambda_3 I$, que sabemos que es singular, y consideramos la restricción de T_1 a Col(T_1). Vamos a calcular la forma y base canónica de esta restricción.

Paso 1. Cálculo en Col(T_1). Mediante la forma reducida por filas se tiene que

$$\text{Col}(T_1) = \langle \boldsymbol{a}_1 = \begin{pmatrix} -1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{a}_2 = \begin{pmatrix} 2 \\ -2 \\ -1 \\ -1 \end{pmatrix}, \boldsymbol{a}_3 = \begin{pmatrix} 2 \\ -2 \\ -1 \\ 1 \end{pmatrix} \rangle$$

es una base. Se tiene que

$$\begin{aligned} T_1(\boldsymbol{a}_1) &= 0 \cdot \boldsymbol{a}_1, \\ T_1(\boldsymbol{a}_2) &= 3\boldsymbol{a}_1 - 3\boldsymbol{a}_2, \\ T_1(\boldsymbol{a}_3) &= \boldsymbol{a}_1 - \boldsymbol{a}_2 - 2\boldsymbol{a}_3, \end{aligned}$$

por lo que la matriz de la restricción de T_1 respecto de esta base es

$$M(T_1) = \begin{pmatrix} 0 & 3 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & -2 \end{pmatrix}.$$

Los autovalores de $M(T_1)$ son $\mu_1 = \lambda_1 - 2 = -2, \mu_2 = \lambda_2 - 2 = -3, \mu_3 = \lambda_3 - 2 = 0$, con multiplicidades algebraicas iguales a 1. Entonces $M(T_1)$ es diagonalizable, y una base de autovectores es

$$\boldsymbol{w}_1 = -\boldsymbol{a}_1 + \boldsymbol{a}_2 - \boldsymbol{a}_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{w}_2 = -\boldsymbol{a}_1 + \boldsymbol{a}_2 = \begin{pmatrix} 3 \\ -3 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{w}_3 = \boldsymbol{a}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Paso 2. Cálculo de $\mathrm{Col}(T_1) \cap \mathrm{null}(T_1)$. Se tiene que $\mathrm{Col}(T_1) \cap \mathrm{null}(T_1) = \langle \boldsymbol{w}_3 \rangle$, y como está en la imagen de la aplicación existe $\boldsymbol{y}_3$ tal que $T_1(\boldsymbol{y}_3) = \boldsymbol{w}_3$. En este caso,

$$\boldsymbol{y}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Paso 3. Adición de $n - r - p$ vectores. En nuestro caso, no hay más vectores que añadir.

Entonces una base canónica es $\mathscr{B}_J = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \boldsymbol{y}_3\}$, y se verifica que

$$\begin{array}{rcl} T_1(\boldsymbol{w}_1) & = & -2\boldsymbol{w}_1, \\ T_1(\boldsymbol{w}_2) & = & -3\boldsymbol{w}_2, \\ T_1(\boldsymbol{w}_3) & = & 0 \cdot \boldsymbol{w}_3, \\ T_1(\boldsymbol{y}_3) & = & \boldsymbol{w}_3. \end{array}$$

Como $T_1 = A - \lambda_3 I$, tenemos que

$$\begin{array}{rcl} A\boldsymbol{w}_1 & = & 0 \cdot \boldsymbol{w}_1, \\ A\boldsymbol{w}_2 & = & -\boldsymbol{w}_2, \\ A\boldsymbol{w}_3 & = & 2\boldsymbol{w}_3, \\ A\boldsymbol{y}_3 & = & \boldsymbol{w}_3 + 2\boldsymbol{y}_3. \end{array}$$

Por tanto, la forma canónica es la matriz de la aplicación lineal A respecto de la base $\mathscr{B}_J$, esto es,

$$M_{\mathscr{B}_J} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

7.8. Potencias de matrices

Tal como hemos visto en sistemas dinámicos discretos, el comportamiento a largo plazo del sistema depende de las potencias de la matriz asociada. Vamos a aprovechar la forma canónica de Jordan de la matriz para estudiar ese comportamiento.

Sea A una matriz cuadrada y J su forma canónica de Jordan. Entonces existe P no singular tal que $J = P^{-1}AP$, o bien $A = PJP^{-1}$, y

$$A^m = (PJP^{-1})(PJP^{-1})\dots(PJP^{-1}) = PJ^mP^{-1}.$$

Así, se reduce el cálculo de la potencia m-ésima de A al de la potencia m-ésima de J, que, como veremos, es más sencilla de calcular.

Potencia de un bloque de Jordan

Sea B un bloque de Jordan de orden s, con λ su elemento diagonal. Entonces

$$B^m = \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \binom{m}{2}\lambda^{m-2} & \dots & \binom{m}{s-1}\lambda^{m-s+1} \\ 0 & \lambda^m & \binom{m}{1}\lambda^{m-1} & \dots & \binom{m}{s-2}\lambda^{m-s+2} \\ 0 & 0 & \lambda^n & \dots & \binom{m}{s-3}\lambda^{m-s+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda^m \end{pmatrix}.$$

con el convenio de $\binom{m}{r} = 0$ si $m < r$.

PRUEBA: Podemos expresar B como la suma $B = \text{diag}(\lambda,\dots,\lambda) + N = D_\lambda + N$, donde

$$N_{s\times s} = \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Como D_λ conmuta con cualquier matriz cuadrada de orden s, y $N^{s-1} \neq 0$ y

$N^m = 0, m \geq s$, se tiene que

$$\begin{aligned} B^m &= (D_\lambda + N)^m \\ &= D_\lambda^m + \binom{m}{1} D_\lambda^{m-1} N + \binom{m}{2} D_\lambda^{m-2} N^2 + \cdots + \binom{m}{s-1} D_\lambda^{m-s+1} N^{s-1} \\ &= \lambda^m I_s + \binom{m}{1} \lambda^{m-1} N + \binom{m}{2} \lambda^{m-2} N^2 + \cdots + \binom{m}{s-1} \lambda^{m-s+1} N^{s-1}, \end{aligned}$$

de donde se sigue la expresión buscada. □

Por consiguiente, la expresión general de la potencia m-ésima de A es

$$A^m = P \begin{pmatrix} B_1^m & 0 & \dots & 0 \\ 0 & B_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & B_t^m \end{pmatrix} P^{-1},$$

donde P es la matriz de paso a la forma canónica de Jordan, y cada B_j^m es la potencia m-ésima de un bloque de Jordan.

Ejemplo 7.8.1. La matriz

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

tiene los autovalores $\lambda = i$ y $\overline{\lambda} = -i$. Su forma canónica compleja es

$$J = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \text{ con matriz de paso } P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}.$$

Entonces

$$\begin{aligned} A^m = PJ^mP^{-1} &= \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} i^m & 0 \\ 0 & (-i)^m \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} i^m + (-i)^m & i^{m+1} + (-i)^{m+1} \\ -i^{m+1} - (-i)^{m+1} & -i^{m+2} - (-i)^{m+2} \end{pmatrix}. \end{aligned}$$

7.9. Relaciones de recurrencia

Ecuaciones de diferencias

Dados $a_1,\dots,a_p \in \mathbb{R},\mathbb{C}$, con $a_p \neq 0$, llamamos ***ecuación lineal de diferencias finitas*** con coeficientes constantes de orden p a una relación de recurrencia del tipo

$$x_{n+p} - a_1 x_{n+p-1} - \cdots - a_p x_n = \varphi(n), \text{ para todo } n \geq 1,$$

donde $\varphi : \mathbb{N} \to \mathbb{R}$ es una función.
Si $\varphi(n) = 0$ para todo $n \in \mathbb{N}$, decimos que la ecuación de diferencias es *homogénea*.

Una solución de la ecuación de diferencias es una sucesión $\{x_n\}_{n\geq 1}$ que la satisfaga.

Vamos a calcular una expresión explícita de x_n en función de n para el caso homogéneo. Dada la ecuación de diferencias

$$x_{n+p} - a_1 x_{n+p-1} - \cdots - a_p x_n = 0, n \geq 1,$$

podemos escribir el siguiente sistema de ecuaciones lineales

$$\begin{array}{lcccccl} x_{n+p} & = & a_1 x_{n+p-1} & + \cdots + & a_{p-1} x_{n+1} + & a_p x_n, \\ x_{n+p-1} & = & x_{n+p-1}, & & & & \\ & \vdots & & & & & \\ x_{n+1} & = & & & & x_{n+1}, & \end{array}$$

cuya matriz de coeficientes es

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Esta matriz será la ***matriz asociada a la ecuación de diferencias***. Si escribimos $\boldsymbol{x}_n = (x_{n+p}, x_{n+p-1}, \dots, x_{n+1})^t$ entonces

$$\boldsymbol{x}_n = A\boldsymbol{x}_{n-1} = \cdots = A^n \boldsymbol{x}_0.$$

Es claro de lo anterior que el término general x_{n+p} de cualquier solución de la ecuación de diferencias es una *combinación lineal* de los elementos de la primera fila de A^n. Como sabemos calcular una expresión general de las potencias de una matriz cuadrada a partir de sus autovalores, podemos decir algo más.

Término general de la ecuación de diferencias

El término general de una ecuación de diferencias

$$x_{n+p} = a_1 x_{n+p-1} + \cdots + a_p x_n, \text{ para todo } n \geq 1,$$

es una combinación lineal de

$$\lambda^n, n\lambda^n, \ldots, n^{s-1}\lambda^n,$$

para cada autovalor λ de multiplicidad s de la matriz de la ecuación.

PRUEBA: Sea A la matriz de la ecuación, y $J = P^{-1}AP$ su forma canónica de Jordan. Entonces $A^n = PJ^nP^{-1}$, de donde los elementos de la primera fila de A son combinación lineal de los elementos de J^n. Sabemos que estos elementos son de la forma

$$\lambda^n, \binom{n}{1}\lambda^{n-1}, \binom{n}{2}\lambda^{n-2}, \ldots, \binom{n}{s-1}\lambda^{n-s+1},$$

para cada autovalor λ de A, con m su multiplicidad. Recordemos que los bloques de Jordan son a lo sumo de orden s.

Ahora, para cada $k = 1, 2, \ldots, s-1$,

$$\binom{n}{k}\lambda^{n-k} = \frac{\lambda^{-k}}{k!}(n(n-1)\cdots(n-k+1))\lambda^n = \frac{\lambda^{-k}}{k!}(n^k + b_{1k}n^{s-1} + \cdots + b_{k,k-1}n)\lambda^n,$$

para ciertos escalares $b_{1k}, \ldots, b_{k-1,k}$. Concluimos entonces que los elementos de PJ^nP^{-1} son combinaciones lineales de

$$\lambda^n, n\lambda^n, n^2\lambda^n, \ldots, n^{s-1}\lambda^n,$$

para cada autovalor λ de A, con multiplicidad s.

□

El caso en que la matriz de la ecuación en diferencias sea diagonalizable es particularmente sencillo. Si $\lambda_1, \ldots, \lambda_r$ son los autovalores distintos de A, entonces el término general de la ecuación es de la forma

$$x_{n+p} = c_1\lambda_1^n + c_2\lambda_2^n + \cdots + c_r\lambda_r^n,$$

donde $c_1, c_2, \ldots, c_r$ son escalares.

La determinación de las constantes que aparecen en las combinaciones lineales se hará a partir de las condiciones iniciales que nos proporcionen. Darán lugar a un sistema de ecuaciones, y sus soluciones serán los coeficientes buscados.

Ejemplo 7.9.1. Sucesión de Fibonacci. Este es el ejemplo clásico que se usa para ilustrar las ecuaciones en diferencias finitas homogéneas. Consideremos la sucesión definida por la relación de recurrencia

$$a_0 = 1, a_1 = 1, a_{n+1} = a_n + a_{n-1}, n \geq 2.$$

La ecuación característica es $z^2 = z + 1$, de raíces $r_1 = \frac{1}{2}(1+\sqrt{5}), r_2 = \frac{1}{2}(1-\sqrt{5})$. Entonces la forma general de la solución es

$$a_n = c_1 r_1^n + c_2 r_2^n,$$

y tenemos que calcular los valores de c_1 y c_2. Vienen dados por la condiciones iniciales, por lo que obtenemos el sistema de ecuaciones

$$\begin{cases} n=0, & a_0 = 1 = c_1 + c_2, \\ n=1, & a_1 = 1 = c_1 r_1 + c_2 r_2. \end{cases}$$

Las soluciones son

$$c_1 = \frac{1}{\sqrt{5}} r_1, c_2 = -\frac{1}{\sqrt{5}} r_2,$$

por lo que

$$\begin{aligned} a_n &= \frac{1}{\sqrt{5}} r_1 \cdot r_1^n - \frac{1}{\sqrt{5}} r_2 \cdot r_2^n = \frac{1}{\sqrt{5}} (r_1^{n+1} - r_2^{n+1}) \\ &= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right). \end{aligned}$$

Ejemplo 7.9.2. En el estudio de la teoría de colas, aparece el modelo

$$\lambda p_0 = \mu p_1, (\lambda + \mu) p_n = \lambda p_{n-1} + \mu p_{n+1}, n \geq 1, \lambda < \mu,$$

y los p_i indican una distribución de probabilidad. La ecuación la podemos escribir como

$$p_{n+1} = \frac{\lambda + \mu}{\mu} p_n - \frac{\lambda}{\mu} p_{n-1}, n \geq 1,$$

y la ecuación característica es

$$z^2 - \frac{\lambda + \mu}{\mu} z + \frac{\lambda}{\mu} = 0,$$

de soluciones $\rho = \frac{\lambda}{\mu}$ y 1. Entonces la solución general tiene la forma

$$p_n = c_1\rho^n + c_2, n \geq 1.$$

Como $\sum_{n\geq 0} p_n = 1$, se deduce que $c_2 = 0$, y que $p_0 + c_1 \sum_{n\geq 1} \rho^n = 1$. Entonces

$$p_0 + c_1 \frac{\rho}{1-\rho} = 1, c_1 = \frac{1-\rho}{\rho}(1-p_0).$$

Por tanto, $p_n = \frac{1-\rho}{\rho}\rho^n(1-p_0)$, y la otra condición es $p_1 = \rho p_0$. Concluimos que

$$p_1 = \frac{1-\rho}{\rho}\rho(1-p_0) = \rho p_0,$$

de donde $p_0 = 1-\rho$. Finalmente,

$$p_n = \frac{1-\rho}{\rho}\rho^n\rho = (1-\rho)\rho^n.$$

Ejemplo 7.9.3. Veamos ahora un problema de probabilidad asociado a una cadena de Markov. Supongamos que en un juego, el jugador A tiene y monedas y el jugador B x monedas. El jugador A gana una moneda del jugador B si acierta el resultado del lanzamiento de una moneda, y la pierde en caso contrario. ¿Cuál es la probabilidad de que el jugador A gane las x monedas de B antes de que el jugador B gane las y monedas de A?
La serie de lanzamientos la gana el jugador A si gana x monedas más que B antes de que B gane y monedas más que A, y es ganada por B en caso contrario. Sea p_n la probabilidad de que A gane la serie en el estado en que ha ganado n juegos más que B, donde $-y \leq n \leq x$. Esta definición es la adecuada porque las condiciones para ganar el juego tienen que ver únicamente con la diferencia entre los juegos ganados por A y B y no con los totales ganados. Vamos a establecer una ecuación en diferencias para p_n. Sea p la probabilidad de que A gane un juego y q la probabilidad de que lo haga B ($p+q=1$). Sea S_n el estado en el que el jugador A tiene n monedas y $p_n = P(S_n)$ la probabilidad de ganar cuando se encuentra en el estado S_n. Entonces

$$P(S_n) = P(S_n|W)P(W) + P(S_n|\bar{W})P(\bar{W}),$$

donde W es el suceso en el que el jugador A gana. Entonces $P(W) = p, P(\bar{W}) = q$ y, además,

$$\begin{aligned} P(S_n|W) &= \text{ probabilidad de que A gane con } n+1 \text{ monedas,} \\ P(S_n|\bar{W}) &= \text{ probabilidad de que A gane con } n-1 \text{ monedas.} \end{aligned}$$

Es decir, $P(S_n|W) = p_{n+1}$ y $P(S_n|\bar{W}) = p_{n-1}$, y obtenemos la ecuación en diferencias

$$p_n = pp_{n+1} + qp_{n-1}.$$

Para resolverla,

$$0 = pp_{n+1} - p_n + qp_{n-1}, p_{n+1} = \frac{1}{p}p_n + \frac{q}{p}p_{n-1},$$

con condiciones iniciales $p_x = 1, p_{-y} = 0$. La ecuación característica es $0 = pz^2 - z + q$. Sus raíces son $z = 1, z = q/p$. Si $p \neq q$, tenemos dos raíces distintas, y entonces

$$p_n = \alpha + \beta(q/p)^n$$

para ciertos α, β. Si $p = q$, hay una raíz doble, y en este caso

$$p_n = \alpha + \beta n.$$

Apliquemos las condiciones iniciales $p_x = 1, p_{-y} = 0$. En el caso $p \neq q$ obtenemos

$$1 = \alpha + \beta(q/p)^x, 0 = \alpha + \beta(q/p)^{-y}$$

lo que lleva a

$$\alpha = \frac{1}{1-(q/p)^{x+y}}, \beta = -\frac{(q/p)^y}{(1-(q/p))^{x+y}}$$

y

$$p_n = \frac{1-(q/p)^{n+y}}{1-(q/p)^{x+y}}, \text{ si } p \neq q.$$

En particular,

$$p_0 = \frac{1-r^y}{1-r^{x+y}}, \text{ donde } r = \frac{q}{p}.$$

Si $p = q$ las ecuaciones para α y β son

$$1 = \alpha + \beta x, 0 = \alpha - \beta y$$

por lo que

$$\alpha = \frac{y}{x+y}, \beta = \frac{1}{x+y}$$

y

$$p_n = \frac{y+n}{x+y}, \text{ si } p = q.$$

En este último caso, $p_0 = \frac{y}{y+x}$.

7.10. * Análisis de componentes principales

Un ejemplo de una matriz de datos con dos entradas es el conjunto de pesos y alturas de N estudiantes de un colegio. Sea X_j el vector de observación en $\mathbb{R}^2$ que contiene el peso y la altura del estudiante j-ésimo. Si p es el peso y h la altura, entonces la matriz de observaciones tiene la forma

$$\begin{array}{cccc} p_1 & p_2 & \dots & p_N \\ h_1 & h_2 & \dots & h_N \\ \uparrow & \uparrow & \dots & \uparrow \\ X_1 & X_2 & \dots & X_N. \end{array}$$

Otro ejemplo consiste en las fotos tomadas por un satélite con diferentes cámaras (infrarroja, térmica, color), que la podemos considerar como una imagen con varias componentes, que son fotos a diferentes longitudes de onda. Cada fotografía proporciona información sobre una misma zona. Consideremos tres fotos sobre una región determinada, cada una de ellas de 2000×2000 píxeles. Así, hay cuatro millones de píxeles en cada imagen, y los datos los podemos ver almacenados en una matriz de tres filas y 4 millones de columnas. Si los visualizamos como puntos en $\mathbb{R}^3$, nos queda una nube con cierta correlación.

Media y covarianza. Para preparar el análisis de componentes principales, sea $\begin{bmatrix} X_1 & \dots X_N \end{bmatrix}$ una matriz $p \times N$ de observaciones. La media muestral $\boldsymbol{\mu}$ de las observaciones de los vectores $X_1, \dots, X_N$ está dada por

$$\boldsymbol{\mu} = \frac{1}{N}(X_1 + \cdots + X_N).$$

Para $k = 1, \dots, N$ sea

$$\hat{X}_k = X_k - \boldsymbol{\mu},$$

y formamos la matriz

$$B = \begin{bmatrix} \hat{X}_1 & \hat{X}_2 & \dots & \hat{X}_N \end{bmatrix}.$$

Las columnas de B, es decir, las nuevas variables $\hat{X}_k$, tienen media igual a cero (variables centradas). La matriz de covarianza muestral es la matriz de orden $p \times p$ definida por

$$S = \frac{1}{N-1} BB^t.$$

Sabemos que S es una matriz simétrica semi-definida positiva. Por ejemplo, consideremos los vectores de observación

$$X_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, X_2 = \begin{pmatrix} 4 \\ 2 \\ 13 \end{pmatrix}, X_3 = \begin{pmatrix} 7 \\ 8 \\ 1 \end{pmatrix}, X_4 = \begin{pmatrix} 8 \\ 4 \\ 5 \end{pmatrix}.$$

El vector de media es

$$\boldsymbol{\mu} = \frac{1}{4}\left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ 2 \\ 13 \end{pmatrix} + \begin{pmatrix} 7 \\ 8 \\ 1 \end{pmatrix} + \begin{pmatrix} 8 \\ 4 \\ 5 \end{pmatrix}\right) = \begin{pmatrix} 5 \\ 4 \\ 5 \end{pmatrix}.$$

Las variables centradas son

$$\hat{X}_1 = \begin{pmatrix} -4 \\ -2 \\ -4 \end{pmatrix}, \hat{X}_2 = \begin{pmatrix} -1 \\ -2 \\ 8 \end{pmatrix}, \hat{X}_3 = \begin{pmatrix} 2 \\ 4 \\ -4 \end{pmatrix}, \hat{X}_4 = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix},$$

y

$$B = \begin{pmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{pmatrix},$$

de donde

$$S = \frac{1}{3}\begin{pmatrix} -4 & -1 & 2 & 3 \\ -2 & -2 & 4 & 0 \\ -4 & 8 & -4 & 0 \end{pmatrix}\begin{pmatrix} -4 & -2 & -4 \\ -1 & -2 & 8 \\ 2 & 4 & -4 \\ 3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{pmatrix}.$$

Sea $S = (s_{ij})$, y representemos por $x_1, \ldots, x_p$ las componentes de los vectores X. Entonces x_1 es un escalar que varía en las primeras componentes de los vectores de observación $X_1, \ldots, X_N$. Para $j = 1, \ldots, p$, la entrada s_{jj} de S es la varianza de x_j. La varianza de x_j es una medida de la dispersión de los valores de x_j. La varianza total es la suma de las varianzas de la diagonal de S, que es la traza de S.

La entrada s_{ij}, con $i \neq j$ es la covarianza entre x_i y x_j. Si es igual a cero, decimos que las variables no están correlacionadas. el análisis multivariante de los datos se simplifica cuando la mayoría de las variables $x_1, \ldots, x_p$ no están correlacionadas, esto es, cuando la matriz de covarianza de $X_1, \ldots, X_n$ es diagonal, o aproximadamente diagonal.

Análisis de componentes principales. Por simplicidad, supongamos que la matriz $\begin{pmatrix} X_1 & X_2 & \ldots & X_N \end{pmatrix}$ tiene las variables centradas (media cero). El objetivo del análisis de componentes principales es calcular una matriz ortogonal $P = \begin{pmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_p \end{pmatrix}$ de orden $p \times p$ que nos permita hacer un cambio de variable $X = PU$, o

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_p \end{pmatrix}\begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_p \end{pmatrix},$$

con la propiedad de que las nuevas variables no estén correlacionadas y se encuentren ordenadas en orden decreciente de covarianza.

El cambio ortogonal de variables $X = PU$ significa que cada vector de observación X_k recibe un nuevo nombre U_k, tal que $X_k = PU_k$. Como P es ortogonal, se tiene que $U_k = P^{-1}X_k = P^tX_k$, $k = 1,2,\ldots,N$.

Sabemos que

$$\text{cov}(U) = \text{cov}(P^tX) = P^t\,\text{cov}(X)P,$$

por lo que la matriz ortogonal P que buscamos es la que hace P^tSP diagonal. Como S es simétrica semi-definida positiva, sabemos cómo calcular esta matriz. Sea D la matriz diagonal con los autovalores $\lambda_1,\ldots,\lambda_p$ de S en la diagonal, de forma que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Sea P una matriz cuyas columnas son una base ortonormal de autovectores correspondientes a los autovalores. Entonces $S = PDP^t, D = P^tSP$.

Los autovectores ortonormales $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_p$ de S, que forman la matriz P, se denominan componentes principales de los datos. La primera componente principal es el autovector correspondiente al mayor autovalor de S, la segunda componente principal es el autovector procedente del segundo autovalor, y así sucesivamente.

La primera componente principal $\boldsymbol{v}_1$ determina la nueva variable u_1 de la siguiente forma. Sea

$$\boldsymbol{v}_1 = \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix}.$$

Como $\boldsymbol{v}_1^t$ es la primera fila de P^t, la ecuación $U = P^tX$ nos da que

$$u_1 = \boldsymbol{v}_1^tX = c_1x_1 + c_2x_2 + \cdots + c_px_p.$$

Así, u_1 es combinación lineal de las variables originales $x_1,\ldots,x_p$, con las componentes del autovector $\boldsymbol{v}_1$ como pesos. De manera similar se determina u_2 en función de $\boldsymbol{v}_2$.

Por ejemplo, si la matriz de covarianza es

$$S = \begin{pmatrix} 2382{,}78 & 2611{,}84 & 2136{,}20 \\ 2611{,}84 & 3106{,}47 & 2553{,}90 \\ 2136{,}20 & 2553{,}90 & 2650{,}71 \end{pmatrix},$$

los autovalores son $\lambda_1 = 7614{,}23$, $\lambda_2 = 427{,}63$, $\lambda_3 = 98{,}10$, y los autovectores normalizados son

$$\boldsymbol{v}_1 = \begin{pmatrix} {,}5417 \\ {,}6295 \\ {,}5570 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} -{,}4894 \\ -{,}3026 \\ {,}8179 \end{pmatrix}, \boldsymbol{v}_3 \begin{pmatrix} {,}6834 \\ -{,}7157 \\ {,}1441 \end{pmatrix}.$$

Si nos quedamos con dos decimales, por simplicidad, la variable para la primera componente principal es

$$u_1 = ,54x_1 + ,63x_2 + ,56x_3.$$

La nueva matriz de covarianza, con las variables u_1, u_2, u_3, es

$$D = \begin{pmatrix} 7614{,}23 & & \\ & 427{,}63 & \\ & & 98{,}10 \end{pmatrix}.$$

La varianza total de S y la de la matriz D coinciden, pues la traza no se altera por semejanza de matrices. Por tanto,

$$\begin{aligned} \text{varianza total de } S &= \text{traza}(S) = \text{traza}(D) \\ &= \text{varianza total de } D = 7614{,}23 + 427{,}63 + 98{,}10 = 8139{,}96. \end{aligned}$$

Reducción de la dimensionalidad. El análisis de componentes principales es adecuado para aplicaciones en las que la mayor parte de la variación en los datos se debe a variaciones de unas pocas de las nuevas variables $u_1, \ldots, u_p$. El cociente λ_j / traza(S) mide la fracción de la varianza total explicada o capturada pro u_j.

Así, en el ejemplo anterior,

- Primera componente: $\frac{7614{,}23}{8139{,}96} = 93{,}5\,\%$.
- Segunda componente: $\frac{427{,}63}{8139{,}96} = 5{,}3\,\%$.
- Tercera componente: $\frac{98{,}10}{8139{,}96} = 1{,}2\,\%$.

Los datos apenas tienen varianza en la tercera componente, y los valores de u_3 son prácticamente cero. Algo parecido ocurre con u_2, y los datos aparecen próximos a una recta determinada por u_1.

Caracterización de las variables de componentes principales. Si $u_1, \ldots, u_p$ proceden de un análisis de componentes principales de una matriz $p \times N$ de observaciones, entonces la varianza de u_1 es tan grande como sea posible en el siguiente sentido. Si $\boldsymbol{v}$ es un vector unitario y $\boldsymbol{u} = \boldsymbol{v}^t X$, entonces la varianza de los valores de $\boldsymbol{u}$ cuando X varía sobre los datos originales $X_1, \ldots, X_N$ es $\boldsymbol{v}^t S \boldsymbol{v}$. Se puede probar que

$$\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}^t S \boldsymbol{v} = \lambda_1, \text{ el mayor autovalor de } S,$$

y esta varianza se alcanza cuando $\boldsymbol{v}$ es el autovector correspondiente. De la misma forma, u_2 tiene la máxima varianza entre las variables $\boldsymbol{u} = \boldsymbol{v}^t X$ que no están correlacionadas con u_1. Los mismo para las restantes variables $u_3, \ldots, u_p$.

Nota numérica: la descomposición en valores singulares es la mejor herramienta para realizar el análisis de componentes principales en la práctica. Si B es una matriz $p \times N$ de observaciones centradas, y $A = \frac{1}{\sqrt{N-1}} B^t$, entonces $A^t A$ es la matriz de covarianza S. Los cuadrados de los valores singulares de A son los p autovalores de S, y los vectores de la matriz V son las componentes principales de los datos.

El cálculo iterado de la descomposición en valores singulares de A es más rápido y preciso que una descomposición de autovalores de S. Esto es particularmente cierto cuando p es grande.

Capítulo 8

Número de condición de un sistema

8.1. Normas matriciales

Como $\mathbb{C}^{m\times n}$ es un espacio vectorial de dimensión mn, el tamaño de una matriz $A \in \mathbb{C}^{m\times n}$ se puede medir mediante cualquier norma vectorial de $\mathbb{C}^{mn}$. Así, podríamos definir $\|A\| = \|\text{vec}(A)\|$, donde vec$(A)$ es el vector con $m \times n$ componentes que se obtiene al apilar las columnas de A. Por ejemplo, si consideramos la norma euclídea en $\mathbb{R}^4$, la norma de la matriz

$$A = \begin{pmatrix} 2 & -1 \\ -4 & -2 \end{pmatrix}$$

sería $\|A\| = \sqrt{2^2 + 1^2 + 4^2 + 2^2} = 5$. Esta es la noción más simple de norma matricial, y es la que llamaremos norma de Frobenius.

Norma de Frobenius

La **norma de Frobenius** de $A \in \mathbb{C}^{m\times n}$ es

$$\|A\|_F^2 = \sum_{i,j} |a_{ij}|^2 = \sum_i \|A_{i*}\|_2^2 = \sum_j \|A_{*j}\|_2^2 = \text{traza}(A^* A).$$

Por la propia definición, $\|A\|_F^2 = \|A^*\|_F^2$, de donde $\|A\|_F^2 = \text{traza}(AA^*)$. La norma de Frobenius es buena para algunas aplicaciones, pero no para todas. Así, de manera similar a las normas vectoriales, exploraremos otras alternativas. Pero antes de ello tenemos que dar una definición general de norma matricial. El objetivo es comenzar con las propiedades que definen una norma vectorial y preguntarse qué se debe añadir a la lista.

La multiplicación matricial diferencia el espacio de las matrices de otros espacios vectoriales, donde no tiene que haber un producto definido. Por ello, necesitamos una propiedad que relacione $\|AB\|$ con $\|A\|\,\|B\|$. La norma de Frobenius sugiere cómo debe ser esta relación.

La desigualdad CBS nos dice que

$$\|A\boldsymbol{x}\|_2^2 = \sum_i |A_{i*}\boldsymbol{x}|^2 \leq \sum_i \|A_{i*}\|_2^2 \|\boldsymbol{x}\|_2^2 = \|A\|_F^2 \|\boldsymbol{x}\|_2^2,$$

esto es, que

$$\|A\boldsymbol{x}\|_2 \leq \|A\|_F \|\boldsymbol{x}\|_2.$$

Entonces, si A y B son matrices que se pueden multiplicar

$$\begin{aligned}\|AB\|_F^2 &= \sum_j \left\|[AB]_{*j}\right\|_2^2 = \sum_j \left\|AB_{*j}\right\|_2^2 \leq \sum_j \|A\|_F^2 \left\|B_{*j}\right\|_2^2 \\ &= \|A\|_F^2 \sum_j \left\|B_{*j}\right\|_2^2 = \|A\|_F^2 \|B\|_F^2, \text{ luego } \|AB\|_F \leq \|A\|_F \|B\|_F.\end{aligned}$$

Esto sugiere que la propiedad $\|AB\| \leq \|A\|\,\|B\|$ debe añadirse a las propiedades de norma vectorial para definir una norma matricial.

Norma matricial

Una **norma matricial** es una aplicación $\|*\|$ del conjunto de matrices complejas en $\mathbb{R}$ que satisface las siguientes propiedades:

- $\|A\| \geq 0$ y $\|A\| = 0 \Leftrightarrow A = 0$.
- $\|\alpha A\| = |\alpha|\,\|A\|$ para todo escalar α.
- $\|A+B\| \leq \|A\| + \|B\|$ para todas las matrices del mismo orden.
- $\|AB\| \leq \|A\|\,\|B\|$ para las matrices ajustadas.

La norma de Frobenius satisface las propiedades anteriores, y además es invariante por transformaciones unitarias, es decir, $\|A\|_F = \|UA\|_F = \|AV\|_F$ para U y V matrices unitarias. En efecto,

$$\begin{aligned}\|UA\|_F^2 &= \text{traza}((UA)^*(UA)) = \text{traza}(A^*U^*UA) = \text{traza}(A^*A) = \|A\|_F^2, \\ \|AV\|_F^2 &= \text{traza}((AV)(AV)^*) = \text{traza}(AVV^*A^*) = \text{traza}(AA^*) = \|A\|_F^2.\end{aligned}$$

Norma matricial inducida

Una norma vectorial que está definida en $\mathbb{C}^m$ y $\mathbb{C}^n$ ***induce*** una norma matricial en $\mathbb{C}^{m\times n}$ mediante

$$\|A\| = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| \text{ con } A \in \mathbb{C}^{m\times n}, \boldsymbol{x} \in \mathbb{C}^{n\times 1}.$$

En esta situación, $\|A\boldsymbol{x}\| \leq \|A\|\,\|\boldsymbol{x}\|$.

PRUEBA: La definición de esta norma tiene sentido, porque una función continua sobre un compacto alcanza el máximo. Las tres primeras condiciones de norma se verifican fácilmente. Veamos ahora que $\|A\boldsymbol{x}\| \leq \|A\|\,\|\boldsymbol{x}\|$. Para $\boldsymbol{x} = \mathbf{0}$ es trivial. Sea ahora $\boldsymbol{x}_0 \neq \mathbf{0}$ y $\boldsymbol{z} = \frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|}$. Entonces

$$\|A\| = \max_{\|\boldsymbol{v}\|=1} \|A\boldsymbol{v}\| \geq \|A\boldsymbol{z}\| = \frac{\|A\boldsymbol{x}_0\|}{\|\boldsymbol{x}_0\|} \Rightarrow \|A\boldsymbol{x}_0\| \leq \|A\|\,\|\boldsymbol{x}_0\|.$$

La propiedad multiplicativa se deduce entonces:

$$\|AB\boldsymbol{x}\| \leq \|A\|\,\|B\|\,\|\boldsymbol{x}\|.$$

□

A estas normas también se las llama *subordinadas*. Vamos a estudiar las normas matriciales inducidas por las normas vectoriales que conocemos, esto es, las normas $\|*\|_1, \|*\|_2$ y $\|*\|_\infty$. Queremos obtener expresiones de estas normas que puedan ser calculadas a partir de la matriz, y no como máximo de una función. Comencemos con las más fáciles.

Normas matriciales $\|*\|_1$ y $\|*\|_\infty$

Las normas matriciales inducidas por las normas vectoriales $\|*\|_1$ y $\|*\|_\infty$ verifican:

- $\|A\|_1 = \max_{\|\boldsymbol{x}\|_1=1} \|A\boldsymbol{x}\|_1 = \max_j \sum_i |a_{ij}|$ = la mayor de las 1-normas de las columnas.
- $\|A\|_\infty = \max_{\|\boldsymbol{x}\|_\infty=1} \|A\boldsymbol{x}\|_\infty = \max_i \sum_j |a_{ij}|$ = la mayor de las 1-normas de las filas.

PRUEBA: Si $\boldsymbol{x}$ es un vector con $\|\boldsymbol{x}\|_1 = 1$, entonces

$$\begin{aligned}\|A\boldsymbol{x}\|_1 &= \sum_{i=1}^{m} |A_{i*}\boldsymbol{x}| = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij}x_j \right| \leq \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}||x_j| \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} |a_{ij}||x_j| = \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |a_{ij}| \leq \|\boldsymbol{x}\|_1 \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}| \\ &= \max_{1\leq j\leq n} \|A_{*j}\|_1\end{aligned}$$

Sea k el índice donde se alcanza el máximo y tomemos $\boldsymbol{x} = \boldsymbol{e}_k$. Entonces $\|\boldsymbol{e}_k\|_1 = 1$ y

$$\|A\boldsymbol{e}_k\|_1 = \|A_{*k}\|_1 .$$

Para la norma $\|*\|_\infty$, sea $\boldsymbol{x}$ vector con $\|\boldsymbol{x}\|_\infty = 1$. Entonces

$$\|A\boldsymbol{x}\|_\infty = \max_{1\leq i\leq m} \left| \sum_{j=1}^{n} a_{ij}x_j \right| \leq \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}||x_j| \leq \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}| .$$

Sea k el índice de la fila A_{k*} donde se alcanza el máximo, y definimos el vector $\boldsymbol{u}$ como

$$u_j = \begin{cases} 1 & \text{si } a_{kj} = 0 \\ \frac{1}{|a_{kj}|}\overline{a_{kj}} & \text{si } a_{kj} \neq 0 \end{cases}$$

Entonces $\|\boldsymbol{u}\|_\infty = 1$ y

$$|(A\boldsymbol{u})_i| \leq \sum_{j=1}^{n} |a_{ij}||u_j| = \sum_{j=1}^{n} |a_{ij}| \leq \sum_{j=1}^{n} |a_{kj}|$$

$$|(A\boldsymbol{u})_k| = \left| \sum_{j=1}^{n} a_{kj}u_j \right| = \left| \sum_{a_{kj}\neq 0} a_{kj}u_j \right| = \left| a_{kj}\frac{1}{|a_{kj}|}\overline{a_{kj}} \right| = \sum |a_{kj}|$$

de donde

$$\|A\boldsymbol{u}\|_\infty = \sum |a_{kj}| = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|\, \|\boldsymbol{u}\|_\infty .$$

□

La norma $\|*\|_2$ es algo más difícil de tratar. Necesitamos varios conceptos. En primer lugar, el *radio espectral* de una matriz cuadrada A es el número

$$\rho(A) = \max\{|\lambda| \mid \lambda \text{ autovalor de } A\}.$$

Es de gran importancia para el estudio de sucesiones de matrices.

Norma matricial $\|*\|_2$

La norma matricial inducida por la norma vectorial $\|*\|_2$ verifica que

$$\|A\|_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)} = \left\|A^*\right\|_2.$$

Cuando A es no singular,

$$\left\|A^{-1}\right\|_2 = \frac{1}{\sqrt{\lambda_{\text{mín}}}},$$

donde $\lambda_{\min}$ es el menor autovalor de A^*A.

PRUEBA: Como A^*A es hermitiana, es diagonalizable por una matriz unitaria: $D = U^*A^*AU = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Entonces $A^*A = UDU^*$ y

$$\|Av\|_2^2 = v^*A^*Av = v^*UDU^*v = (U^*v)^*D(U^*v) = \sum_{i=1}^{n} \lambda_i |w_i|^2$$

donde $w = U^*v$. De aquí

$$\|Av\|_2^2 \leq \rho(A^*A) \sum_{i=1}^{n} |w_i|^2 = \rho(A^*A) \left\|U^*v\right\|_2^2 = \rho(A^*A) \|v\|_2^2$$

y entonces $\|A\|_2 \leq \rho(A^*A)^{1/2}$. Sea λ el autovalor máximo de A^*A, que es no negativo, pues A^*A es semi-definida positiva, y u un autovector asociado. Entonces

$$\|Au\|_2^2 = u^*A^*Au = \lambda \|u\|_2^2$$

y tenemos la igualdad.

Como los autovalores no nulos de A^*A y AA^* coinciden, también lo hacen sus radios espectrales.

Si A es no singular, entonces AA^* es no singular, y

$$\left\|A^{-1}\right\|_2 = \sqrt{\rho((A^{-1})^*A^{-1})} = \sqrt{\rho((A^*)^{-1}A^{-1})} = \sqrt{\rho((AA^*)^{-1})}.$$

Pero los autovalores de $(AA^*)^{-1}$ son los inversos de los autovalores de AA^*. Por tanto,

$$\left\|A^{-1}\right\|_2 = \frac{1}{\sqrt{\lambda_{\text{mín}}}}.$$

□

Propiedades de $\|*\|_2$

- $\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \max_{\|\boldsymbol{y}\|=1} |\boldsymbol{y}^* A\boldsymbol{x}|$.
- $\|*\|_2$ es invariante por transformaciones unitarias, esto es, $\|U^* AV\|_2 = \|A\|_2$ si $U^* U = I$ y $V^* V = I$.
- Si B es hermitiana entonces $\|B\|_2 = \rho(B)$.
- Si U es unitaria, entonces $\|U\|_2 = 1$.
- $\|A^* A\|_2 = \|A\|_2^2$.
- $\left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right\|_2 = \max\{\|A\|_2, \|B\|_2\}$.

PRUEBA: Sean $\boldsymbol{x}, \boldsymbol{y}$ vectores unitarios respecto de la norma $\|\cdot\|_2$. Entonces

$$|\boldsymbol{y}^* A\boldsymbol{x}| = |A\boldsymbol{x} \bullet \boldsymbol{y}| \leq \|A\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \leq \|A\|_2 .$$

Vamos a ver que tal valor se alcanza. Sea $\boldsymbol{x}_0$ un vector unitario tal que $\|A\boldsymbol{x}_0\|_2 = \|A\|_2$ (autovector unitario asociado a $\lambda_{\text{máx}}$ de $A^* A$), y consideremos $\boldsymbol{y}_0 = \frac{A\boldsymbol{x}_0}{\|A\|_2}$, que es unitario. Entonces

$$\boldsymbol{y}_0^* A\boldsymbol{x}_0 = \frac{\boldsymbol{x}_0^* A^* A\boldsymbol{x}_0}{\|A\|_2} = \frac{\|A\boldsymbol{x}_0\|_2^2}{\|A\|_2} = \frac{\|A\|_2^2}{\|A\|_2} = \|A\|_2 ,$$

por lo que el máximo se alcanza.

Sea ahora U unitaria. Entonces

$$\|A\|_2^2 = \rho(A^* A) = \rho(A^* U^* UA) = \|UA\|_2^2 .$$

Análogamente, para V unitaria del tamaño adecuado, $\|A\|_2 = \|AV\|_2$.

Si B es hermitiana, existe U unitaria y D diagonal tal que $D = U^* BU$, y los autovalores de B forman la diagonal de la matriz D. Entonces

$$\|B\|_2^2 = \|D\|_2^2 = \rho(D^* D) = |\lambda_{\text{máx}}|^2 = \rho(B)^2 .$$

Si aplicamos este resultado a $B = A^* A$, que es hermitiana, tenemos que $\|A^* A\|_2 = \rho(A^* A)^2 = \|A\|_2^2$.

Por último, si

$$C = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

el mayor autovalor de C^*C es el máximo entre el mayor autovalor de A^*A y B^*B.

□

Valores singulares y $\|*\|_2$

Si $A_{n\times n}$ es una matriz no singular con valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, entonces

- $\sigma_1 = \|A\|_2$.
- $\sigma_n = \frac{1}{\|A^{-1}\|_2}$.

PRUEBA: Si $A = U\Sigma V^*$ es su descomposición en valores singulares, entonces $A^{-1} = V\Sigma^{-1}U^*$, y

$$\|A\|_2 = \left\|U\Sigma V^*\right\|_2 = \|\Sigma\|_2 = \sigma_1,$$
$$\left\|A^{-1}\right\|_2 = \left\|V\Sigma^{-1}U^*\right\|_2 = \left\|\Sigma^{-1}\right\|_2 = \frac{1}{\sigma_n}.$$

□

Nota 8.1.1. Existen acotaciones entre las diferentes normas de matrices. Por ejemplo,

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2\,.$$

La primera desigualdad se obtiene de $\|A\boldsymbol{x}\|_2 \leq \|A\|_F\,\|\boldsymbol{x}\|_2$. Para la segunda, recordemos que $\left\|A_{*j}\right\|_2^2 = \left\|Ae_j\right\|_2^2 \leq \|A\|_2^2$. Entonces, si j_0 es la columna donde se alcanza el máximo de máx$\{\left\|A_{*j}\right\|_2^2\}$, tenemos que

$$\|A\|_F^2 = \sum_{i,j}\left|a_{ij}\right|^2 = \sum_j \left\|A_{*j}\right\|_2^2 \leq n\left\|A_{*j_0}\right\|_2^2 \leq n\,\|A\|_2^2\,.$$

Otras desigualdades son

1. máx$\left|a_{ij}\right| \leq \|A\|_2 \leq \sqrt{mn}\,$máx$\left|a_{ij}\right|$.
2. $\frac{1}{\sqrt{n}}\,\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\,\|A\|_\infty$.
3. $\frac{1}{\sqrt{m}}\,\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_1$.

En cualquier caso, estas relaciones indican que no es necesario obtener un valor exacto de la norma de una matriz, sino que nos puede valer una estimación.

Nota 8.1.2. Para toda norma matricial se verifica que $\rho(A) \leq \|A\|$.

En efecto, sea $\boldsymbol{v}$ autovector asociado al autovalor λ de A de módulo máximo, y $\boldsymbol{w} \in V$ vector tal que la matriz cuadrada $\boldsymbol{v}\boldsymbol{w}^t$ sea no nula. Entonces

$$\rho(A)\left\|\boldsymbol{v}\boldsymbol{w}^t\right\| = |\lambda|\left\|\boldsymbol{v}\boldsymbol{w}^t\right\| = \left\|\lambda\boldsymbol{v}\boldsymbol{w}^t\right\| = \left\|A\boldsymbol{v}\boldsymbol{w}^t\right\| \leq \|A\|\left\|\boldsymbol{v}\boldsymbol{w}^t\right\|,$$

de donde se sigue el resultado por ser $\left\|\boldsymbol{v}\boldsymbol{w}^t\right\| > 0$.

8.2. Aproximaciones de matrices

Existen resultados en los que se aproximan matrices por otras, y esa medida la realizamos con normas matriciales.

Mejor aproximación de rango k

Sea $A = U\Sigma V^*$ descomposición en valores singulares de A. Para cada $0 \leq k \leq r = \text{rango}(A)$ definimos

$$A_k = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^*.$$

Si $k = p = \text{mín}\{m, n\}$ definimos $\sigma_{k+1} = 0$. Entonces

$$\min_{\text{rango}(B)\leq k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}.$$

PRUEBA: Tenemos que $A - A_k = \sum_{i=k+1}^{r} \sigma_i \mathbf{u}_i \boldsymbol{v}_i^*$, por lo que $A - A_k = U_1 \Sigma_1 V_1^*$ donde U_1, V_1 son unitarias y $\Sigma_1 = \text{diag}(\sigma_{k+1}, \ldots, \sigma_r, 0, \ldots, 0)$. Entonces $\|A - A_k\|_2 = \sigma_{k+1}$. Supongamos que B es una matriz con rango$((B) \leq k$ y $\|A - B\|_2 < \|A - A_k\|_2 = \sigma_{k+1}$. Entonces $W = \text{null}(B)$ es un subespacio de dimensión, al menos, $n - k$. Si $\boldsymbol{w} \in W$, entonces

$$\|A\boldsymbol{w}\|_2 = \|(A - B)\boldsymbol{w}\|_2 \leq \|A - B\|_2 \|\boldsymbol{w}\|_2 < \sigma_{k+1} \|\boldsymbol{w}\|_2$$

Sea $L = \langle \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k+1} \rangle$. Es un subespacio vectorial de dimensión $k+1$, y $A\boldsymbol{v}_i = \sigma_i \mathbf{u}_i$. Si $\boldsymbol{w} \in L$, con $\boldsymbol{w} = \sum_{j=1}^{k+1} \alpha_j \boldsymbol{v}_j$, entonces

$$\|A\boldsymbol{w}\|_2 = \left\| \sum_{j=1}^{k+1} \alpha_j \sigma_j \boldsymbol{v}_j \right\|_2 = \left(\sum_{j=1}^{k+1} \alpha_j^2 \sigma_j^2 \right)^{1/2} \geq \sigma_{k+1} \|\boldsymbol{w}\|_2$$

Como la suma de las dimensiones de W y L es mayor que n, tiene que haber un vector común, lo que es contradictorio. □

Un subconjunto Y de un espacio métrico X es denso si cada entorno de un punto de X contiene un punto de Y. Esto es equivalente a decir que todo punto de X es límite de una sucesión de puntos de Y.

El espacio $\mathscr{M}(n \times n)$ de las matrices $n \times n$ con coeficientes complejos es un espacio métrico si definimos la distancia entre matrices como $d(A,B) = \|A-B\|_2$. Vamos a ver que ciertos subconjuntos de este espacio son densos. El argumento de cada caso tendrá un ingrediente común. La propiedad que caracteriza al subconjunto Y será una que no cambia por semejanza unitaria. Así, si $A = UTU^*$ y probamos la existencia de un elemento de Y en un entorno de radio ε de una matriz triangular superior T, entonces habremos probado la existencia de un elemento de Y en un entorno de A.

- Las matrices no singulares son densas. Una matriz es no singular si y solamente si no tiene el autovalor cero. Esta propiedad no es afectada por una transformación de semejanza. Queremos probar que si A tiene el autovalor cero, entonces, para cada $\varepsilon > 0$ existe una matriz no singular B tal que $\|A-B\|_2 < \varepsilon$. Sea $A = UTU^*$, con T triangular superior. Si A es singular, algunas de las entradas de la diagonal de T es nula. Las cambiamos por números positivos no nulos pequeños, de forma que la nueva matriz T' verifique que $\|T-T'\|_2 < \varepsilon$. Entonces T' es no singular, y también lo es $A' = UT'U^*$. Además,

$$\|A-A'\|_2 = \|U(T-T')U^*\|_2 = \|T-T'\|_2 < \varepsilon.$$

- Las matrices con todos los autovalores distintos son densas. Usamos el mismo argumento que en el caso anterior. Si dos entradas de la diagonal de T son iguales, cambiamos una ligeramente.

- Las matrices diagonalizables son densas. Sabemos que toda matriz con sus autovalores distintos es diagonalizable (pero no al revés). Por ello, el conjunto de matrices diagonalizable contiene un conjunto que ya es denso, y por tanto es denso también.

Nota 8.2.1. El teorema de aproximación nos dice que si aproximamos A por las primeras k componentes de la SVD, perdemos una aportación del orden del valor singular σ_{k+1}. Esto es lo que se usa en las aplicaciones que mencionamos a continuación.

1. Deficiencia del rango ([**?**, sec. 2.5.5]): tratamiento de matrices donde pequeños cambios en los valores provocan alteración del rango.

2. Reducción de ruido en el procesamiento digital de señales.

3. Restauración de imágenes.

4. Análisis de series temporales.

5. Extracción de información de bases de datos.

6. Compresión de imágenes.

8.3. Límites de potencias

Para escalares α sabemos que $\alpha^k \to 0$ si y solamente si $|\alpha| < 1$, por lo que es natural preguntarnos si algo parecido ocurre con las matrices. La primera tentación es cambiar el módulo $|*|$ por una norma matricial $\|*\|$, pero esto no funciona para las normas habituales. Por ejemplo, si

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix},$$

entonces

$$A^k = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, k \geq 2,$$

pero $\|A\|_1 = 2$.

Lo primero que necesitamos es definir el concepto de límite para sucesiones de matrices. Una norma matricial es una norma vectorial para el espacio de las matrices cuadradas. Por tanto, tiene sentido hablar de la convergencia de una sucesión de matrices. Decimos que $\lim_{k\to\infty} A_k = A$ si

$$\lim_{k\to\infty} \|A_k - A\| = 0.$$

La sucesión que vamos a estudiar es la $\{A^k\}$.

Convergencia a cero

Sea A una matriz cuadrada. Son equivalentes:

1. $\lim_{k\to\infty} A^k = 0$.
2. $\lim_{k\to\infty} A^k \boldsymbol{v} = \mathbf{0}$ para todo $\boldsymbol{v} \in V$.
3. $\rho(A) < 1$.
4. La serie de Neumann $I + A + A^2 + \ldots$ converge.

En tal caso, $(I-A)^{-1}$ existe y $\sum_{k=0}^{\infty} A^k = (I-A)^{-1}$.

PRUEBA:

- 1) $\Rightarrow$ 2) Sea $\|*\|$ norma matricial subordinada a la norma vectorial. Entonces $\lim \left\|A^k\right\| = 0$, y para todo $\boldsymbol{v} \in V$ tenemos que $\left\|A^k \boldsymbol{v}\right\| \leq \left\|A^k\right\| \|\boldsymbol{v}\|$, por lo que $\lim \left\|A^k \boldsymbol{v}\right\| = 0$.
- 2) $\Rightarrow$ 3) Si $\rho(A) > 1$, sea λ el autovalor de módulo máximo y $\boldsymbol{v}$ autovector asociado. Entonces $A^k \boldsymbol{v} = \lambda^k \boldsymbol{v}$ y $\lim \left\|A^k \boldsymbol{v}\right\| = \|\boldsymbol{v}\| \lim |\lambda|^k \neq 0$.
- 3) $\Rightarrow$ 1) Si $P^{-1}AP = J$ es la forma canónica de Jordan de A, entonces

$$A^k = P J^k P^{-1} = P \begin{pmatrix} \ddots & & \\ & J_*^k & \\ & & \ddots \end{pmatrix} P^{-1},$$

donde

$$J_* = \begin{pmatrix} \lambda & 1 & \\ & \ddots & \ddots \\ & & \lambda \end{pmatrix}$$

denota un bloque de Jordan en J. Claramente, $A^k \to 0$ si y solamente si $J_*^k \to 0$ para cada bloque de Jordan, por lo que basta probar que si $\rho(A) < 1$ entonces $J_*^k \to 0$. Si usamos la convención de $\binom{k}{j} = 0$ si $j > k$, tenemos

$$J_*^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \ldots & \binom{k}{m-1}\lambda^{k-m+1} \\ & \lambda^k & \binom{k}{1}\lambda^{k-1} & \ddots & \vdots \\ & & \ddots & \ddots & \binom{k}{2}\lambda^{k-2} \\ & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix}, \tag{8.3.1}$$

con m el tamaño del bloque. Vamos a ver que si $|\lambda| < 1$ entonces

$$\lim_{k\to\infty} \binom{k}{j} \lambda^{k-j} = 0 \text{ para cada valor fijado de } j.$$

Notemos que

$$\binom{k}{j} = \frac{k(k-1)\cdots(k-j+1)}{j!} \leq \frac{k^j}{j!}.$$

Entonces

$$\left| \binom{k}{j} \lambda^{k-j} \right| \leq \frac{k^j}{j!} |\lambda|^{k-j}.$$

El término de la derecha tiende a cero cuando $k \to \infty$, porque k^j tiende a infinito con velocidad polinomial, mientras que $|\lambda|^{k-j}$ tiende a cero con velocidad exponencial. Por tanto, si $|\lambda| < 1$ entonces $J_*^k \to 0$.

- 1) $\Rightarrow$ 4). Se tiene que

$$(I-A)(I+A+A^2+\cdots+A^{n-1}) = I - A^n \to I \text{ cuando } n \to \infty,$$

de donde tenemos la implicación. Además, la matriz $I-A$ es no singular y

$$(I-A)^{-1} = I + A + A^2 + \ldots$$

- 4) $\Rightarrow$ 3). Si $\sum_{k=0}^{\infty} A^k$ converge entonces $\sum_{k=0}^{\infty} J_*^k$ debe converger para cada bloque de Jordan de la forma canónica de A. Por la expresión de J_*^k que hemos visto en la ecuación 8.3.1, esto implica que

$$\left[\sum_{k=0}^{\infty} J_*^k \right]_{ii} = \sum_{k=0}^{\infty} \lambda^k$$

converge para cada autovalor λ de A. Esta serie geométrica converge si y solamente si $|\lambda| < 1$. Por tanto, la convergencia de $\sum_{k=0}^{\infty} A^k$ implica que $\rho(A) < 1$.

□

Nos centramos ahora en la posibilidad de que exista el límite $\lim A^k$ pero que no sea cero. Como ya sabemos, $\lim A^k$ existe si y solamente si existe $\lim J_*^k$ para cada bloque de Jordan de A. También es claro que $\lim J_*^k$ no existe cuando $|\lambda| > 1$, y conocemos el resultado cuando $|\lambda| < 1$. Por ello, debemos examinar el caso $|\lambda| = 1$. Si $|\lambda| = 1$, con $\lambda \neq 1$, es decir, $\lambda = \exp(i\theta)$, con $0 < \theta < 2\pi$, entonces

los términos diagonales λ^k oscilan indefinidamente, y esto hace que no exista $\lim J_*^K$. Cuando $\lambda = 1$,

$$J_*^k = \begin{pmatrix} 1 & \binom{k}{1} & \cdots & \binom{k}{m-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \binom{k}{1} \\ & & & 1 \end{pmatrix}_{m\times m}$$

tiene un valor límite si y solamente si $m = 1$, que es equivalente a decir que la multiplicidad algebraica y geométrica de $\lambda = 1$ coinciden, pues el bloque 1×1 se repetirá p veces, su multiplicidad. Tenemos probado entonces el siguiente resultado:

Límite de potencias

Existe $\lim A^k$ si y solamente si la forma canónica de Jordan es de la forma

$$J = P^{-1}AP = \begin{pmatrix} I_{p\times p} & 0 \\ 0 & K \end{pmatrix}, \tag{8.3.2}$$

donde p es la multiplicidad algebraica (geométrica) de 1, y $\rho(K) < 1$.

Supuesta la existencia de $\lim A^k$, queremos describir dicho límite. Si $p = 0$, ya lo sabemos, dicho límite es la matriz nula. Si $p > 0$, entonces consideremos

$$P = \left(\begin{array}{c|c} P_1 & P_2 \end{array}\right), P^{-1} = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix},$$

con P_1 matriz de orden $n \times p$, Q_1 de orden $p \times n$. Entonces

$$\begin{aligned} \lim A^k &= \lim P \begin{pmatrix} I_{p\times p} & 0 \\ 0 & K^k \end{pmatrix} P^{-1} = P \begin{pmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{pmatrix} P^{-1} \\ &= \left(\begin{array}{c|c} P_1 & P_2 \end{array}\right) \begin{pmatrix} I_{p\times p} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = P_1 Q_1. \end{aligned}$$

Si la multiplicidad algebraica es $p = 1$, entonces $\lim A^k = \boldsymbol{v}\boldsymbol{w}^t$, donde $\boldsymbol{v}$ es autovector de A asociado a 1, y $\boldsymbol{w}^t$ es la primera fila de P^{-1}. Observemos que, en tal caso,

$$\text{si } P^{-1}AP = \begin{pmatrix} 1 & \\ & K \end{pmatrix} \text{ entonces } \begin{pmatrix} 1 & \\ & K^t \end{pmatrix} = P^t A^t (P^t)^{-1} = Q^{-1} A^t Q,$$

donde $Q = (P^{-1})^t$. La primera columna de Q es autovector de A^t asociado al autovalor 1, y esa columna es la primera fila traspuesta de P^{-1}, esto es, $\boldsymbol{w}^t$. Como $P^{-1}P = I$, se sigue que $\boldsymbol{w}^t\boldsymbol{v} = 1$.

En conclusión, $\lim A^k = \boldsymbol{v}\boldsymbol{w}^t$, donde $\boldsymbol{v}$ es autovector de A asociado a 1, y $\boldsymbol{w}$ es autovector de A^t asociado a 1, con $\boldsymbol{w}^t\boldsymbol{v} = 1$.

8.4. Número de condición

Los sistemas de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$ que aparecen en la práctica vienen casi siempre con incertidumbres debidas a errores de modelado (ciertas simplificaciones son siempre necesarias), errores en la recolección de datos (medidas), y errores de redondeo (porque $\sqrt{2}$ o π no se pueden dar exactamente). Además, los errores de redondeo en los cálculos en coma flotante son una continua fuente de variaciones en la solución. En todos los casos es importante estimar el grado de incertidumbre en la solución de $A\boldsymbol{x} = \boldsymbol{b}$. Esto no es difícil cuando A se conoce exactamente y todos los posibles errores están en el lado derecho.

Sea $A\boldsymbol{x} = \boldsymbol{b}$ un sistema en el que se conoce A exactamente, pero el vector $\boldsymbol{b}$ está sujeto a un error $\boldsymbol{e}$, y consideremos $A\hat{\boldsymbol{x}} = \boldsymbol{b} - \boldsymbol{e} = \hat{\boldsymbol{b}}$. Se trata de estimar el error relativo $\|\Delta\boldsymbol{x}\| / \|\boldsymbol{x}\| = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| / \|\boldsymbol{x}\|$ de $\boldsymbol{x}$ en función del error relativo $\|\Delta\boldsymbol{b}\| / \|\boldsymbol{b}\| = \left\|\boldsymbol{b} - \hat{\boldsymbol{b}}\right\| / \|\boldsymbol{b}\| = \|\boldsymbol{e}\| / \|\boldsymbol{b}\|$ de $\boldsymbol{b}$. Vamos a considerar una norma matricial inducida por la norma vectorial. Entonces

$$\|\boldsymbol{b}\| = \|A\boldsymbol{x}\| \leq \|A\|\,\|\boldsymbol{x}\|, \text{ y } \boldsymbol{x} - \hat{\boldsymbol{x}} = A^{-1}\boldsymbol{e},$$

de donde

$$\frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} = \frac{\left\|A^{-1}\boldsymbol{e}\right\|}{\|\boldsymbol{x}\|} \leq \frac{\|A\|\left\|A^{-1}\right\|\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|} = \text{cond}(A)\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|},$$

donde $\text{cond}(A) = \|A\|\left\|A^{-1}\right\|$. En el caso de la norma 2,

$$\text{cond}_2(A) = \frac{\sigma_1}{\sigma_n}.$$

Por otro lado,

$$\|\boldsymbol{e}\| = \|A(\boldsymbol{x} - \hat{\boldsymbol{x}})\| \leq \|A\|\,\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \text{ y } \|\boldsymbol{x}\| \leq \left\|A^{-1}\right\|\|\boldsymbol{b}\|.$$

Entonces

$$\frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \geq \frac{\|\boldsymbol{e}\|}{\|A\|\,\|\boldsymbol{x}\|} \geq \frac{\|\boldsymbol{e}\|}{\|A\|\left\|A^{-1}\right\|\|\boldsymbol{b}\|} = \frac{1}{\text{cond}(A)}\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|}.$$

Tenemos así las siguientes cotas para la incertidumbre:

$$\text{cond}(A)^{-1}\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} = \frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \text{cond}(A)\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|},$$

donde

$$\text{cond}(A) = \|A\| \left\|A^{-1}\right\|.$$

En otras palabras, cuando A está *bien condicionada* (cond(A) pequeño), pequeños errores en $\boldsymbol{b}$ no pueden afectar mucho a la solución; pero cuando A está *mal condicionada* (cond(A) grande), una pequeña variación en $\boldsymbol{b}$ puede dar lugar a una gran variación en $\boldsymbol{x}$. Además, estas cotas se pueden alcanzar para algunas direcciones de variación, aunque para otras puede que no tengan apenas efecto en la solución. Como la dirección del error $\boldsymbol{e}$ es desconocida, adoptamos una estrategia conservadora y debemos proceder con cuidado en sistemas mal condicionados.

¿Qué ocurre si tenemos errores tanto en la matriz como en el término derecho? Se puede probar [**?**, ej. 5.12.11, 5.12.12] que si E es la matriz error, entonces

$$\frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A)\,\|E\| / \|A\|} \left(\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{b}\|} + \frac{\|E\|}{\|A\|} \right).$$

De nuevo, si A está bien condicionada, pequeñas variaciones en A y $\boldsymbol{b}$ producen pequeñas variaciones en la solución.

Propiedades del número de condición

- cond$(A) \geq 1$ para una norma matricial inducida.
- cond$(A) =$ cond(A^{-1}).
- cond$(\lambda A) =$ cond(A) para todo $\lambda \neq 0$.
- Sea A matriz hermitiana no singular, con autovalores $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Entonces
$$\text{cond}_2(A) = \frac{|\lambda_1|}{|\lambda_n|}.$$
- Si U es unitaria y A es una matriz arbitraria, entonces $\text{cond}_2(A) = \text{cond}_2(UA) = \text{cond}_2(AU) = \text{cond}_2(U^* AU)$, es decir, cond_2 es invariante por transformaciones unitarias, y $\text{cond}_2(U) = 1$.
- $\text{cond}_2(A) = \frac{\sigma_1}{\sigma_n}$, donde σ_1 y σ_n son los valores singulares mayor y menor, respectivamente, de la matriz A.

PRUEBA: Si la norma matricial es inducida, entonces $1 = \|I\| \leq \|A\| \left\|A^{-1}\right\| =$ cond(A). Las dos siguientes propiedades son inmediatas.

Si A es hermitiana no singular, entonces sus valores singulares son los valores absolutos de sus autovalores. □

Nota 8.4.1. Sea $\|\cdot\|$ norma matricial subordinada y A matriz hermitiana. Entonces

$$\text{cond}(A) = \|A\| \left\|A^{-1}\right\| \geq \rho(A)\rho(A^{-1}) = \text{cond}_2(A).$$

Esto significa que, para matrices hermitianas, cond_2 es el menor de todos los números de condición.

Ejemplo 8.4.2. Consideremos el sistema $A\boldsymbol{x} = \boldsymbol{b}$ con

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

y supongamos que tenemos una variación en $\boldsymbol{b}$ dada por

$$\Delta\boldsymbol{b} = \begin{pmatrix} 0{,}1 \\ -0{,}1 \\ 0{,}1 \\ -0{,}1 \end{pmatrix}.$$

La solución exacta del sistema es

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

mientras que la del sistema alterado $A\boldsymbol{x} = \boldsymbol{b} + \Delta\boldsymbol{b}$ es

$$\boldsymbol{u} + \Delta\boldsymbol{u} = \begin{pmatrix} 9{,}2 \\ -12{,}6 \\ 4{,}5 \\ -1{,}1 \end{pmatrix}.$$

Para la norma $\|\cdot\|_2$ calculamos los errores relativos y tenemos

$$\frac{\|\Delta\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2} \approx 8{,}2, \frac{\|\Delta\boldsymbol{b}\|_2}{\|\boldsymbol{b}_2\|} \approx 0{,}003.$$

Esto era de esperar porque $\text{cond}_2(A) \approx 2984{,}1$.

Nota 8.4.3. El cálculo del número de condición a partir de la definición implica a la inversa de la matriz, por lo que no es un buen método. En la práctica, se suele calcular como subproducto del proceso de resolución de un sistema. También se suelen dar cotas a $\|A^{-1}\|$

Nota 8.4.4. Una "regla del pulgar"sobre la influencia del número de condición en la validez de la solución es la siguiente. Supongamos que se usa eliminación gaussiana con pivoteo parcial sobre un sistema bien escalado $A\boldsymbol{x} = \boldsymbol{b}$ con aritmética de t-dígitos en coma flotante. Supongamos también que no hay otras fuentes de error. Entonces, si $\text{cond}(A)$ es del orden de 10^p, la solución calculada es precisa con $t-p$ dígitos significativos. En otras palabras, esperamos una pérdida de unos p dígitos. Por ejemplo, consideremos el siguiente sistema:

$$\begin{array}{rll} ,835x & +,667y & = ,168 \\ ,333x & +,266y & = ,067. \end{array}$$

Entonces

$$A^{-1} = \begin{pmatrix} -266000 & 667000 \\ 333000 & -835000 \end{pmatrix}, \|A\|_1 = 1{,}168, \left\|A^{-1}\right\|_1 = 1502000,$$

$$\text{cond}_1(A) = 1754336 \approx 1{,}7 \times 10^6.$$

El valor de $\text{cond}_1(A)$ no es tan importante como su orden de magnitud. Lo anterior indica que el cambio relativo en la solución puede ser del orden de un millón de veces el cambio relativo en A. Así, si usamos eliminación gaussiana con 8 dígitos para resolver el sistema, únicamente podemos esperar $t-p = 8-6 = 2$ dígitos significativos. Esto no significa que podamos tener suerte y alcancemos mayor precisión.

Justifiquemos la anterior afirmación. Tenemos la relación

$$\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}\, A\frac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} \right).$$

Supongamos, por simplicidad, que

$$\frac{\|\Delta A\|}{\|A\|} = \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} = 10^{-d}, \text{ y } \text{cond}(A)\frac{\|\Delta A\|}{\|A\|} << 1.$$

Entonces $\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}$ es aproximadamente menor o igual que $2 \times \text{cond}(A) \times 10^{-d}$.

Entonces, si los datos tienen un error relativo del orden de 10^{-d} y si el error relativo de la solución tiene que garantizarse que sea menor o igual que 10^{-t}, entonces $\text{cond}(A)$ tiene que ser menor o igual que $\frac{1}{2} \times 10^{d-t}$. Así, el carácter bien o mal condicionado de un sistema depende de la precisión de los datos y de la magnitud del error que se puede tolerar en la solución.

Ejemplo 8.4.5. Estudio de la matriz de Hilbert. El ejemplo más famoso de matriz mal condicionada es la matriz de Hilbert, definida como

$$H = (h_{ij} = \frac{1}{i+j-1}).$$

Si notamos por H_n la matriz de Hilbert de orden n, entonces

$$H_4 = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}.$$

Estas matrices son simétricas, definidas positivas, y su número de condición crece con n. Por ejemplo,

$$\text{cond}_2(H_4) = 1{,}5514e+4, \text{cond}_2(H_8) = 1{,}5258e+10.$$

Los números de condición no varían mucho aunque se usen diferentes normas:

n	$\text{cond}_1(H_n)$	$\text{cond}_2(H_n)$	$\text{cond}_\infty(H_n)$
3	748	524.06	748
6	$2{,}907e+7$	$1{,}4951e+7$	$2{,}907e+7$
9	$1{,}0997e+12$	$4{,}9315e+11$	$1{,}0997e+12$
12	$3{,}7983e+16$	$1{,}6995e+16$	$3{,}7983e+16$

En general, no es necesario calcular exactamente el número de condición. Existen estimadores de los mismos, que se pueden calcular más fácilmente. por ejemplo, en MATLAB tenemos la función `condest`, que ofrece los siguientes resultados.

n	$\text{cond}_1(H_n)$	`condest`$(H_n, 1)$
3	748	748
6	$2{,}907e+7$	$2{,}907e+7$
9	$1{,}0997e+12$	$1{,}0997e+12$
12	$3{,}7983e+16$	$3{,}7983e+16$

Sea $\boldsymbol{z}$ el vector cuyas componentes son iguales a 1, y sea $\boldsymbol{b} = H_n\boldsymbol{z}$. Si resolvemos el sistema $H_n\boldsymbol{x} = \boldsymbol{b}$, deberíamos obtener $\boldsymbol{z}$ como solución, en teoría. Vamos a ver lo que ocurre en los casos $n = 4, 8, 12, 16$, y comparamos el error relativo de la solución con la cota dada por el número de condición. Los cálculos los hacemos con la norma $\|\cdot\|_2$.

n	$\frac{\|x-z\|}{\|z\|}$	$\text{cond}_2(H_n)\frac{\|r\|}{b}$
4	$1{,}8713e-13$	$6{,}3033e-13$
8	$1{,}0147e-7$	$1{,}9161e-6$
12	0,0819	2,259
16	2,9134	$1{,}102e+2$

El ejemplo anterior exagera en cierta forma el carácter intratable de las matrices de Hilbert. Para $n = 12$ hemos obtenido una solución no muy buena. En realidad, la bondad de la solución depende no solamente de la matriz de coeficientes, sino también del vector $\boldsymbol{b}$. La mayoría de las elecciones de $\boldsymbol{b}$ no nos dará tan mal resultado. Vamos a realizar unos ejemplos para $n = 12$.

1. Sea $\boldsymbol{z}$ el vector de unos, y resolvemos el sistema $H_{12}\boldsymbol{y} = \boldsymbol{z}$.

$$\boldsymbol{y} = \begin{bmatrix} -11{,}73898 \\ 1683{,}25293 \\ -59039{,}42645 \\ 887108{,}96161 \\ -7106863{,}06612 \\ 33868482{,}96454 \\ -101706871{,}33084 \\ 197355775{,}18879 \\ -246878130{,}40362 \\ 192141309{,}97641 \\ -84590742{,}76364 \\ 16087442{,}23886 \end{bmatrix}.$$

   Observemos que $\|\boldsymbol{y}\|$ es grande.

2. Sea $\boldsymbol{b} = H_{12}\boldsymbol{y}$. En principio, $\boldsymbol{b}$ debería ser igual a $\boldsymbol{z}$, pero por los errores de redondeo, es algo distinto. En concreto,

$$\boldsymbol{b} = \begin{bmatrix} 1{,}0000000054 \\ 1{,}0000000040 \\ 1{,}0000000044 \\ 1{,}0000000033 \\ 1{,}0000000010 \\ 1{,}0000000016 \\ 0{,}9999999980 \\ 0{,}9999999987 \\ 1{,}0000000020 \\ 1{,}0000000003 \\ 1{,}0000000019 \\ 1{,}0000000036 \end{bmatrix}, \text{ y } \|\boldsymbol{b}-\boldsymbol{z}\|_2 = 1{,}021672962136328e-008.$$

3. Consideremos ahora el sistema $H_{12}\boldsymbol{x} = \boldsymbol{b}$. En la forma en la que hemos definido $\boldsymbol{b}$, la solución del sistema debe ser $\boldsymbol{y}$. Sin embargo, la experiencia del ejemplo anterior nos sugiere que la solución calculada $\hat{\boldsymbol{x}}$ puede ser bastante diferente de $\boldsymbol{y}$. Veamos lo que ocurre:

$$\hat{\boldsymbol{x}} = \begin{bmatrix} -11{,}1229779077 \\ 1602{,}5823404322 \\ -56438{,}9935515628 \\ 850989{,}2113072880 \\ -6837963{,}3603667002 \\ 32672039{,}6311306580 \\ -98338256{,}5577669890 \\ 191204241{,}5733344900 \\ -239611909{,}8317356100 \\ 186785154{,}0078478500 \\ -82351251{,}5060592290 \\ 15681947{,}6979349870 \end{bmatrix}, \|\hat{\boldsymbol{x}} - \boldsymbol{y}\|_2 = 1{,}172007429552991e+007.$$

Observemos que coinciden en los dos primeros dígitos significativos. No es tan malo como ocurría en el ejemplo anterior. El error relativo es

$$\frac{\|\hat{\boldsymbol{x}} - \boldsymbol{y}\|_2}{\|\boldsymbol{y}\|_2} = 0{,}029695021063446.$$

4. Calculemos ahora la norma del residuo $\hat{\boldsymbol{r}} = \boldsymbol{b} - H_{12}\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{r}} = \begin{bmatrix} -0{,}0000000095 \\ 0{,}0000000030 \\ -0{,}0000000019 \\ 0{,}0000000015 \\ 0{,}0000000028 \\ 0{,}0000000017 \\ 0{,}0000000023 \\ 0{,}0000000027 \\ -0{,}0000000006 \\ 0{,}0000000006 \\ -0{,}0000000008 \\ -0{,}0000000014 \end{bmatrix}.$$

Cota superior del error relativo: $\text{cond}_2(H_{12})\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} = 5{,}653362621489827e+007$. Observemos que es un valor muy pesimista comparado con el verdadero error relativo.

8.5. Un poco de historia

Referencias históricas en

Joseph F. Grcar, John von Neumann' analysis of gaussian elimination and the origins of modern numerical analysis, SAIM Review, vol. 53(4), 607–682, 2011.

D.A. Grier, When computers were human, Princeton Universit Press, 2005, p. 304 (primer párrafo)

Some simultaneous equation calculations posed unusual problems for Blanch's computers. On these calculations, the computers could follow every step of the computing plan, check the work with a desk calculator to ensure than every step was done properly, and still produce values that were wildly incorrect. Though some blamed the computing plan, Blanch discovered a difficulty that would eventually be called "ill conditioning." Ill-conditioned simultaneous equation problems are fundamentally unstable, just as a coin balanced on its edge is unstable. Rounding the values of an ill-conditioned problem, a simple and innocuous act, can cause the calculation to collapse into a meaningless mess of figures. The only way to fix this problem is to reorganize the computing plan, producing a plan that is algebraically equivalent to the original calculation but avoids certain combinations of the four arithmetic operations.

Capítulo 9

Ajuste por mínimos cuadrados

9.1. Soluciones mínimo cuadráticas

Sea $(V,\bullet)$ un espacio euclídeo (sobre $\mathbb{R}$ o sobre $\mathbb{C}$), y L un subespacio de V. Todo vector $\boldsymbol{w}\in V$ puede escribirse de manera única como $\boldsymbol{w}=\boldsymbol{u}+\boldsymbol{v}$, con $\boldsymbol{u}\in L,\boldsymbol{v}\in L^{\perp}$.

En la situación anterior, llamamos al vector $\boldsymbol{u}$ la *proyección ortogonal* del vector $\boldsymbol{w}$ sobre la variedad L. Escribiremos $\boldsymbol{u}=p_L(\boldsymbol{w})$.

Notaremos, como es habitual, $\|\cdot\|$ a la norma inducida por el producto escalar.

Cálculo de la proyección ortogonal

Sea $\mathscr{B}_L=\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_r\}$ una base ortogonal de la variedad L. Si $\boldsymbol{w}\in V$ entonces

$$p_L(\boldsymbol{w})=\frac{\boldsymbol{w}\bullet\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1+\cdots+\frac{\boldsymbol{w}\bullet\boldsymbol{u}_r}{\|\boldsymbol{u}_r\|^2}\boldsymbol{u}_r.$$

PRUEBA: A partir de $\mathscr{B}_L$ ampliamos a $\mathscr{B}=\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_r,\ldots,\boldsymbol{u}_n\}$ base ortogonal de V. Si $\boldsymbol{w}=\sum_{i=1}^n\alpha_i\boldsymbol{u}_i$, entonces $\boldsymbol{w}\bullet\boldsymbol{u}_j=\alpha_j(\boldsymbol{u}_j\bullet\boldsymbol{u}_j)=\alpha_j\left\|\boldsymbol{u}_j\right\|^2$. De aquí, $\alpha_j=\frac{\boldsymbol{w}\bullet\boldsymbol{u}_j}{\|\boldsymbol{u}_j\|^2}$. Como $L^{\perp}=\langle\boldsymbol{u}_{r+1},\ldots,\boldsymbol{u}_n\rangle$, entonces la parte de $\boldsymbol{w}$ que está en L es $\alpha_1\boldsymbol{u}_1+\cdots+\alpha_r\boldsymbol{u}_r$. □

Tenemos un resultado clásico en espacios vectoriales con producto escalar: el teorema de Pitágoras.

Sean $\boldsymbol{w}_1,\boldsymbol{w}_2\in V$ vectores ortogonales. Entonces

$$\|\boldsymbol{w}_1+\boldsymbol{w}_2\|^2=\|\boldsymbol{w}_1\|^2+\|\boldsymbol{w}_2\|^2.$$

En efecto,

$$\begin{aligned}\|\boldsymbol{w}_1+\boldsymbol{w}_2\|^2 &= (\boldsymbol{w}_1+\boldsymbol{w}_2)\bullet(\boldsymbol{w}_1+\boldsymbol{w}_2)\\ &= \|\boldsymbol{w}_1\|^2+\|\boldsymbol{w}_2\|^2+2(\boldsymbol{w}_1\bullet\boldsymbol{w}_2)=\|\boldsymbol{w}_1\|^2+\|\boldsymbol{w}_2\|^2.\end{aligned}$$

Mejor aproximación a un vector

Sea L subespacio de V y $\boldsymbol{w}\in V$. Entonces $p_L(\boldsymbol{w})$ es el único vector $\boldsymbol{u}\in L$ que minimiza la expresión $\|\boldsymbol{w}-\boldsymbol{u}\|$.

PRUEBA: Dado $\boldsymbol{u}\in L$, entonces $p_L(\boldsymbol{w})-\boldsymbol{u}\in L$ y $\boldsymbol{w}-p_L(\boldsymbol{w})\in L^\perp$. Por el teorema de Pitágoras,

$$\|\boldsymbol{w}-\boldsymbol{u}\|^2=\left\|p_L(\boldsymbol{w})-\boldsymbol{u}\right\|^2+\left\|\boldsymbol{w}-p_L(\boldsymbol{w})\right\|^2\geq\left\|\boldsymbol{w}-p_L(\boldsymbol{w})\right\|$$

y la igualdad se da si y solamente si $p_L(\boldsymbol{w})-\boldsymbol{u}=\boldsymbol{0}$. □

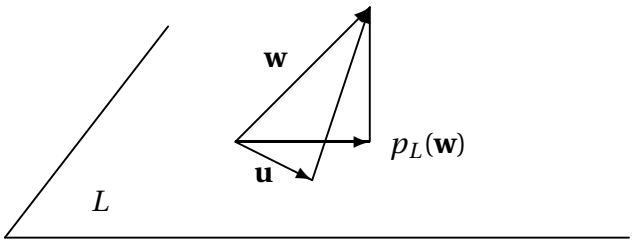

Figura 9.1: Proyección ortogonal

Consideremos el sistema de ecuaciones lineales $A\boldsymbol{x}=\boldsymbol{b}$, con A una matriz de orden $m\times n$. Como el espacio de columnas determinado por A es $\mathrm{Col}(A)=\{A\boldsymbol{v}\mid\boldsymbol{v}\in\mathbb{K}^n\}$, el sistema será compatible si y solamente si el vector $\boldsymbol{b}\in\mathrm{Col}(A)$. Esto es lo que nos dice el teorema de Rouché-Frobenius. Cuando el sistema es incompatible, esto es, cuando $\boldsymbol{b}\notin\mathrm{Col}(A)$, nos interesa buscar un valor de $\boldsymbol{x}$ lo "mejor" posible. ¿Cómo medimos este concepto? A través de una norma vectorial.

Soluciones mínimo cuadráticas

Consideremos $\mathbb{K}^m$ con la estructura euclídea natural. Llamamos ***solución mínimo cuadrática*** del sistema $A\boldsymbol{x}=\boldsymbol{b}$ a un vector $\boldsymbol{u}\in\mathbb{K}^n$ que haga mínima la norma $\|A\boldsymbol{u}-\boldsymbol{b}\|_2$, o lo que es equivalente, que se minimice $(A\boldsymbol{u}-\boldsymbol{b})^*(A\boldsymbol{u}-\boldsymbol{b})$.

En el caso de sistemas compatibles, cualquier solución es mínimo cuadrática. La cuestión es dar un procedimiento para calcularlas en cualquier caso.

Ecuaciones normales

Las soluciones mínimo cuadráticas del sistema $A_{m\times n}\boldsymbol{x}=\boldsymbol{b}$ coinciden con las soluciones del sistema $A^*A\boldsymbol{x}=A^*\boldsymbol{b}$, que es compatible. Este sistema de ecuaciones recibe el nombre de ***ecuaciones normales***.
Si A es de rango pleno por columnas, esto es, rango$(A)=n$, entonces existe una única solución mínimo cuadrática determinada por $\boldsymbol{x}=(A^*A)^{-1}A^*\boldsymbol{b}$.

PRUEBA: La norma $\|A\boldsymbol{u}-\boldsymbol{b}\|_2$ se minimiza cuando $A\boldsymbol{u}$ es la proyección ortogonal de $\boldsymbol{b}$ sobre el espacio Col(A). Esto es equivalente a que $A\boldsymbol{u}-\boldsymbol{b}$ sea ortogonal a Col(A), que está generado por los vectores $A\boldsymbol{e}_i, i=1,\ldots,n$, donde $\boldsymbol{e}_i$ son los vectores de la base estándar. Entonces

$$\begin{aligned} A\boldsymbol{u}-\boldsymbol{b}\perp \text{Col}(A) &\Leftrightarrow (A\boldsymbol{u}-\boldsymbol{b})\bullet A\boldsymbol{e}_i=0, i=1,\ldots,n\\ &\Leftrightarrow \boldsymbol{e}_i^*A^*(A\boldsymbol{u}-\boldsymbol{b})=0, i=1,\ldots,n\\ &\Leftrightarrow \boldsymbol{e}_i^*A^*A\boldsymbol{u}=\boldsymbol{e}_i^*A^*\boldsymbol{b}, i=1,\ldots,n\\ &\Leftrightarrow A^*A\boldsymbol{u}=A^*\boldsymbol{b}, \end{aligned}$$

de donde $\boldsymbol{u}$ es solución mínimo cuadrática del sistema $A\boldsymbol{x}=\boldsymbol{b}$ si y solamente si $\boldsymbol{u}$ es solución de $A^*A\boldsymbol{x}=A^*\boldsymbol{b}$. Este sistema es compatible, porque $A^*\boldsymbol{b}\in$ Col$(A^*)=$ Col(A^*A).

Por último, si A es de rango pleno por columnas, la matriz A^*A es no singular, y el sistema $A^*A\boldsymbol{x}=A^*\boldsymbol{b}$ es compatible determinado, con solución $(A^*A)^{-1}A^*\boldsymbol{b}$. □

Ejemplo 9.1.1. Consideremos el sistema $A\boldsymbol{x}=\boldsymbol{b}$, donde

$$A=\begin{pmatrix} 3 & 2 & 1\\ 1 & 1 & 0\\ -1 & 0 & -1\end{pmatrix}, \boldsymbol{b}=\begin{pmatrix} 2\\ 0\\ 1\end{pmatrix}.$$

Como rango$(A)=2$, rango$(A|\boldsymbol{b})=3$, el sistema es incompatible. Pasamos a resolver $A^tA\boldsymbol{x}=A^t\boldsymbol{b}$. En este caso,

$$B=A^tA=\begin{bmatrix} 11 & 7 & 4\\ 7 & 5 & 2\\ 4 & 2 & 2\end{bmatrix}, \boldsymbol{c}=A^t\boldsymbol{b}=\begin{bmatrix} 5\\ 4\\ 1\end{bmatrix}, \text{ y } \begin{pmatrix} B & \boldsymbol{c}\end{pmatrix}\overset{\text{rref}}{\longrightarrow}\begin{bmatrix} 1 & 0 & 1 & -1/2\\ 0 & 1 & -1 & 3/2\\ 0 & 0 & 0 & 0\end{bmatrix}$$

por lo que el sistema de ecuaciones normales es compatible indeterminado. Las soluciones son de la forma

$$\begin{cases} x &= -\frac{1}{2} - \lambda \\ y &= \frac{3}{2} + \lambda \\ z &= \lambda \end{cases}$$

Ejemplo 9.1.2. Vamos a considerar el problema de ajuste por mínimos cuadrados de una curva. Se trata de calcular un polinomio

$$p(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_{n-1} t^{n-1}$$

de un grado determinado que sea próximo, en el sentido de mínimos cuadrados, a un conjunto de puntos

$$\mathscr{D} = \{(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m)\},$$

donde los t_i son números distintos y $n \leq m$. Si llamamos $\epsilon_i = p(t_i) - b_i$, se trata de minimizar la suma de cuadrados

$$\sum_{i=1}^{m} \epsilon_i^2 = \sum_{i=1}^{m} (p(t_i) - b_i)^2 = (A\boldsymbol{x} - \boldsymbol{b})^t (A\boldsymbol{x} - \boldsymbol{b}),$$

donde

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \ldots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & t_m & t_m^2 & \ldots & t_m^{n-1} \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{n-1} \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

En otras palabras, el polinomio de grado $n-1$ que verifica la propiedad de mínimos cuadrados se obtiene como solución del problema de mínimos cuadrados asociado al sistema $A\boldsymbol{x} = \boldsymbol{b}$. Este polinomio es único porque la matriz A es de Vandermonde, con $n \leq m$ y rango$(A) = n$.

Ejemplo 9.1.3. Consideremos el siguiente conjunto de datos:

t_i	1,0	1,5	2,0	2,5	3,0
y_i	1,1	1,2	1,3	1,3	1,4

Queremos calcular la recta $y = \alpha_0 + \alpha_1 t$ de mejor ajuste. Para ello, planteamos las condiciones

$$\begin{aligned} \alpha_0 + 1{,}0 \cdot \alpha_1 &= 1{,}1, \\ \alpha_0 + 1{,}5 \cdot \alpha_1 &= 1{,}2, \\ \alpha_0 + 2{,}0 \cdot \alpha_1 &= 1{,}3, \\ \alpha_0 + 2{,}5 \cdot \alpha_1 &= 1{,}3, \\ \alpha_0 + 3{,}0 \cdot \alpha_1 &= 1{,}4. \end{aligned}$$

Esto se traduce en el sistema $A\boldsymbol{\alpha} = \boldsymbol{y}$, donde

$$A = \begin{bmatrix} 1{,}0 & 1{,}0 \\ 1{,}0 & 1{,}5 \\ 1{,}0 & 2{,}0 \\ 1{,}0 & 2{,}5 \\ 1{,}0 & 3{,}0 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 1{,}1 \\ 1{,}2 \\ 1{,}3 \\ 1{,}3 \\ 1{,}4 \end{bmatrix}.$$

La solución mínimo cuadrática se obtiene como solución del sistema de ecuaciones normales $A^t A\boldsymbol{\alpha} = A^t \boldsymbol{y}$:

$$\begin{bmatrix} 5{,}00 & 10{,}00 \\ 10{,}00 & 22{,}50 \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 6{,}3 \\ 12{,}95 \end{pmatrix}.$$

La solución de este sistema es

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0{,}98 \\ 0{,}14 \end{pmatrix}.$$

Podemos dibujar los datos.

```
t = [1.0 1.5 2.0 2.5 3.0]';
y = [1.1 1.2 1.3 1.3 1.4]';
A1 = [ones(5,1), t];
alpha0 = 0.98;alpha1 = 0.14;
escala = 0:0.01:3;
recta = alpha0 + alpha1*escala;
plot(escala, recta, t,y,'bd','MarkerFaceColor','g','MarkerSize',10)
```

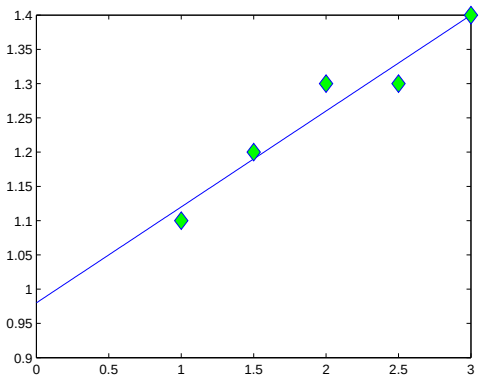

Figura 9.2: Ajuste lineal

El error cuadrático medio es la norma del residuo:

$$\boldsymbol{r} = \boldsymbol{y} - A\hat{\boldsymbol{\alpha}} \text{ y } \|\boldsymbol{r}\| = 0{,}0548.$$

Ahora podemos realizar una interpolación de tipo cuadrático, mediante el ajuste con un polinomio de la forma $y = \alpha_0 + \alpha_1 t + \alpha_2 t^2$. El sistema a resolver es

$$\begin{aligned}
\alpha_0 + 1{,}0 \cdot \alpha_1 + (1{,}0)^2\alpha_2 &= 1{,}1,\\
\alpha_0 + 1{,}5 \cdot \alpha_1 + (1{,}5)^2\alpha_2 &= 1{,}2,\\
\alpha_0 + 2{,}0 \cdot \alpha_1 + (2{,}0)^2\alpha_2 &= 1{,}3,\\
\alpha_0 + 2{,}5 \cdot \alpha_1 + (2{,}5)^2\alpha_2 &= 1{,}3,\\
\alpha_0 + 3{,}0 \cdot \alpha_1 + (3{,}0)^2\alpha_2 &= 1{,}4.
\end{aligned}$$

La matriz de coeficientes es

$$A = \begin{bmatrix} 1{,}00 & 1{,}00 & 1{,}00 \\ 1{,}00 & 1{,}50 & 2{,}25 \\ 1{,}00 & 2{,}00 & 4{,}00 \\ 1{,}00 & 2{,}50 & 6{,}25 \\ 1{,}00 & 3{,}00 & 9{,}00 \end{bmatrix},$$

y el sistema de ecuaciones normales $A^tA\boldsymbol{\alpha} = A^t\boldsymbol{y}$ es

$$\begin{bmatrix} 5{,}000 & 10{,}000 & 22{,}500 \\ 10{,}000 & 22{,}500 & 55{,}000 \\ 22{,}500 & 55{,}000 & 142{,}125 \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{bmatrix} 6{,}300 \\ 12{,}950 \\ 29{,}725 \end{bmatrix}, \text{ con solución } \boldsymbol{\alpha} = \begin{bmatrix} 0{,}88000 \\ 0{,}25429 \\ -0{,}02857 \end{bmatrix}.$$

El valor del residuo es $\|\boldsymbol{r}\|_2 = \|\boldsymbol{y} - A\boldsymbol{\alpha}\|_2 = 0{,}0478$, que es menor que en el ajuste lineal, como era de esperar, pues el espacio vectorial de los polinomios de grado menor o igual que 2 contiene a las expresiones lineales.

```
A2 = [ones(5,1), t, t.^2];
alpha = A\y;
parabola = alpha(1) + alpha(2)*escala + alpha(3)*escala.^2;
plot(escala, recta, escala, parabola, ...
t,y,'bd','MarkerFaceColor','g','MarkerSize',10)
```

Ejemplo 9.1.4. La interpolación de un conjunto de datos no tiene que ser con polinomios. Consideremos los puntos del ejemplo anterior, pero ahora vamos a realizar un ajuste con las funciones $1, \exp(t), \exp(-t)$. Entonces el sistema queda

$$\begin{aligned}
\alpha_0 + \alpha_1 e^{1{,}0} + \alpha_2 e^{-1{,}0} &= 1{,}1,\\
\alpha_0 + \alpha_1 e^{1{,}5} + \alpha_2 e^{-1{,}5} &= 1{,}2,\\
\alpha_0 + \alpha_1 e^{2{,}0} + \alpha_2 e^{-2{,}0} &= 1{,}3,\\
\alpha_0 + \alpha_1 e^{2{,}5} + \alpha_2 e^{-2{,}5} &= 1{,}3,\\
\alpha_0 + \alpha_1 e^{3{,}0} + \alpha_2 e^{-3{,}0} &= 1{,}4.
\end{aligned}$$

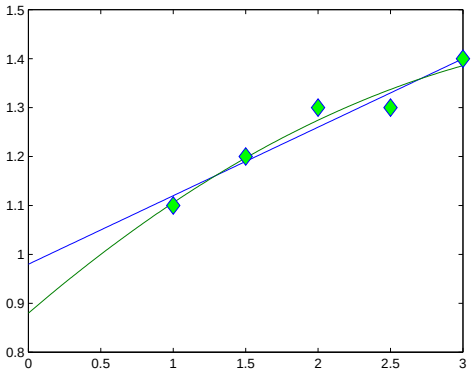

Figura 9.3: Ajuste lineal y cuadrático

La matriz de coeficientes es

$$A = \begin{bmatrix} 1{,}00000 & 2{,}71828 & 0{,}36788 \\ 1{,}00000 & 4{,}48169 & 0{,}22313 \\ 1{,}00000 & 7{,}38906 & 0{,}13534 \\ 1{,}00000 & 12{,}18249 & 0{,}08208 \\ 1{,}00000 & 20{,}08554 & 0{,}04979 \end{bmatrix},$$

y el sistema de ecuaciones normales $A^t A\boldsymbol{\alpha} = A^t \boldsymbol{y}$ tiene como solución

$$\boldsymbol{\alpha} = \begin{bmatrix} 1{,}32479 \\ 0{,}00483 \\ -0{,}64107 \end{bmatrix}.$$

El valor del residuo es $\|\boldsymbol{y} - A\boldsymbol{\alpha}\|_2 = 0{,}0421$.

```
A3 = A3 = [ones(5,1), exp(t), exp(-t)];
beta = A3\y;
expon = beta(1) + beta(2)*exp(escala) + beta(3)*exp(-escala);
plot(escala, recta, escala, parabola,escala, expon, t,y,...
'bd','MarkerFaceColor','g','MarkerSize',10)
```

Ejemplo 9.1.5. Mediante la generación de matrices aleatorias, se han obtenido los tiempos de ejecución de la descomposición de Cholesky en MATLAB de 38 matrices con dimensiones entre 1000 y 6000, con índice asociado entre 1 y 38. Los resultados se encuentran en el fichero `simulachol`. Para cargarlos, procedemos como sigue:

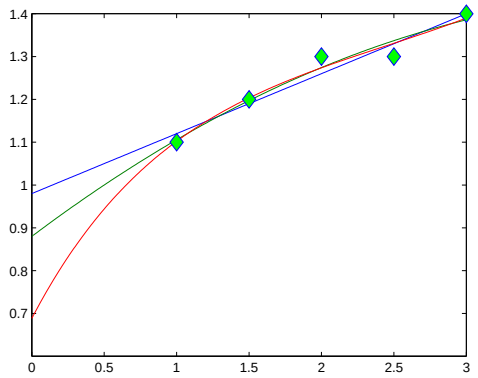

Figura 9.4: Ajuste lineal, cuadrático y exponencial

```
>> load simulachol
>> varx = (1:38)'; % valores de las variables de regresi\'{o}n
>> plot(varx, simulachol, 'og');
```

En el gráfico 9.5 aparecen los puntos del experimento, y se trata de hacer un ajuste de los mismos.

Parece claro que un ajuste lineal no es el más adecuado. Comencemos con uno de tipo cuadrático $y = \alpha_0 + \alpha_1 x + \alpha_2 x^2$. Para ello, la matriz de coeficientes del sistema es

$$A_2 = \begin{pmatrix} 1 & k & k^2 \end{pmatrix}_{k=1,\ldots,38}.$$

y el término independiente está formado por el vector `simulachol`. Construimos el sistema:

```
>> A2 = [ones(38,1), varx, varx.^2];
>> solu2 = A2 \ simulachol

solu2 =

    1.0537
   -0.1388
    0.0225
```

Podemos dibujar la curva resultante, que aparece en la figura 9.6:

```
>> escala = (1:0.1:38)';
>> curva2 = solu2(1) + solu2(2) * escala + solu2(3) * escala.^2;
>> plot(varx, simulachol, 'og', escala, curva2)
```

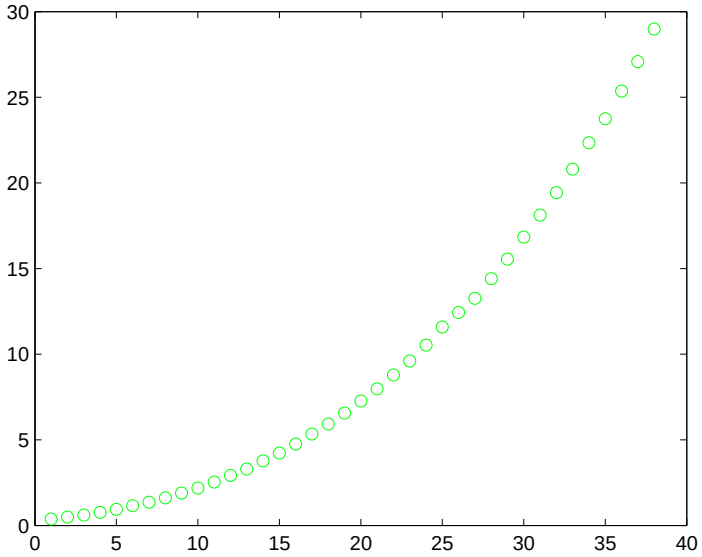

Figura 9.5: Muestra de tiempos de ejecución del algoritmo de Cholesky.

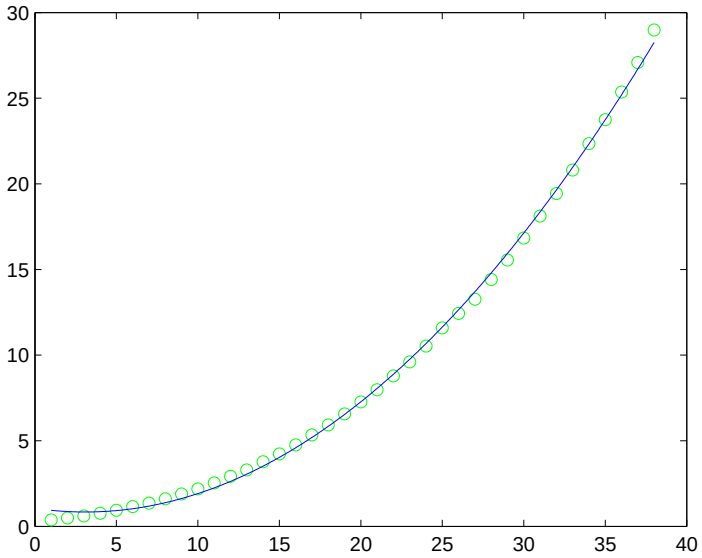

Figura 9.6: Ajuste cuadrático.

El ajuste parece adecuado, pero hay que analizar los residuos. El vector de residuos debe seguir una distribución normal de media cero y varianza σ^2. Representemos gráficamente el resultado (figura 9.7):

```
>> residuo2 = simulachol - ...
   (solu2(1) + solu2(2) * varx + solu2(3) * varx.^2);
>> plot(varx,residuo2,'or')
```

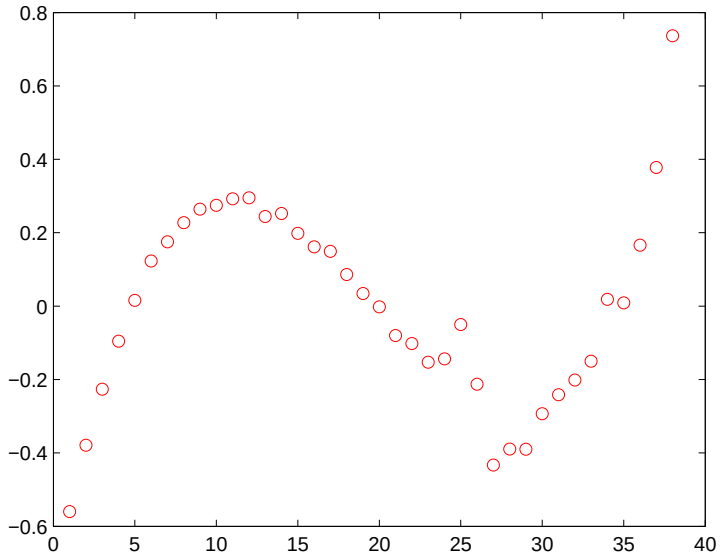

Figura 9.7: Residuos del ajuste cuadrático

La forma del residuo indica que hay una relación funcional entre ellos, por lo que probamos el ajuste cúbico, que se representa en la figura 9.8.

```
>> A3 = [ones(38,1), varx, varx.^2, varx.^3];
>> solu3 = A3 \simulachol;
>> curva3 = solu3(1) + solu3(2) * escala + ...
   solu3(3) * escala.^2 + solu3(4) * escala.^3;
>> plot(varx, simulachol, 'og', escala, curva2, 'r', ...
   escala, curva3, 'b')
```

Se ve que el ajuste es más adecuado, y la forma de los residuos también es diferente (figura 9.9):

```
>> residuo3 = simulachol - ...
   (solu3(1) + solu3(2) * varx + solu3(3) * varx.^2 + solu3(4) * varx.^3);
>> plot(varx,residuo2,'or', varx, residuo3, '+g')
```

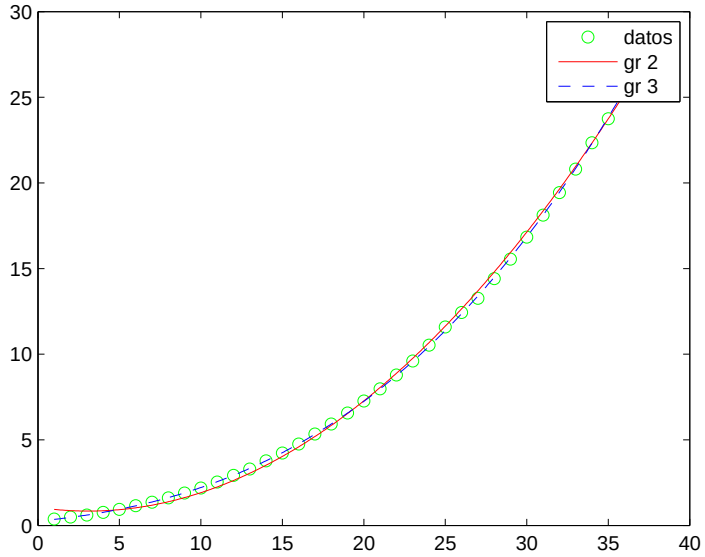

Figura 9.8: Ajustes cuadrático y cúbico

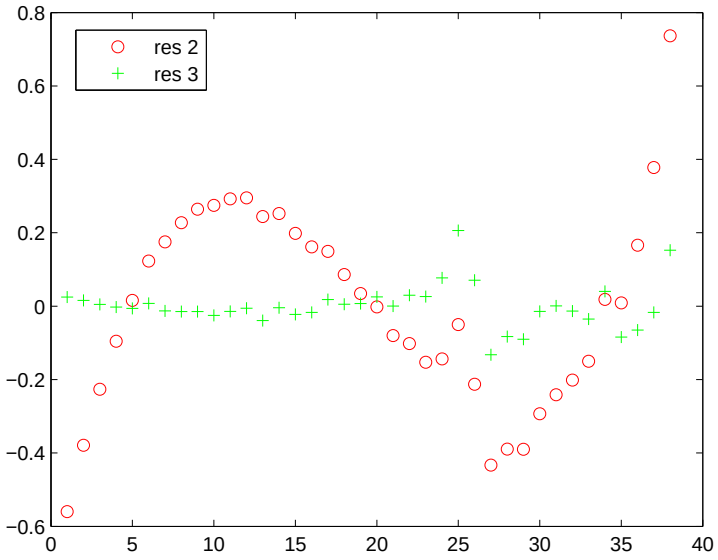

Figura 9.9: Residuos del ajuste cuadrático y cúbico

Por tanto, el ajuste cúbico es más adecuado.

9.2. * Teorema de Gauss-Markov

Las características numéricas de un fenómeno físico se denominan parámetros, y el objetivo es diseñar funciones o reglas que llamamos estimadores que usen observaciones para estimar los parámetros de interés. Por ejemplo, la media de una característica numérica de una población es un parámetro, y la media de una población es un estimador de ese parámetro.

Los buenos estimadores deben ser insesgados y de mínima varianza. Tomemos X, Y variables aleatorias, y notaremos, como es habitual,

$E[X] = \mu_X$ la media de X,
$\text{var}[X] = E[(X-\mu_X)^2] = E[X^2] - \mu_X^2$ la varianza de X,
$\text{cov}[X,Y] = E((X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y$ la covarianza de X, Y.

Estimadores insesgados de mínima varianza

Un estimador $\hat{\theta}$, considerado como variable aleatoria, de un parámetro θ se dice ***insesgado*** cuando $E[\hat{\theta}] = \theta$. Decimos además que $\hat{\theta}$ es estimador insesgado de ***mínima varianza*** de θ cuando $\text{var}[\hat{\theta}] \leq \text{var}[\hat{\phi}]$ para todos los estimadores insesgados $\hat{\phi}$ de θ.

Estas ideas permiten demostrar por qué el método de mínimos cuadrados es la mejor forma para determinar la relación entre los datos. Sea Y una variable que suponemos relacionada linealmente con otras $X_1, X_2, \ldots, X_n$ mediante una ecuación

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

donde las β_i son parámetros desconocidos. Supongamos que los valores asumidos para las X_i no están sujetos a error o variación y pueden ser exactamente calculados pero, por problemas de medida, los valores de Y no pueden ser exactamente medidos. Entonces nos queda

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon$$

donde ε es una variable aleatoria que mide el error de medida. El problema es determinar los parámetros β_i mediante la observación de valores de Y en m puntos diferentes $X_{i*} = (x_{i1}, x_{i2}, \ldots, x_{in}) \in \mathbb{R}^n$, donde x_{ij} es el valor de X_j que se

usa para hacer la i-ésima observación. Si llamamos y_i la variable aleatoria que representa la salida de la i-ésima observación de Y, tenemos que

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \varepsilon_i, i = 1, 2, \ldots, m, \tag{9.2.1}$$

donde ε_i es una variable aleatoria que mide el error de la i-ésima observación. Se supone en general que los errores de observación no están correlacionados entre sí, pero tienen la misma varianza (desconocida) y media cero. En otras palabras, tenemos que,

$$E[\varepsilon_i] = 0 \text{ para cada } i, \operatorname{cov}[\varepsilon, \varepsilon_j] = \begin{cases} \sigma^2 & \text{cuando } i = j, \\ 0 & \text{cuando } i \neq j \end{cases}$$

Si

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1n} \\ 1 & x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \ldots & x_{mn} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

las ecuaciones 9.2.1 se pueden escribir como $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. En la práctica, los puntos X_{i*} en donde se hacen las observaciones se pueden tomar, casi siempre, que hagan que la matriz X tenga rango $n+1$. El modelo estándar lineal queda

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ con } \begin{cases} \operatorname{rango}(X) = n+1, \\ E[\boldsymbol{\varepsilon}] = \mathbf{0}, \\ \operatorname{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I, \end{cases}$$

donde hemos adoptado la convención

$$E[\boldsymbol{\varepsilon}] = \begin{pmatrix} E[\varepsilon_1] \\ E[\varepsilon_2] \\ \vdots \\ E[\varepsilon_m] \end{pmatrix}, \operatorname{cov}[\boldsymbol{\varepsilon}] = \begin{pmatrix} \operatorname{cov}[\varepsilon_1, \varepsilon_1] & \operatorname{cov}[\varepsilon_1, \varepsilon_2] & \ldots & \operatorname{cov}[\varepsilon_1, \varepsilon_m] \\ \operatorname{cov}[\varepsilon_2, \varepsilon_1] & \operatorname{cov}[\varepsilon_2, \varepsilon_2] & \ldots & \operatorname{cov}[\varepsilon_2, \varepsilon_m] \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}[\varepsilon_m, \varepsilon_1] & \operatorname{cov}[\varepsilon_m, \varepsilon_2] & \ldots & \operatorname{cov}[\varepsilon_m, \varepsilon_m] \end{pmatrix}.$$

El problema es determinar el mejor (mínima varianza) estimador lineal insesgado para las componentes de $\boldsymbol{\beta}$. Esto nos lo da el siguiente resultado.

Teorema de Gauss-Markov

Para el modelo lineal estándar, el estimador lineal insesgado de mínima varianza para β_i es la i-ésima componente $\hat{\beta}_i$ del vector

$$\hat{\boldsymbol{\beta}} = (X^t X)^{-1} X^t \boldsymbol{y} = X^+ \boldsymbol{y}.$$

En otras palabras, el mejor estimador lineal insesgado de $\boldsymbol{\beta}$ es la solución mínimo cuadrática de $X\hat{\boldsymbol{\beta}} = \boldsymbol{y}$.

PRUEBA: Es claro que $\hat{\boldsymbol{\beta}} = X^+ \boldsymbol{y}$ es un estimador lineal de $\boldsymbol{\beta}$ porque cada componente $\hat{\beta}_i = \sum_k [X^+]_{ik} y_k$ es una función lineal de las observaciones.

El que $\hat{\boldsymbol{\beta}}$ sea insesgado se sigue de

$$E[\boldsymbol{y}] = E[X\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = E[X\boldsymbol{\beta}] + \mathbf{0} = X\boldsymbol{\beta}$$

por lo que

$$E[\hat{\boldsymbol{\beta}}] = E[X^+ \boldsymbol{y}] = X^+ E[\boldsymbol{y}] = X^+ X\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Para verificar que $\hat{\boldsymbol{\beta}} = X^+ \boldsymbol{y}$ tiene mínima varianza entre los estimadores lineales insesgados de $\boldsymbol{\beta}$, sea $\boldsymbol{\beta}'$ un estimador lineal insesgado arbitrario de $\boldsymbol{\beta}$. Como es lineal, existe una matriz L tal que $\boldsymbol{\beta}' = L\boldsymbol{y}$. El carácter insesgado implica que

$$\boldsymbol{\beta} = E[\boldsymbol{\beta}'] = E[L\boldsymbol{y}] = LE[\boldsymbol{y}] = LX\boldsymbol{\beta}.$$

Queremos que $\boldsymbol{\beta} = LX\boldsymbol{\beta}$ se verifique independientemente de los valores de β, por lo que $LX = I$. Para $i \neq j$, se tiene que

$$0 = \operatorname{cov}[\varepsilon_i, \varepsilon_j] = E[\varepsilon_i \varepsilon_j] - E[\varepsilon_i]E[\varepsilon_j] \Rightarrow E[\varepsilon_i \varepsilon_j] = E[\varepsilon_i]E[\varepsilon_j] = 0.$$

Por otro lado,

$$\operatorname{cov}[y_i, y_j] = \begin{cases} E[(y_i - E[y_i])^2] = E[\varepsilon_i^2] = Var[\varepsilon_i] = \sigma^2 & \text{cuando } i = j, \\ E[(y_i - E[y_i])(y_j - E[y_j])] = E[\varepsilon_i \varepsilon_j] = 0 & \text{cuando } i \neq j. \end{cases}$$

Recordemos que si $\operatorname{cov}[W, Z] = 0$ entonces $\operatorname{var}[aW + bZ] = a^2 \operatorname{var}[W] + b^2 \operatorname{var}[Z]$. Entonces

$$\operatorname{var}[\beta_i'] = \operatorname{var}[L_{i*}\boldsymbol{y}] = \operatorname{var}[\sum_{k=1}^{m} l_{ik} y_k] = \sigma^2 \sum_{k=1}^{m} l_{ik}^2 = \sigma^2 \|L_{i*}\|_2^2.$$

Dado que $LX = I$, se sigue que $\operatorname{var}[\beta_i']$ es mínima si y sólo si L_{i*} es la solución mínimo cuadrática del sistema $\boldsymbol{z}^t X = \boldsymbol{e}_i^t$, equivalente a $X^t \boldsymbol{z} = \boldsymbol{e}_i$. Su solución mínimo cuadrática es

$$\boldsymbol{z} = (X^t)^+ \boldsymbol{e}_i, \text{ de donde } \boldsymbol{z}^t = \boldsymbol{e}_i^t X^+ = X_{i*}^+.$$

Entonces var$[\beta_i^*]$ es mínima si y sólo si $L_{i*} = X_{i*}^+$, para $i = 1,2,\ldots,m$, que es lo mismo que $L = X^+$. Por tanto, las componentes de $\hat{\boldsymbol{\beta}} = X^+\boldsymbol{y}$ son los únicos estimadores lineales insesgados de mínima varianza para los parámetros en $\boldsymbol{\beta}$.
□

Una cuestión que podemos responder aquí de una forma aproximada es la confianza de los coeficientes del vector $\hat{\boldsymbol{\beta}}$. Una forma de medir dicha confianza es a través de su varianza, que podemos calcular. Consideremos el modelo lineal estándar que hemos descrito anteriormente:

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ con } \begin{cases} \text{rango}(X) = n+1, \\ E[\boldsymbol{\varepsilon}] = \mathbf{0}, \\ \text{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I, \end{cases}$$

Llamemos $\hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{\varepsilon}$. Entonces $\text{cov}(\hat{\boldsymbol{y}}) = \text{cov}(-\boldsymbol{\varepsilon}) = \sigma^2 I$, y

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\beta}}) &= \text{cov}((X^tX)^{-1}X^t\hat{\boldsymbol{y}}) \\ &= (X^tX)^{-1}X^t\,\text{cov}(\hat{\boldsymbol{y}})X(X^tX)^{-1} \\ &= (X^tX)^{-1}X^t\sigma^2 IX(X^tX)^{-1} = \sigma^2(X^tX)^{-1}. \end{aligned}$$

9.3. * Variaciones de mínimos cuadrados

9.3.1. Mínimos cuadrados ponderados

Consideremos el modelo de regresión lineal múltiple

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

donde ahora var$\boldsymbol{\epsilon} \neq \sigma^2 I_m$. En este caso, nuestro estimador $\hat{\boldsymbol{\beta}} = (X^tX)^{-1}X^t\boldsymbol{y}$ es todavía el estimador de mínimos cuadrados de $\boldsymbol{\beta}$, pero pierde el carácter de insesgado y mínima varianza. Vamos a suponer, en primer lugar, que las variables aleatorias ε_i no están correlacionadas, pero sus varianzas no son las mismas. En términos matriciales, var$\boldsymbol{\varepsilon} = \Omega = \sigma^2 C$, donde $C = \text{diag}(c_1^2,\ldots,c_m^2)$, y cada c_i es una constante conocida. Este problema de regresión se conoce como mínimos cuadrados ponderados. El estimador para mínimos cuadrados ponderados de $\boldsymbol{\beta}$ se obtiene mediante una simple transformación de tal forma que la regresión múltiple clásica se aplica al modelo transformado. Consideremos la matriz $C^{-1/2} = \text{diag}(c_1^{-1},\ldots,c_m^{-1})$ y transformemos el problema original multiplicando ambos lados de la ecuación por $C^{-1/2}$. El nuevo modelo es

$$C^{-1/2}\boldsymbol{y} = C^{-1/2}X\boldsymbol{\beta} + C^{-1/2}\boldsymbol{\varepsilon},$$

o de forma equivalente

$$\boldsymbol{y}_* = X_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*,$$

donde

$$\boldsymbol{y}_* = C^{-1/2}\boldsymbol{y}, X_* = C^{-1/2}X \text{ y } \boldsymbol{\varepsilon}_* = C^{-1/2}\boldsymbol{\epsilon}.$$

Por un lado se tiene que $E[\boldsymbol{\epsilon}_*] = C^{-1/2}E[\boldsymbol{\varepsilon}] = 0$. Por otro, la matriz de covarianza de $\boldsymbol{\varepsilon}_*$ es igual a

$$\begin{aligned}\operatorname{var}\boldsymbol{\varepsilon}_* &= \operatorname{var}(C^{-1/2}\boldsymbol{\varepsilon}) = C^{-1/2}(\operatorname{var}\boldsymbol{\varepsilon})C^{-1/2}\\ &= C^{-1/2}\sigma^2 CC^{-1/2} = \sigma^2 I_m.\end{aligned}$$

Por tanto, para el modelo transformado se aplica el modelo de mínimos cuadrados clásico, y el estimador de $\boldsymbol{\beta}$ queda

$$\hat{\boldsymbol{\beta}} = (X_*^t X_*)^{-1} X_*^t \boldsymbol{y}_*.$$

Si reemplazamos por los valores originales, obtenemos

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (X^t C^{-1/2} C^{-1/2} X)^{-1} X^t C^{-1/2} C^{-1/2}\boldsymbol{y}\\ &= (X^t C^{-1} X)^{-1} X^t C^{-1}\boldsymbol{y}.\end{aligned}$$

El tratamiento anterior de mínimos cuadrados ponderados es equivalente a suponer que las medidas tomadas en $\boldsymbol{y}$ no son de la misma fiabilidad. Entonces se puede asignar un peso a los errores cometidos, dando más importancia a las medidas en las que tengamos más confianza. Si los pesos se denotan por $w_1^2, w_2^2, \ldots, w_m^2$, la suma ponderada de los errores será de la forma

$$w_1^2(y_1 - \hat{y}_1)^2 + w_2^2(y_2 - \hat{y}_2)^2 + \cdots + w_m^2(y_m - \hat{y}_m)^2.$$

Entonces estamos en la situación inicial con $\operatorname{var}\varepsilon_i = \sigma_i^2 = \frac{1}{w_i^2}$, es decir, la varianza del error es la inversa de los pesos (mayor peso implica menor varianza).

9.3.2. Mínimos cuadrados generalizado

Vamos a considerar ahora un problema de regresión más general. En caso anterior suponíamos que $\operatorname{var}\boldsymbol{\varepsilon} = \sigma^2 \operatorname{diag}(c_1^1, \ldots, c_m^2)$. Ahora consideramos $\operatorname{var}\boldsymbol{\varepsilon} = \sigma^2 C$, donde C es una matriz $m \times m$ definida positiva. Así, los errores aleatorios no solamente pueden tener varianzas distintas, sino que pueden estar correlacionados. Como en el caso ponderado, la solución pasa por transformar el problema original a uno clásico de mínimos cuadrados. Para ello, calculemos T una matriz $m \times m$ tal que $TT^t = C$ (factorización de Cholesky). Entonces $C^{-1} = (T^t)^{-1}T^{-1}$. Transformamos el modelo de regresión original en

$$\boldsymbol{y}_* = X_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*,$$

donde

$$\boldsymbol{y}_* = T^{-1}\boldsymbol{y}, X_* = T^{-1}X \text{ y } \varepsilon_* = T^{-1}\varepsilon.$$

Observemos que $E[\varepsilon_*] = T^{-1}E[\varepsilon] = 0$, y

$$\begin{aligned}\text{var}(\varepsilon_*) &= \text{var}(T^{-1}\varepsilon)\\ &= T^{-1}\,\text{var}(\varepsilon)(T^{-1})^t\\ &= T^{-1}\sigma^2 C(T^t)^{-1} = \sigma^2 I_m.\end{aligned}$$

Por tanto, el estimador $\hat{\boldsymbol{\beta}}_*$ de $\boldsymbol{\beta}$ en el problema de mínimos cuadrados generalizado es

$$\begin{aligned}\hat{\boldsymbol{\beta}}_* &= (X_*^t X_*)^{-1} X_*^t \boldsymbol{y}_*\\ &= (X^t(T^t)^{-1}T^{-1}X)^{-1}X^t(T^t)^{-1}T^{-1}\boldsymbol{y}\\ &= (X^t C^{-1}X)^{-1}X^t C^{-1}\boldsymbol{y}.\end{aligned}$$

9.4. Métodos numéricos de cálculo

Consideremos el sistema $A\boldsymbol{x} = \boldsymbol{b}$ con A de rango completo. Para resolverlo tenemos que encontrar la solución del sistema $A^*A\boldsymbol{x} = A^*\boldsymbol{b}$ (ecuaciones normales). Tenemos varios algoritmos.

9.4.1. Mediante Cholesky

Si A es de rango completo, entonces A^*A es hermitiana definida positiva, y el método estándar para encontrar la solución del sistema es la factorización de Cholesky: construimos una factorización $A^*A = R^*R$, con R triangular superior. El sistema se reduce a

$$R^*R\boldsymbol{x} = A^*\boldsymbol{b}$$

y el proceso es

1. Construye $M = A^*A$ y $\boldsymbol{c} = A^*\boldsymbol{b}$.
2. $M = R^*R$ mediante Cholesky.
3. Resuelve el sistema triangular inferior $R^*\boldsymbol{y} = \boldsymbol{c}$.
4. Resuelve el sistema triangular superior $R\boldsymbol{x} = \boldsymbol{y}$.

Si A es de orden $m \times n$, entonces la complejidad de cálculo es aproximadamente igual a $(mn^2 + 1/3n^3)$ [**?**, p.82].

Ejemplo 9.4.1. Vamos a ajustar el siguiente conjuntos de datos:

t_i	1000	1050	1060	1080	1110	1130
$\boldsymbol{y}_1$	6010	6153	6421	6399	6726	6701

Mostramos la ejecución de los comandos en MATLAB.

```
>> t = [1000 1050 1060 1080 1110 1130]';
>> y1 = [6010 6153 6421 6399 6726 6701]';
>> A = [ones(numel(t),1), t];
>> M = A'*A

M =

           6        6430
        6430     6901500
>> cond(M,2)

ans =

  7.4307e+008
```

Observemos que el número de condición de la matriz M es muy elevado. Veremos después cómo se puede rebajar.

```
>> RM = chol(M)

RM =

  1.0e+003 *

    0.0024    2.6250
         0    0.1034

>> betaC = RM'\c

betaC =

  1.0e+004 *

    1.5681
    0.0611
```

```
>> alphaC = RM\betaC

alphaC =

   71.6443
    5.9067
```

Veamos una forma de conseguir que la matriz M tenga un mejor número de condición. Una forma es la aplicación del escalado a las filas de la matriz, pero entonces perdemos el carácter simétrico. Se trata de centrar la variable de datos t mediante una transformación

$$z = \frac{t-\mu}{\sigma},$$

donde μ es la media de t, y σ su desviación estándar. El problema de ajuste se reduce a uno de la forma $y = \beta_0 + \beta_1 z$, cuya solución se calcula de la forma siguiente:

```
>> z = (t-mean(t))/std(t);
>> Am = [ones(numel(z)), z];
>> M2 = Am'*Am
M2 =

    6.0000   -0.0000
   -0.0000    5.0000

>> cond(M2,2)

ans =

    1.2000
```

El número de condición ha disminuido considerablemente, y las soluciones son más fiables. La razón de la mejora del número de condición es que en el primer caso, los valores de t_i se concentran en el intervalo $[1000, 1130]$, y las dos columnas de la matriz de coeficientes son aproximadamente proporcionales. En contraste, la variable z tiene una distribución en $[-1{,}5504, 1{,}2620]$, y las columnas de A_m no son tan parecidas.

```
      >> R2 = chol(M2);
      >> gamma = R2' \ (Am'*y1);
      >> >> beta = R2 \ gamma

beta =

  1.0e+003 *

    6.4017
    0.2730
```

9.4.2. Mediante QR reducida

Sea $A = \hat{Q}\hat{R}$ factorización QR *reducida* de la matriz A, con $\hat{Q}$ del mismo orden que A y $\hat{R}$ triangular superior cuadrada. Las ecuaciones normales quedan entonces

$$A^*A\boldsymbol{x} = A^*\boldsymbol{b} \Rightarrow \hat{R}^*\hat{Q}^*\hat{Q}\hat{R}\boldsymbol{x} = \hat{R}^*\hat{Q}^*\boldsymbol{b}$$

Como $\hat{Q}$ tiene columnas ortonormales, la matriz $\hat{Q}^*\hat{Q}$ es igual a la identidad de orden n. Además, $\hat{R}$ es invertible, luego obtenemos el sistema

$$\hat{R}\boldsymbol{x} = \hat{Q}^*\boldsymbol{b}.$$

El proceso queda entonces

1. $A = \hat{Q}\hat{R}$ descomposición QR reducida.
2. $\boldsymbol{c} = \hat{Q}^*\boldsymbol{b}$.
3. Resuelve el sistema triangular superior $\hat{R}\boldsymbol{x} = \boldsymbol{c}$.

La complejidad de cálculo es del orden de $2mn^2 - 2/3n^3$, si calculamos la descomposición QR por Householder [**?**, p.83].

Ejemplo 9.4.2. Vamos a ajustar el siguiente conjuntos de datos:

t_i	1000	1050	1060	1080	1110	1130
$\boldsymbol{y}_1$	6010	6153	6421	6399	6726	6701

Mostramos la ejecución de los comandos en MATLAB.

```
>> t = [1000 1050 1060 1080 1110 1130]';
>> y1 = [6010 6153 6421 6399 6726 6701]';
>> A = [ones(numel(t),1), t];
```

```
>> [QA,RA] = qr(A,0)

QA =

   -0.4082   -0.6934
   -0.4082   -0.2096
   -0.4082   -0.1129
   -0.4082    0.0806
   -0.4082    0.3709
   -0.4082    0.5644


RA =

  1.0e+003 *

   -0.0024   -2.6250
         0    0.1034
>> beta = QA'*y1

beta =

  1.0e+004 *

   -1.5681
    0.0611

>> alphaQR = RA\beta

alphaQR =

   71.6443
    5.9067
```

9.4.3. * Mediante SVD reducida

Consideremos la descomposición SVD *reducida* $A = \hat{U}\hat{\Sigma}V^*$, donde $\hat{U}$ es de columnas ortonormales del mismo orden que A, $\hat{\Sigma}$ es diagonal y V es unitaria.

Entonces las ecuaciones normales $A^*A\boldsymbol{x} = A^*\boldsymbol{b}$ implican

$$\Sigma V^*\boldsymbol{x} = \hat{U}^*\boldsymbol{b}$$

El proceso queda entonces

1. $A = \hat{U}\hat{\Sigma}V^*$ con U rectangular, Σ cuadrada.
2. $\boldsymbol{c} = \hat{U}^*\boldsymbol{b}$.
3. Resuelve el sistema diagonal $\hat{\Sigma}\boldsymbol{y} = \boldsymbol{c}$.
4. $\boldsymbol{x} = V\boldsymbol{y}$.

El coste es del orden de $2mn^2 + 11n^3$ [**?**, p.84].

Ejemplo 9.4.3. Vamos a ajustar el siguiente conjuntos de datos:

t_i	1000	1050	1060	1080	1110	1130
$\boldsymbol{y}_1$	6010	6153	6421	6399	6726	6701

Mostramos la ejecución de los comandos en MATLAB.

```
>> t = [1000 1050 1060 1080 1110 1130]';
>> y1 = [6010 6153 6421 6399 6726 6701]';
>> A = [ones(numel(t),1), t];
>> [UA,SA,VA]=svd(A,0)

UA =

    0.3807    0.7089
    0.3997    0.2255
    0.4035    0.1288
    0.4111   -0.0645
    0.4225   -0.3545
    0.4301   -0.5479


SA =

  1.0e+003 *

    2.6271         0
         0    0.0001
```

```
VA =

    0.0009    1.0000
    1.0000   -0.0009
>> beta = UA'*y1;
>> gamma = SA\beta

gamma =

    5.9735
   71.6388
alphaSVD = VA*gamma

alphaSVD =

   71.6443
    5.9067
```

9.4.4. Conclusiones

Si se busca velocidad, la elección es Cholesky. Sin embargo, la resolución del sistema no es siempre estable por los errores de redondeo. Por ese motivo, ha sido tradicional el uso de QR (ver [**?**, p. 245]), que es algo más barato que SVD. Sin embargo, si la matriz A tiene valores singulares próximos a cero (deficiencia de rango), el uso de SVD es el más aconsejable. Precisamente para los problemas en los que A tiene más columnas que filas (indeterminado) se usa el análogo de SVD ([**?**, p.263]).

En el documento "Comparación de mínimos cuadrados" se muestra una comparativa de los métodos anteriores sobre un caso concreto. También se explican cómo se resuelve este problema en diferentes herramientas, como MATLAB, MAPLE, SCILAB, OCTAVE, SAGE.

No hay que olvidar la técnica de centrado, muy habitual en la regresión entre dos variables. Si la matriz A es de la forma

$$A = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_n \end{pmatrix}, \boldsymbol{t} \text{ el vector con los datos experimentales,}$$

hacemos el cambio $y_i = \frac{1}{s}(t_i - E[\boldsymbol{y}])$, donde s es la desviación estándar, la muestral o la poblacional, según se quiera. Entonces se realiza el ajuste $\alpha_0 + \alpha_1 y_i = b_i$, y la nueva matriz es

$$B = \begin{pmatrix} 1 & y_1 \\ 1 & y_2 \\ \vdots & \\ 1 & y_n \end{pmatrix}, \text{ que verifica } B^t B = \begin{pmatrix} n & 0 \\ 0 & k \end{pmatrix}, k = n \text{ o } n-1.$$

El número de condición de esta matriz es muy próximo a 1.

9.5. * Colinealidad en la matriz de datos

Esta sección es complementaria a los métodos numéricos para la resolución del problema de mínimos cuadrados, y es un compendio de material diverso. Se trata de afrontar el caso en el que la matriz de coeficientes asociada al problema de mínimos cuadrados genera unas ecuaciones normales con un número de condición grande. Empecemos mostrando un escenario donde se presentan estos problemas.

Supongamos que estamos realizando un estudio médico sobre el efecto de ciertas fármacos sobre el nivel de azúcar en la sangre. Recogemos datos de cada paciente, numerados de $i = 1$ a m, guardando su nivel inicial de azúcar en sangre a_{i1}, su nivel final b_{i1}, la cantidad de fármaco administrado a_{i2}, y otras cantidades de tipo médico, como el peso diario, en un tratamiento de una semana (a_{i3} hasta a_{i9}). En total, hay $n < m$ valores médicos medidos para cada paciente. Nuestro objetivo es predecir b_i dados a_{i1} hasta a_{in}, y formulamos el problema como un ajuste por mínimos cuadrados $\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|_2$. Usaremos $\boldsymbol{x}$ para predecir el valor de azúcar en sangre b_j de un futuro paciente j mediante el resultado del ajuste $\sum_{k=1}^{n} a_{jk}x_k$.

Como el peso de una persona no cambia significativamente de un día para otro, es probable que las columnas 3 a la 9 de la matriz A, que contienen los pesos, sean muy similares. Por simplificar el razonamiento, supongamos que son iguales. Esto significa que la matriz A es deficiente de rango, y el vector $\boldsymbol{u}_0 = \boldsymbol{e}_3 - \boldsymbol{e}_4$ es un vector de null(A). Si $\boldsymbol{u}$ es una solución mínimo cuadrática de norma mínima, entonces $\boldsymbol{u} + \beta\boldsymbol{u}_0$ también es solución mínimo cuadrática, para cualquier escalar β, como pueden ser $\beta = 0$ o $\beta = 10^6$. ¿Existe alguna razón para preferir una a otra? El valor 10^6 no parece una buena elección, ya que el futuro paciente j, que ganan medio kilo entre los días 1 y 2, tendrá esa diferencia de 0,5 kilogramos multiplicada por 10^6 en la predicción $\sum_{k=1}^{n} a_{jk}x_k$ del nivel de azúcar final. Es más razonable tomar $\beta = 0$, que corresponde a la solución de norma mínima $\boldsymbol{u}$.

9.5.1. Uso de la SVD en análisis de regresión

Lo que sigue es un extracto del artículo 'Use of Singular Value Decomposition in Regression Analysis', por John Mandel, The American Statiscian, Feb 1982, vol. 36, n. 1, 15–24.

Así como la regresión lineal múltiple por mínimos cuadrados se ha usado durante mucho tiempo como una importante técnica estadística para *ajustar ecuaciones a datos*, las implicaciones completas, limitaciones y problemas inherentes asociados han sido tratados en artículos y libros únicamente de forma reciente. Además de aclarar estas cuestiones, una gran parte del trabajo ha

proporcionado modificaciones de la técnica con el objeto de incrementar su fiabilidad como una herramienta de análisis de datos.

Sin duda, la mayor fuente de dificultades en el uso de mínimos cuadrados es la existencia de *colinealidad* en muchos conjuntos de datos, y la mayoría de las modificaciones al método de mínimos cuadrados ordinario es un intento de ocuparse del problema de la colinealidad. Entre estas modificaciones se pueden citar la regresión de componentes principales (Draper,Smith, 1981; Hocking, Speed, Lynn, 1976), regresión de raíces latentes (Webster, Gunst, Mason, 1974), contracción (Hocking, Speed, Lynn, 1976; Stein, 1960), 'ridge regression' (Chatterjee, Price, 1977; Draper,Smith, 1981; Hocking, Speed, Lynn, 1976; Hoerl, Kennard, 1970; Marquardt, 1970; Marquardt, Snee, 1973), y otras variantes de estas técnicas.

Aquí no pretendemos discutirlas todas ellas, o comparar sus méritos relativos. El propósito es presentar la naturaleza de estos problemas a través de una cuidadosa explicación de la matemática adecuada y los aspectos conceptuales. Es casi indispensable, para alcanzar este objetivo, usar la notación matricial y recurrir al método de las componentes principales o técnicas relacionadas. Usaremos la descomposición en valores singulares (SVD) de la matriz de diseño, una técnica que tiene mucho que ver con el método de las componentes principales, para aclarar el problema de la colinealidad.

Haremos una exposición general, sin garantía de completitud en el tratamiento. Para una discusión más amplia y avanzada se puede consultar (Belsley, Kuh, Welsch, 1980), o la nueva edición de (Draper, Smith, 1981).

El modelo

Suponemos el modelo lineal estándar

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}, \tag{9.5.1}$$

donde $\boldsymbol{y}, \boldsymbol{e}$ son vectores de N elementos, $X = (x_{ij})$ es una matrix de orden $N \times p$, y $\boldsymbol{\beta}$ es un vector de p elementos. La matrix X es dada, y el vector $\boldsymbol{y}$ contiene las medidas y_i. Los errores e_i se suponen no correlacionados, de media cero y varianza σ^2, que es desconocida. Algunas de las ideas generales las ilustraremos con los datos de la tabla 9.10, en la que $N = 8, p = 3$.

En este caso hay tres variables regresoras x_1, x_2, x_3, de las cuales la primera es igual a 1 para todo i. La ecuación de regresión queda de la forma

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i,$$

con un término independiente β_1.

Punto	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{y}$
1	1	16,85	1,46	41,38
2	1	24,81	−4,61	31,01
3	1	18,85	−0,21	37,41
4	1	12,63	4,93	50,05
5	1	21,38	−1,36	39,17
6	1	18,78	−0,08	38,86
7	1	15,58	2,98	46,14
8	1	16,30	1,73	44,17

Figura 9.10: Tabla de datos A

En muchos casos (ver Schott), se realiza previamente una estandarización de todas las variables regresoras, excepto el término independiente. La estandarización del regresor x_j consiste en reemplazarlo en la ecuación de regresión por

$$x_j = \bar{x}_j + s_j t_j,$$

donde $\bar{x}_j$ es la media, y s_j es la desviación estándar de los elementos x_{ij} en la columna x_j. La regresión lineal es ahora entre $\boldsymbol{y}$ y los t_j, donde tenemos que $\bar{t}_j = 0$ (variables centradas), y $s_{t_j} = 1$ (variables escaladas). Los usos y utilidad del centrado y escalado se discuten con detalle en (Draper, Smith, 1981). Por simplicidad en la presentación omitimos este paso en esta exposición. Hemos visto un caso en el ejemplo 9.4.1.

El objeto del análisis de regresión es estimar los coeficientes β_j, $j = 1,\ldots,p$, así como σ^2, predecir el valor de $\boldsymbol{y}$ para cualquier valor de las variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$, y estimar el error del valor predicho $\hat{\boldsymbol{y}}$. Para evitar confusión, un conjunto de valores $(x_1, x_2, \ldots, x_p)$ para los cuales se calcula un valor de $\boldsymbol{y}$ se denominará un punto en el espacio X, o simplemente un punto, en lugar de un vector.

SVD de la matriz X

Dada una matriz X de orden $N \times p$, es posible expresar cada elemento x_{ij} de X como

$$x_{ij} = \sigma_1 u_{1i} v_{1j} + \sigma_2 u_{2i} v_{2j} + \ldots + \sigma_r u_{ri} v_{rj} = \sum_{k=1}^{r} \sigma_k u_{ki} v_{kj}, \tag{9.5.2}$$

donde $\sigma_1 \geq \sigma 2 \geq \ldots \geq \sigma_r$. Esto se conoce como la descomposición en valores singulares (SVD) compacta de X. El número de términos r es el rango de la matriz X, por lo que $r \leq \text{mín}\{N, p\}$.

Supondremos siempre que $N \geq p$ (sistema sobredeterminado), y entonces $r \leq p$. Los r vectores $\boldsymbol{u}_k$ forman un sistema ortonormal, al igual que los vectores $\boldsymbol{v}_k$. La notación matricial clásica es

$$X = U\Sigma V^t,$$

con U de orden $N \times r$, Σ diagonal de orden r, y V^t de orden $r \times p$. Las columnas de la matriz U son los vectores $\boldsymbol{u}_k$, y las columnas de la matriz V son los vectores $\boldsymbol{v}_k$. Se tiene entonces que $U^tU = I_r = V^tV$. Los números $\sigma_1, \ldots, \sigma_r$ son las raíces cuadradas de los autovalores no nulos de X^tX (o de XX^t), y son los valores singulares. Las columnas de U son los autovectores de XX^t, y las columnas de V los autovectores de X^tX.

En el caso de la matriz de datos de la tabla 9.10, podemos hacer los siguientes cálculos en MATLAB.

```
>> X =  [1, 16.85, 1.46; ...
      1, 24.81, -4.61; ...
      1, 18.85, -0.21; ...
      1, 12.63, 4.93; ...
      1, 21.38, -1.36; ...
      1, 18.78, -0.08; ...
      1, 15.58, 2.98; ...
      1, 16.30, 1.73];

>> [U,S,V] = svd(X,0)

U =

   -0.3226   -0.1761   -0.1938
   -0.4739    0.6035   -0.0499
   -0.3606    0.0382   -0.3337
   -0.2424   -0.6214    0.0362
   -0.4087    0.1869    0.6592
   -0.3593    0.0216   -0.2418
   -0.2985   -0.3705    0.4536
   -0.3121   -0.2110   -0.3853


S =

   52.3478         0         0
         0    7.8539         0
```

```
         0         0    0.0557
```

```
V =

   -0.0531   -0.0673   -0.9963
   -0.9986    0.0084    0.0526
   -0.0048   -0.9977    0.0677
```

En este caso, $r = 3$, es decir, $r = p$. Esto se conoce como el caso de rango completo. Cada elemento de X se puede recalcular a partir de la fórmula 9.5.2. Por ejemplo, el elemento $x_{43} = 4{,}93$ es igual a

$$\begin{aligned}(-0{,}2424) \times 52{,}3478 \times (-0{,}0048) \\ + (-0{,}6214) \times 7{,}8539 \times (-0{,}9977) \\ + 0{,}0362 \times 0{,}0557 \times 0{,}0677. \quad (9.5.3)\end{aligned}$$

Interpretación geométrica de SVD

Para simplificar la explicación, consideraremos un ejemplo con únicamente dos variables regresoras x_1, x_2, dados por la matriz

$$X = \begin{pmatrix} 1{,}3 & 1{,}2 \\ 4{,}2 & 2{,}8 \\ 6{,}3 & 7{,}4 \\ 8{,}0 & 7{,}1 \\ 9{,}4 & 8{,}2 \end{pmatrix}.$$

Las columnas de X representan a los vectores x_1, x_2. Cada fila de X se puede interpretar como un punto en el espacio de dos dimensiones, con coordenadas (x_1, x_2). Podemos considerar X representada por 5 puntos en el plano. Calculemos su SVD compacta.

```
>> X2 = [ 1.3, 1.2; 4.2, 2.8; 6.3, 7.4; 8.0, 7.1; 9.4, 8.2];
>> [U2,S2,V2] = svd(X2,0)
```

```
U2 =

    0.0895   -0.0021
    0.2521   -0.5460
    0.4881    0.7830
```

```
    0.5409   -0.1555
    0.6306   -0.2541


S2 =

   19.7718         0
         0    1.4654


V2 =

    0.7336   -0.6796
    0.6796    0.7336
```

Las columnas de V, etiquetadas $\boldsymbol{v}_1, \boldsymbol{v}_2$, también representan un vector o un punto. Como vectores, al ser ortonormales, los podemos considerar como un nuevo sistema de referencia en el plano. Si calculamos las coordenadas de los puntos de X respecto a este nuevo sistema, las nuevas coordenadas serán iguales a $\sigma_1 \boldsymbol{u}_1$ y $\sigma_2 \boldsymbol{u}_2$. Por ejemplo, las nuevas coordenadas de (4,2,2,8) serán (19,7718× 0,2521, 1,4654×(−0,5460) = (4,9845, −0,8001). Los tamaños relativos de las coordenadas no son accidentales. Las proyecciones de los puntos sobre el eje $\boldsymbol{v}_1$ cubren un mayor rango que las proyecciones sobre el eje $\boldsymbol{v}_2$. En otras palabras, los cinco puntos de la matriz de diseño descansan fundamentalmente sobre el eje $\boldsymbol{v}_1$, y menos sobre el eje $\boldsymbol{v}_2$. Notemos que si tuviéramos $\sigma_2 = 0$, entonces las coordenadas sobre el eje $\boldsymbol{v}_2$ serían 0. En tal caso, los cinco puntos estarían sobre la recta determinada por $\boldsymbol{v}_1$. Vemos entonces que el propósito de SVD es reorientar los ejes coordenados de tal forma que siguen los más aproximadamente posible la forma dibujada por los puntos de la matriz X. La SVD nos ayuda a entender la estructura de la matriz X.

9.5.2. Regresión de componentes principales

El objetivo principal de esta sección se puede ahora enunciar en términos más precisos. Con la ayuda de la SVD, vamos a ver las ventajas de reemplazar X por su SVD en el cálculo de la regresión de $\boldsymbol{y}$ sobre X. Este procedimiento se denomina regresión de componentes principales. Veremos que mientras esta técnica se puede usar en cualquier regresión dada por el modelo estándar, es particularmente interesante en los casos de *colinealidad o* casi colinealidad. Estos términos se explicarán más adelante.

Cambiamos en la ecuación (9.5.1) la matriz X por su SVD:

$$\boldsymbol{y} = U\Sigma V^t\boldsymbol{\beta} + \boldsymbol{e}. \tag{9.5.4}$$

Escrita de esta forma, hablaremos del modelo de regresión de componentes principales.

La ecuación anterior la podemos escribir como

$$\boldsymbol{y} = U(\Sigma V^t\boldsymbol{\beta}) + \boldsymbol{e},$$

donde $\Sigma V^t\boldsymbol{\beta}$ es un vector de orden r, que llamaremos $\boldsymbol{\alpha}$. Entonces la ecuación (9.5.4) queda

$$\boldsymbol{y} = U\boldsymbol{\alpha} + \boldsymbol{e}.$$

El vector $\boldsymbol{y}$ y la matriz U son conocidas. La solución mínimo cuadrática se calcula de la forma habitual:

$$\hat{\boldsymbol{\alpha}} = (U^tU)^{-1}U^t\boldsymbol{y} = U^t\boldsymbol{y}.$$

Por ejemplo, con los datos de la tabla 9.10, nos quedaría

```
>>  y = [ 41.38 31.01 37.41 50.05 39.17 38.86 46.14 44.47];
>>  alphag = U'*y

alphag =

  -111.2858
   -36.5653
    -0.0185
```

Sigamos con el estudio teórico. Se tiene que

$$\hat{\boldsymbol{\alpha}} = U^t(U\boldsymbol{\alpha} + \boldsymbol{e}) = \boldsymbol{\alpha} + U^t\boldsymbol{e},$$

o bien $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = U^t\boldsymbol{e}$. Entonces $E(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = 0$, de donde $E(\hat{\boldsymbol{\alpha}}) = E(\boldsymbol{\alpha})$. Así, $\hat{\boldsymbol{\alpha}}$ es insesgado. Además,

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\alpha}}) &= \text{var}(\boldsymbol{\alpha} + U^t\boldsymbol{e}) \\ &= \text{var}(U^t\boldsymbol{e}) = U^t\,\text{var}(\boldsymbol{e})U = \sigma^2 I. \end{aligned}$$

Por tanto, las componentes $\hat{\alpha}_j$ son mutuamente no correlacionadas. Observemos que el número de elementos de $\boldsymbol{\beta}$ es p, y el de α es r, que puede ser menor que p.

Tenemos que $\hat{\boldsymbol{\beta}} = X^+\boldsymbol{y} = V\Sigma^{-1}U^t\boldsymbol{y} = V\Sigma^{-1}\hat{\boldsymbol{\alpha}}$. Recordemos que V no es cuadrada, pero sus columnas forman un sistema ortonormal, es decir, $V^tV = I$. Entonces $V^t\hat{\boldsymbol{\beta}} = \Sigma^{-1}\hat{\boldsymbol{\alpha}}$, y de aquí

$$\Sigma V_1^t\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\alpha}}. \tag{9.5.5}$$

Por ejemplo, con los datos de la tabla 9.10 y la SVD calculada, nos queda

$$\hat{\beta}_1 = -0{,}0531\frac{\hat{\alpha}_1}{52{,}3478} + (-0{,}0673)\frac{\hat{\alpha}_2}{7{,}8539} + (-0{,}9963)\frac{\hat{\alpha_3}}{0{,}0557}.$$

En general, la ecuación es de la forma

$$\hat{\beta}_j = \sum_{k=1}^{p} v_{jk}\frac{\hat{\alpha}_k}{\sigma_k}.$$

Como las componentes de $\hat{\boldsymbol{\alpha}}$ no están correlacionadas, y tienen varianza igual a σ^2, se tiene que

$$\text{var}(\hat{\beta}_j) = \sigma^2\sum_{k=1}^{p}\frac{v_{jk}^2}{\sigma_k^2}.$$

Con los datos que estamos tomando, obtenemos

$$\text{var}(\hat{\beta}_1) = \sigma^2\left[\left(-\frac{0{,}0531}{52{,}3478}\right)^2 + \left(-\frac{0{,}0673}{7{,}8539}\right)^2 + \left(-\frac{0{,}9963}{0{,}0557}\right)^2\right].$$

El numerador de cada término son los cuadrados de los elementos de la primera fila de la matriz V, por lo que están entre 0 y 1. Sin embargo, los denominadores son los cuadrados de los valores singulares. En el ejemplo, σ_3 es mucho más pequeño que σ_1 y σ_2, por lo que el tercer sumando es el que más contribuye a la varianza total. Lo mismo ocurre con $\text{var}(\hat{\beta}_2)$ y $\text{var}(\hat{\beta}_3)$. La razón de esta situación no deseable es el pequeño valor de σ_3. t'Vemos así que el uso de la SVD nos permite identificar la causa de varianzas grandes para algunos coeficientes. En cierta forma, el valor de σ_3 se puede considerar, a efectos prácticos, igual a cero. Pero un valor singular nulo tiene consecuencias importantes en la interpretación de la regresión.

9.5.3. Efectos en la regresión de la colinealidad

Un valor singular nulo implica un autovalor cero en la matriz X^tX. Como $\text{rango}(X^tX) = \text{rango}(X)$, esto significa que existe una relación lineal entre las columnas de la matriz X. Por tanto, un valor singular próximo a cero implica que existe casi una relación lineal entre las columnas de X.

Vamos a ver lo que ocurre en el ejemplo anterior de manera gráfica. Como la variable $\boldsymbol{x}_1$ tiene el valor constante 1, dibujaremos la distribución de puntos definida por $\boldsymbol{x}_2$ y $\boldsymbol{x}_3$. Suponemos los datos ya cargados en la matriz X.

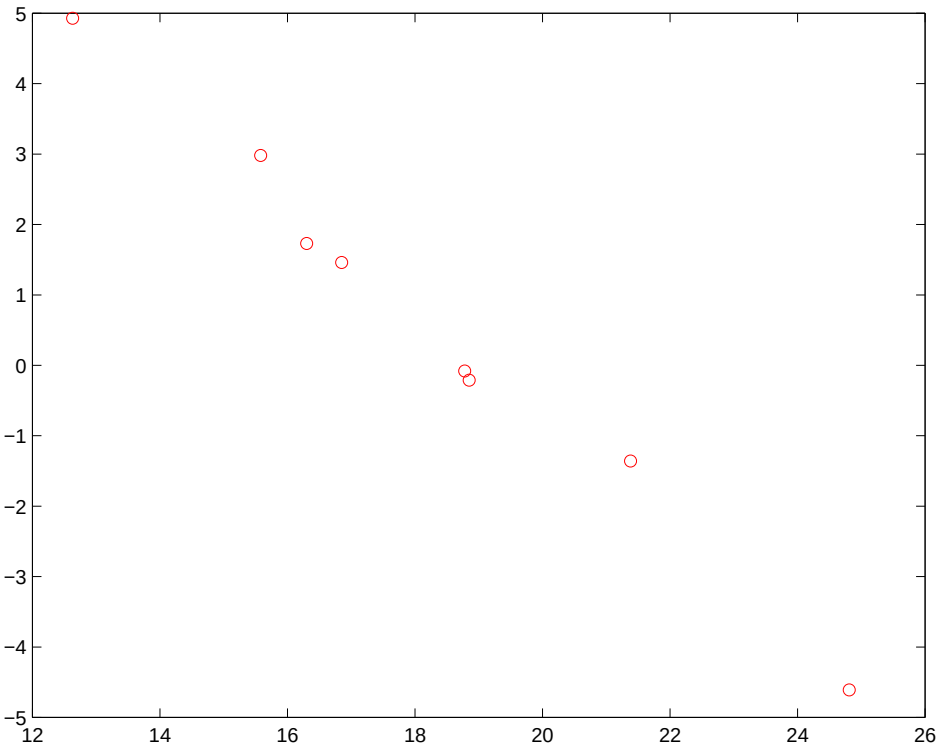

Figura 9.11: Casi colinealidad entre $\boldsymbol{x}_2$ y $\boldsymbol{x}_3$.

```
>> plot(X(:,2),X(:,3),'ro')
```

El gráfico que aparece lo vemos en la figura 9.11 Podemos calcular la recta de regresión asociada a estos datos. Planteamos el problema de encontrar una relación lineal del tipo

$$\boldsymbol{x}_3 = \gamma_1 \boldsymbol{x}_1 + \gamma_2 \boldsymbol{x}_2, \tag{9.5.6}$$

que en forma matricial es

$$\begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \boldsymbol{x}_3.$$

El ajuste por mínimos cuadrados lo podemos calcular con MATLAB mediante el comando

.

```
>> A = [X(:,1),X(:,2)];
>> A\X(:,3)

ans =

   14.5594
   -0.7689
```

Entonces $\gamma_1 = 14{,}5594$ y $\gamma_2 = -0{,}7689$, esto es,

$$x_3 = 14{,}5594 x_1 - 0{,}7689 x_2. \tag{9.5.7}$$

Punto	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{y}$
1	1	16,85	2,3625	41,38
2	1	24,81	−3,6075	31,01
3	1	18,85	0,8625	37,41
4	1	12,63	5,5275	50,05
5	1	21,38	−1,0350	39,17
6	1	18,78	0,9150	38,86
7	1	15,58	3,3150	46,14
8	1	16,30	2,7750	44,47

Figura 9.12: Tabla de datos B. $\boldsymbol{x}_3 = 15\boldsymbol{x}_1 - 0{,}75\boldsymbol{x}_2$

Se pueden obtener análogas relaciones si hacemos las regresiones de x_2 respecto a x_1, x_3, y de x_1 respecto a x_2, x_3. Procediendo como antes se obtiene

$$x_2 = 18{,}9253x_1 - 1{,}2856x_3, x_1 = 0{,}0528x_2 + 0{,}0679x_3$$

respectivamente. Si expresamos en ambos casos x_3 en función de x_1 y x_2, nos queda

$$x_3 = 14{,}7209x_1 - 0{,}7778x_2, \tag{9.5.8}$$

y

$$x_3 = 14{,}7200x_1 - 0{,}7775x_2. \tag{9.5.9}$$

El problema que pretendemos resolver es descubrir estas relaciones sin gráficos. Para ello, tomaremos en primer lugar un conjunto de datos *preparados*, con una relación lineal exacta, y de manera analítica la encontraremos. La tabla de partida aparece en 9.12.

La SVD de la matriz X es la siguiente:

```
>> X1 = [1, 16.85, 2.3625; 1, 24.81, -3.6075; ...
 1, 18.85, 0.8625; 1, 12.63, 5.5275; ...
 1, 21.38, -1.0350; 1, 18.78, 0.9150; ...
 1, 15.58, 3.3150; 1, 16.30, 2.7750];

>> [U1, S1, V1] = svd(X1,0)


U1 =

   -0.3239   -0.1931    0.8253
   -0.4699    0.5981   -0.1172
```

```
   -0.3606    0.0057   -0.3217
   -0.2465   -0.6126   -0.1688
   -0.4070    0.2572    0.2668
   -0.3593   -0.0013   -0.2596
   -0.3006   -0.3194   -0.0445
   -0.3138   -0.2478   -0.1803


S1 =

   52.4063         0         0
         0    8.0365         0
         0         0    0.0000


V1 =

   -0.0531   -0.0639    0.9965
   -0.9974    0.0514   -0.0498
   -0.0480   -0.9966   -0.0664
```

En la forma que estamos usando de SVD, nuestra matriz U está formada por las dos primeras columnas de U_1, Σ es la matriz diagonal de orden 2 con los dos valores singulares no nulos, y V por las dos primeras columnas de V_1. Hay que hacer notar que las matrices U y V no son únicas.

```
>> U = U1(:,1:2)
>> S = S1(1:2,1:2)
>> V = V1(:,1:2)
```

El rango de X es 2, y $p = 3$. A partir de $\hat{\boldsymbol{\alpha}} = U^t \boldsymbol{y}$, nos queda

$$\hat{\boldsymbol{\alpha}} = U^t \boldsymbol{y} = \begin{pmatrix} -111{,}5248 \\ -35{,}6285 \end{pmatrix}.$$

Aunque $\hat{\boldsymbol{\alpha}}$ tiene solamente dos componentes, para $\hat{\boldsymbol{\beta}}$ hay tres componentes, y están relacionados por

$$\Sigma V^t \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\alpha}}.$$

En el ejemplo,

```
>> S * V'
```

```
ans =

   -2.7815  -52.2719   -2.5180
   -0.5133    0.4131   -8.0094
```

luego

$$\begin{pmatrix} -2{,}7815 & -52{,}2719 & -2{,}5180 \\ -0{,}5133 & 0{,}4131 & -8{,}0094 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -111{,}5248 \\ -35{,}6285 \end{pmatrix}.$$

Podemos expresar cualquiera de las tres componentes de $\hat{\boldsymbol{\beta}}$ en función de las otras dos, y esto implica que no hay solución única para $\hat{\boldsymbol{\beta}}$.

Para tratar el caso general, notemos la matriz ΣV^t, de orden $r \times p$, por Z, y dividimos Z en la forma

$$Z = \begin{pmatrix} Z_A & Z_B \end{pmatrix},$$

donde Z_A es de orden $r \times r$, y Z_B de orden $r \times (p-r)$. Hacemos lo mismo con el vector $\hat{\boldsymbol{\beta}}$, y nos queda

$$\begin{pmatrix} Z_A & Z_B \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}}_A \\ \hat{\boldsymbol{\beta}}_B \end{pmatrix} = \hat{\boldsymbol{\alpha}},$$

con $\hat{\boldsymbol{\beta}}_A$ vector de orden r, y $\hat{\boldsymbol{\beta}}_B$ de orden $(p-r)$. Desarrollamos y nos queda

$$Z_A\hat{\boldsymbol{\beta}}_A + Z_B\hat{\boldsymbol{\beta}}_B = \hat{\boldsymbol{\alpha}}.$$

Para los datos de nuestro ejemplo, esta ecuación queda

$$\begin{pmatrix} -2{,}7815 & -52{,}2719 \\ -0{,}5133 & 0{,}4131 \end{pmatrix}\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} + \begin{pmatrix} -2{,}5180 \\ -8{,}0094 \end{pmatrix}\begin{pmatrix} \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -111{,}5248 \\ -35{,}6285 \end{pmatrix}.$$

Podemos suponer que Z_A es no singular; si no, bastaría una reordenación de la variables $\hat{\beta}_i$. Entonces

$$\hat{\boldsymbol{\beta}}_A + Z_A^{-1}Z_B\hat{\boldsymbol{\beta}}_B = Z_A^{-1}\hat{\boldsymbol{\alpha}}. \tag{9.5.10}$$

La ecuación (9.5.10) muestra que, una vez se elija un valor de $\hat{\boldsymbol{\beta}}_B$, el vector $\hat{\boldsymbol{\beta}}_A$ está unívocamente determinado. En nuestro ejemplo, $\hat{\beta}_1$ y $\hat{\beta}_2$ se calculan fijando el valor de $\hat{\beta}_3$.

9.5.4. Predicción en el caso de colinealidad

Consideremos ahora un nuevo punto x para el que deseamos estimar $\hat{\boldsymbol{y}}$. Se verifica que $\hat{\boldsymbol{y}} = x\hat{\boldsymbol{\beta}}$. Si introducimos las particiones como en el apartado anterior,

$$\hat{\boldsymbol{y}} = \begin{pmatrix} x_A & x_B \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\beta}}_A \\ \hat{\boldsymbol{\beta}}_B \end{pmatrix},$$

con x_A de dimensión $1 \times r$, x_B de dimensión $1 \times (p-r)$. Entonces

$$\begin{aligned} \hat{\boldsymbol{y}} &= x_A\hat{\boldsymbol{\beta}}_A + x_B\hat{\boldsymbol{\beta}}_B && (9.5.11)\\ &= x_A(Z_A^{-1}\hat{\boldsymbol{\alpha}} - Z_A^{-1}Z_B\hat{\boldsymbol{\beta}}_B) + x_B\hat{\boldsymbol{\beta}}_B && (9.5.12)\\ &= x_A(Z_A^{-1}\hat{\boldsymbol{\alpha}}) + (x_B - x_AZ_A^{-1}Z_B)\hat{\boldsymbol{\beta}}_B. && (9.5.13)\end{aligned}$$

Recordemos que $\hat{\boldsymbol{\beta}}_B$ puede tomar cualquier valor. Para que la ecuación anterior tenga sentido, el valor de $\hat{\boldsymbol{y}}$ debe permanecer constante para cualquier valor arbitrario de $\hat{\boldsymbol{\beta}}$. Entonces

$$x_B - x_AZ_A^{-1}Z_B = 0 \qquad (9.5.14)$$

Recordemos que $Z = \Sigma V^t$. Partimos $V^t = \begin{pmatrix} V_A^t & V_B^t \end{pmatrix}$, con V_A^t de orden $r \times r$ y V_B^t de orden $r \times (p-r)$. Entonces

$$Z = \begin{pmatrix} Z_A & Z_B \end{pmatrix} = \Sigma\begin{pmatrix} V_A^t & V_B^t \end{pmatrix} = \begin{pmatrix} \Sigma V_A^t & \Sigma V_B^t \end{pmatrix}.$$

De aquí, $Z_A^{-1}Z_B = (V_A^t)^{-1}\Sigma^{-1}\Sigma V_B^t = (V_A^t)^{-1}V_B^t$. Entonces la ecuación (9.5.14) se puede escribir como

$$x_B = x_A\begin{pmatrix} (V_A^t)^{-1} & V_B^t \end{pmatrix}.$$

Si se verifica la ecuación (9.5.14), la solución es

$$\hat{\boldsymbol{y}} = x_A(Z_A^{-1}\hat{\boldsymbol{\alpha}}). \qquad (9.5.15)$$

Es importante resaltar que la ecuación (9.5.11) da lugar a dos relaciones: la condición (9.5.14) y la solución (9.5.15). Y la solución solamente es válida cuando la condición se cumple.

Para nuestro ejemplo,

```
>> Z = S * V'

Z =

   -2.7815  -52.2719   -2.5180
   -0.5133    0.4131   -8.0094

>> ZA = Z(1:2,1:2);
>> ZB = Z(:,3);
>> inv(ZA) * ZB

ans =

   15.0000
   -0.7500
```

La condición queda como

$$x_3 = 15x_1 - 0{,}75x_2, \tag{9.5.16}$$

y la solución, sujeta a la condición (9.5.16), es

```
>> inv(ZA) * alphag

ans =

   68.2063
   -1.4958
```

Esto es,

$$\hat{y} = 68{,}2063x_1 - 1{,}4958x_2.$$

Por ejemplo, podemos obtener una estimación para $x = \begin{pmatrix} 1 & 15 & 3{,}75 \end{pmatrix}$, pues verifica la restricción (9.5.16). En tal caso, $\hat{y} = 45{,}7693$. Sin embargo, no es posible obtener una estimación para $x = \begin{pmatrix} 1 & 20 & 3{,}75 \end{pmatrix}$.

En resumen, si $r < p$ no existe una solución única $\hat{\boldsymbol{y}}$ para cualquier punto $\boldsymbol{x}$, excepto en el caso en el que este $\boldsymbol{x}$ verifique la condición de colinealidad que se obtiene del valor singular nulo.

9.5.5. Tratamiento de la casi colinealidad

Volvemos a los datos de la tabla 9.10, en los que la matriz X es de rango completo, pero uno de los valores singulares es próximo a cero. En principio, es posible obtener una estimación $\hat{\boldsymbol{y}}$ para cualquier punto x. Sin embargo, a causa de la mala elección de los puntos que definen a la matriz X, tal estimación no será muy fiable lejos del hiperplano definido por la regresión entre las columnas de X.

Consideremos un nuevo punto x de p componentes. Si expresamos sus coordenadas respecto al sistema ortonormal definido por la matriz U, tenemos la relación

$$x = u\Sigma V^t, \tag{9.5.17}$$

donde u también es un punto de p componentes. En el caso de rango completo, V es una matriz ortogonal, de donde

$$u = xV\Sigma^{-1}. \tag{9.5.18}$$

El valor previsto para $\hat{\boldsymbol{y}}$, en función de u, es

$$\hat{\boldsymbol{y}} = u\hat{\boldsymbol{\alpha}},$$

de donde

$$\text{var}(\hat{\boldsymbol{y}}) = \sigma^2 \sum_{i=1}^{p} u_j^2.$$

Esta ecuación indica que basta una componente de u grande para que la varianza de $\hat{\boldsymbol{y}}$ crezca sensiblemente. Veamos bajo qué circunstancias esto ocurre.

De la SVD $X = U\Sigma V^t$, escribimos $U\Sigma = XV$. Dividimos la matriz Σ en dos bloques Σ_A y Σ_B, tales que el segundo contiene los valores singulares considerablemente menores que los del primer bloque. En nuestro ejemplo numérico, podemos hacer

$$\Sigma_A = \begin{pmatrix} 52{,}3478 & \\ & 7{,}8539 \end{pmatrix}, \Sigma_B = \begin{pmatrix} 0{,}0557 \end{pmatrix}.$$

En este caso, Σ_B contiene un solo valor, pero contendrá, en general, l valores. Entonces Σ_A es una matriz diagonal de $p-l=t$ valores. Hagamos una partición de $V = \begin{pmatrix} V_A & V_B \end{pmatrix}$, donde V_A e de orden $p \times t$ y V_B de orden $p \times l$. Así,

$$\left(U \begin{pmatrix} \Sigma_A \\ 0 \end{pmatrix} U \begin{pmatrix} 0 \\ \Sigma_B \end{pmatrix} \right) = \left(\begin{array}{c|c} XV_A & XV_B \end{array} \right).$$

Como los valores de Σ_B son pequeños, las columnas representadas por $U\begin{pmatrix} 0 \\ \Sigma_B \end{pmatrix}$ contienen elementos muy pequeños. Recordemos que las columnas de U forman un sistema ortonormal, por lo que todos sus elementos, en módulo, son menores que 1. En consecuencia, lo mismo ocurre para XV_B. Lo escribimos como $XV_B \approx 0$. Esta ecuación representa l ecuaciones lineales. En nuestro ejemplo,

$$XV_B = X \begin{pmatrix} -0{,}9963 \\ 0{,}0526 \\ 0{,}0677 \end{pmatrix},$$

lo que da para cada fila de X una expresión de la forma

$$-0{,}9963x_1 + 0{,}0526x_2 + 0{,}0677x_3 \approx 0.$$

Si la comparamos con la ecuación (9.5.8), se obtiene algo similar tras dividir por el coeficiente de x_3.

$$\frac{1}{0{,}0677}(-0{,}9963x_1 + 0{,}0526x_2 + 0{,}0677x_3) = -14{,}7192x_1 + 0{,}7774x_2 + x_3 \approx 0.$$

Volvamos ahora al problema de predecir $\hat{\boldsymbol{y}}$ para un nuevo punto x. Usamos la ecuación (9.5.17) para escribir

$$u = \begin{pmatrix} xV_A & xV_B \end{pmatrix} \Sigma^{-1} = \begin{pmatrix} xV_A & xV_B \end{pmatrix} \begin{pmatrix} \Sigma_A^{-1} & 0 \\ 0 & \Sigma_B^{-1} \end{pmatrix}.$$

Los últimos l elementos de u se calculan de $xV_B\Sigma_B^{-1}$. Como antes, los elementos de V_B están acotados, en módulo, por 1, y los elementos de Σ_B^{-1} son grandes. Entonces encontraremos en u componentes de módulo grande, que incrementan fuertemente la varianza de $\hat{\boldsymbol{y}}$, a menos que $xV_B \approx 0$, esto es, a menos que el nuevo punto satisfaga la condición de casi linealidad de la matriz X. Cuanto más lejos esté de las relaciones lineales $XV_B = 0$, peor será la precisión del valor predicho. Este es el corazón del problema de la colinealidad, cuando se mira desde el punto de vista de la estimación con intenciones de predicción.

Si las condiciones lineales se verifican exactamente, tenemos

$$xV_B = 0, \text{ y } u = xV_A(\Sigma_A^{-1}),$$

que implica que las últimas l componentes de u son cero y

$$\hat{\boldsymbol{y}} = u_1\hat{\alpha}_1 + \ldots + u_t\hat{\alpha}_t.$$

En resumen: el caso de casi colinealidad se caracteriza por uno o varios valores singulares próximos a cero. El rango de X es p, pero las predicciones se pueden hacer únicamente para los puntos próximos a las relaciones lineales $xV_B = 0$. Cuando se verifica esta relación, el valor predicho es $\hat{\boldsymbol{y}} = xV_A\Sigma_A^{-1}\hat{\boldsymbol{\alpha}}$.

Capítulo 10

Inversas generalizadas y aplicaciones

10.1. Inversa generalizada de Moore-Penrose

Sea A una matriz de orden $m \times n$. Una ***inversa generalizada de Moore-Penrose*** o ***pseudoinversa*** de A es una matriz M de orden $n \times m$ tal que

1. $AMA = A$.
2. $MAM = M$.
3. AM es hermitiana (simétrica).
4. MA es hermitiana (simétrica).

Ejemplo 10.1.1. Si A es cuadrada y no singular, una inversa generalizada de Moore-Penrose es A^{-1}. Otro caso especialmente sencillo es para

$$D = \begin{pmatrix} \alpha_1 & & & & \\ & \alpha_2 & & & \\ & & \ddots & & \\ & & & \alpha_r & \\ & & & & \mathbf{0} \end{pmatrix}_{m \times n},$$

una matriz diagonal de orden $m \times n$, con $\alpha_i \neq 0, i = 1, \ldots, r$. Entonces una inversa generalizada de Moore-Penrose de la matriz D es

$$M = \begin{pmatrix} \alpha_1^{-1} & & & & \\ & \alpha_2^{-1} & & & \\ & & \ddots & & \\ & & & \alpha_r^{-1} & \\ & & & & \mathbf{0} \end{pmatrix}_{n \times m}.$$

Por ejemplo, si

$$D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ entonces } M = \begin{pmatrix} \frac{1}{3} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Inversa generalizada de Moore-Penrose

Sea A una matriz $m \times n$. Entonces:

1. A tiene una única inversa generalizada de Moore-Penrose, que notaremos por A^+.
2. Si la descomposición en valores singulares de A es

$$A_{m\times n} = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}_{m\times n} V^* \text{ entonces}$$

$$A^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix}_{n\times m} U^*.$$

PRUEBA: Para la primera parte, sean M_1, M_2 inversas generalizadas de Moore-Penrose de una matriz A. Entonces

$$M_1^* A^* = AM_1, A^* M_1^* = M_1 A, A^* M_2^* = M_2 A, M_2^* A^* = AM_2.$$

Además

$$\begin{aligned} AM_1 &= (AM_1)^* = M_1^* A^* = M_1^* (AM_2 A)^* = M_1^* A^* (AM_2)^* \\ &= (AM_1)^* AM_2 = AM_1 AM_2 = AM_2 \end{aligned}$$

y

$$\begin{aligned} M_1 A &= (M_1 A)^* = A^* M_1^* = (AM_2 A)^* M_1^* = (M_2 A)^* A^* M_1^* \\ &= M_2 A(M_1 A)^* = M_2 AM_1 A = M_2 A. \end{aligned}$$

Por tanto, nos queda

$$M_1 = M_1 AM_1 = M_1 AM_2 = M_2 AM_2 = M_2.$$

Para la segunda, basta comprobar que dicha matriz verifica las condiciones de la inversa generalizada de Moore-Penrose. Sea $M = V\Sigma^+ U^*$. Entonces

1. $AMA = (U\Sigma V^*)(V\Sigma^+U^*)(U\Sigma V^*) = U(\Sigma\Sigma^+\Sigma)V^* = U\Sigma V^* = A.$
2. $MAM = V(\Sigma^+U^*)(U\Sigma V^*)(V\Sigma^+U^*) = V(\Sigma^+\Sigma\Sigma^+)U^* = V\Sigma^+U^* = M.$
3. $$(AM)^* = (U\Sigma V^*V\Sigma^+U^*)^* = (U\Sigma\Sigma^+U^*)^* = U(\Sigma\Sigma^+)^*U^* = U\Sigma\Sigma^+U^* \\ = U\Sigma V^*V\Sigma^+U^* = AM.$$
4. $$(MA)^* = (V\Sigma^+U^*U\Sigma V^*)^* = (V\Sigma^+\Sigma V^*)^* = V(\Sigma^+\Sigma)^*V^* = V\Sigma^+\Sigma V^* \\ = V\Sigma^+U^*U\Sigma V^* = MA.$$

□

A partir de ahora trabajaremos en $\mathbb{R}$, aunque todo sigue siendo válido en $\mathbb{C}$, cambiando la trasposición por la traspuesta conjugada.

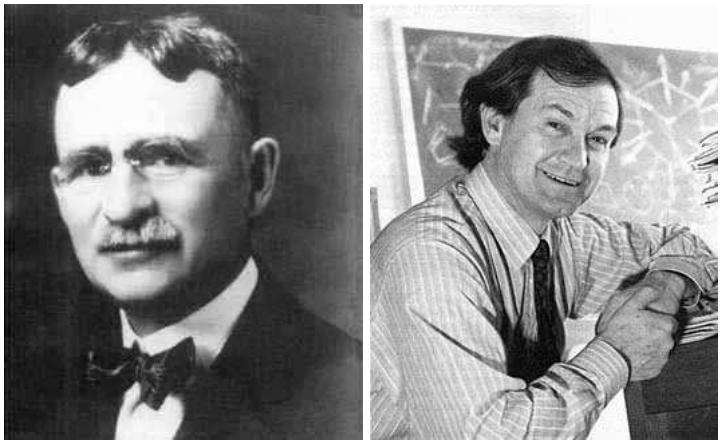

Figura 10.1: E.H. Moore (1862-1932), R. Penrose (1931-)

Ejemplo 10.1.2. Consideremos la matriz

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Calculamos en primer lugar su descomposición en valores singulares. Tenemos que

$$A^tA = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \text{ de autovalores } \lambda_1 = 4, \lambda_2 = 2.$$

Entonces los valores singulares son $\sigma_1 = 2, \sigma_2 = \sqrt{2}$, y

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}.$$

Ahora debemos calcular una base ortonormal de autovectores de $A^t A$.

Para el autovalor λ_1,

$$\text{null}(A^t A - \lambda_1 I) \equiv \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Una base ortonormal de este espacio de autovectores está formada por el vector

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Para el autovalor λ_2,

$$\text{null}(A^t A - \lambda_2 I) \equiv \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \boldsymbol{w}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Una base ortonormal de este espacio de autovectores está formada por el vector

$$\boldsymbol{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Por tanto, la matriz V es

$$V = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{pmatrix} = \begin{bmatrix} 1/2\sqrt{2} & -1/2\sqrt{2} \\ 1/2\sqrt{2} & 1/2\sqrt{2} \end{bmatrix}.$$

Para el cálculo de la matriz $U_{3\times 3}$ necesitamos

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \begin{bmatrix} 1/2\sqrt{2} \\ 0 \\ 1/2\sqrt{2} \end{bmatrix},$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Ahora debemos completar $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ a una base ortonormal de $\mathbb{R}^3$. Para ello, calculamos el subespacio vectorial $\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle^{\perp}$:

$$\langle \boldsymbol{u}_1, \boldsymbol{u}_2 \rangle^{\perp} \equiv \left\{ \begin{array}{rcl} \frac{\sqrt{2}}{2} x_1 + \frac{\sqrt{2}}{2} x_3 & = & 0, \\ x_2 & = & 0. \end{array} \right. \Rightarrow \left\{ \begin{array}{rcl} x_1 & = & -x_3, \\ x_2 & = & 0, \\ x_3 & = & x_3. \end{array} \right. \Rightarrow \boldsymbol{u}_3' = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Entonces basta normalizar para obtener

$$\boldsymbol{u}_3 = \frac{1}{\boldsymbol{u}_3'} \boldsymbol{u}_3' = \begin{bmatrix} -1/2\sqrt{2} \\ 0 \\ 1/2\sqrt{2} \end{bmatrix},$$

de donde

$$U = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{u}_3 \end{pmatrix} = \begin{bmatrix} 1/2\sqrt{2} & 0 & -1/2\sqrt{2} \\ 0 & 1 & 0 \\ 1/2\sqrt{2} & 0 & 1/2\sqrt{2} \end{bmatrix}.$$

Como

$$\Sigma^+ = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \end{pmatrix},$$

tenemos que

$$A^+ = V \Sigma^+ U^t = \begin{bmatrix} 1/4 & -1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}.$$

Propiedades de la pseudoinversa

1. $A^+ = A^{-1}$ cuando A es no singular.
2. $(A^+)^+ = A$.
3. $(A^+)^t = (A^t)^+$.
4. $(PAQ)^+ = Q^t A^+ P^t$ cuando P y Q son matrices ortogonales.
5. $A^+ = \begin{cases} (A^tA)^{-1}A^t & \text{cuando rango}(A_{m\times n}) = n, \\ A^t(AA^t)^{-1} & \text{cuando rango}(A_{m\times n}) = m. \end{cases}$
6. $A^t = A^tAA^+ = A^+AA^t$.
7. $A^+ = A^t(AA^t)^+ = (A^tA)^+A^t$.
8. $\text{Col}(A^+) = \text{Col}(A^t) = \text{Col}(A^+A)$.

Demostración. Las propiedades 1),2),3),4) y 5) se prueban a partir de las propiedades que definen la inversa generalizada de Moore-Penrose. Para 5), observemos que si rango$(A_{m\times n}) = n$, entonces rango(A^tA) = rango$(A) = n$, por lo que A^tA es no singular. Análogamente para AA^t.

6. Recordemos que AA^+ y A^+A son simétricas, es decir, $AA^+ = (AA^+)^t$ y $A^+A = (A^+A)^t$. Entonces
$$A^tAA^+ = A^t(AA^+)^t = (AA^+A)^t = A^t,$$
$$A^+AA^t = (A^+A)^tA^t = (AA^+A)^t = A^t.$$

7. Si tomamos la descomposición en valores singulares de A, entonces
$$AA^t = U\begin{pmatrix} \Sigma & \\ & \mathbf{0} \end{pmatrix}_{m\times n} V^tV\begin{pmatrix} \Sigma^t & \\ & \mathbf{0} \end{pmatrix}_{n\times m} U^t = U\begin{pmatrix} \Sigma^2 & \\ & \mathbf{0} \end{pmatrix}_{m\times m} U^t,$$
de donde (propiedad 4),
$$(AA^t)^+ = U\begin{pmatrix} \Sigma^{-2} & \\ & \mathbf{0} \end{pmatrix}_{m\times m} U^t.$$
Entonces
$$\begin{aligned} A^t(AA^t)^+ &= V\begin{pmatrix} \Sigma^t & \\ & \mathbf{0} \end{pmatrix}_{n\times m} U^tU\begin{pmatrix} \Sigma^{-2} & \\ & \mathbf{0} \end{pmatrix}_{m\times m} U^t \\ &= V\begin{pmatrix} \Sigma^{-1} & \\ & \mathbf{0} \end{pmatrix}_{n\times m} U^t = A^+. \end{aligned}$$

La otra igualdad es análoga.

8. Tenemos la cadena de contenciones

$$\mathrm{Col}(A^t) \overset{6(1)}{\subset} \mathrm{Col}(A^+A) \subset \mathrm{Col}(A^+) \overset{7(1)}{\subset} \mathrm{Col}(A^t),$$

por lo que todas son igualdades.

□

Como hemos visto en los ejemplos, pueden existir múltiples soluciones mínimo cuadráticas del sistema $A\boldsymbol{x} = \boldsymbol{b}$, y nos planteamos elegir una óptima en algún sentido. Nos centramos en las de norma mínima.

Solución mínimo cuadrática de norma mínima

1. Cuando $A\boldsymbol{x} = \boldsymbol{b}$ es compatible, $\boldsymbol{u} = A^+\boldsymbol{b}$ es la solución de norma mínima.
2. Cuando $A\boldsymbol{x} = \boldsymbol{b}$ es incompatible, $\boldsymbol{u} = A^+\boldsymbol{b}$ es la solución mínimo cuadrática de norma mínima.

PRUEBA: Supongamos que el sistema es compatible, y $A\boldsymbol{x}_0 = \boldsymbol{b}$. Cambiamos A por AA^+A para escribir $\boldsymbol{b} = A\boldsymbol{x}_0 = AA^+A\boldsymbol{x}_0 = AA^+\boldsymbol{b}$. Entonces $A^+\boldsymbol{b}$ resuelve el sistema $A\boldsymbol{x} = \boldsymbol{b}$ cuando es compatible. Ahora vamos a ver que es de norma mínima. La solución general del sistema es de la forma $A^+\boldsymbol{b} + \boldsymbol{h}$, con $\boldsymbol{h} \in \mathrm{null}(A)$ (solución particular más solución del homogéneo). Sabemos, por las propiedades de la pseudoinversa, que $\mathrm{Col}(A^+) = \mathrm{Col}(A^t)$, por lo que $A^+\boldsymbol{b} \in \mathrm{Col}(A^t) = \mathrm{null}(A)^\perp$. Entonces $A^+\boldsymbol{b} \perp \boldsymbol{h}$. Por el teorema de Pitágoras,

$$\left\| A^+\boldsymbol{b} + \boldsymbol{h} \right\|_2^2 = \left\| A^+\boldsymbol{b} \right\|_2^2 + \|\boldsymbol{h}\|_2^2 \geq \left\| A^+\boldsymbol{b} \right\|_2^2,$$

y la igualdad se da si y solamente si $\boldsymbol{h} = \boldsymbol{0}$. Así, $A^+\boldsymbol{b}$ es la única solución de norma mínima.

Supongamos ahora que el sistema es incompatible. Las soluciones mínimo cuadráticas son las soluciones del sistema de ecuaciones normales $A^tA\boldsymbol{x} = A^t\boldsymbol{b}$, que es compatible. Por lo anterior, la solución mínimo cuadrática de norma mínima de este sistema es $\boldsymbol{u} = (A^tA)^+A^t\boldsymbol{b} = A^+\boldsymbol{b}$, por la propiedad 7). □

Ejemplo 10.1.3. Consideremos el sistema $A\boldsymbol{x} = \boldsymbol{b}$, con

$$\begin{pmatrix} 1 & 2 \\ -2 & -4 \\ 2 & 4 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ -5 \\ 5 \end{pmatrix}.$$

En primer lugar, calculamos todas las soluciones mínimo cuadráticas. Para ello, resolvemos el sistema de ecuaciones normales $A^tA\boldsymbol{x} = A^t\boldsymbol{b}$. En este caso,

$$\left(\begin{array}{c|c} A^tA & A^t\boldsymbol{b} \end{array} \right) = \left[\begin{array}{cc|c} 9 & 18 & 24 \\ 18 & 36 & 48 \end{array} \right] \overset{\text{rref}}{\longrightarrow} \left[\begin{array}{ccc} 1 & 2 & 8/3 \\ 0 & 0 & 0 \end{array} \right]$$

Entonces el conjunto de soluciones se escribe como

$$\begin{cases} x_1 &= \frac{8}{3} - 2x_2, \\ x_2 &= x_2. \end{cases}$$

Si $\boldsymbol{u}$ es solución, entonces $\|\boldsymbol{u}\|^2 = \left(\frac{8}{3} - \lambda\right)^2 + \lambda^2$, y buscamos minimizar esta expresión. Si la vemos como una función f en λ, calculamos su derivada para calcular los extremos.

$$f'(\lambda) = 10\lambda - 32/3 \Rightarrow \lambda = 16/15.$$

En este valor se alcanza el mínimo, por lo que la solución mínimo cuadrática de norma mínima es

$$\boldsymbol{u}_0 = \begin{pmatrix} \frac{8}{3} - 2\frac{16}{15} \\ \frac{16}{15} \end{pmatrix} = \begin{pmatrix} \frac{8}{15} \\ \frac{16}{15} \end{pmatrix}.$$

Para casos más generales, se obtiene una función cuadrática en varias variables, que hay que minimizar. Las propiedades de A^+ nos permite decir que la solución mínimo cuadrática de norma mínima es $A^+\boldsymbol{b}$.

La descomposición en valores singulares de la matriz A es

$$A = U\Sigma V^t = \begin{pmatrix} -1/3 & 2/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \\ -2/3 & 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 3\sqrt{5} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -1/\sqrt{5} & -2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix},$$

de donde

$$A^+ = V\Sigma^+U^t = \begin{pmatrix} 1/45 & -2/45 & 2/45 \\ 2/45 & -4/45 & 4/45 \end{pmatrix}, \text{ y } A^+\boldsymbol{b} = \begin{pmatrix} \frac{8}{15} \\ \frac{16}{15} \end{pmatrix}.$$

10.2. Mínimos cuadrados sin rango completo

Existen problemas en los que, por su planteamiento, la matriz de coeficientes no es de rango completo, como ocurre en ciertos modelos de clasificación.

Ejemplo 10.2.1. Extraído de Searle, p. 392. En un análisis del peso de una planta productora de caucho, se consideran 6 ejemplares: 3 del tipo I, 2 del tipo II y 1 del tipo III. La tabla de pesos para esta muestra es

Tipo I	Tipo II	Tipo 3
101	84	32
105	88	
94		

Sea b_{ij} el peso de la planta j-ésima dentro del tipo i, con $i = 1,2,3$ y $j = 1,\ldots,n_i$, donde n_i es el número de observaciones de cada tipo. El problema es estimar el efecto del tipo de planta sobre el peso de la misma. Para ello, suponemos que

$$b_{ij} = \mu + \alpha_i + e_{ij},$$

donde μ representa la media de peso de la población, α_i es el efecto del tipo i sobre el peso y e_{ij} es el término residual aleatorio propio de la observación b_{ij}. Suponemos que las variables aleatorias e_{ij} son independientes, con media cero y la misma varianza σ^2. Para desarrollar el método de estimación escribimos las ecuaciones correspondientes a las observaciones:

$$\begin{array}{rllllll}
101 & = b_{11} & = \mu & +\alpha_1 & & & +e_{11} \\
105 & = b_{12} & = \mu & +\alpha_1 & & & +e_{12} \\
94 & = b_{13} & = \mu & +\alpha_1 & & & +e_{13} \\
85 & = b_{21} & = \mu & & +\alpha_2 & & +e_{21} \\
88 & = b_{22} & = \mu & & +\alpha_2 & & +e_{22} \\
32 & = b_{31} & = \mu & & & +\alpha_3 & +e_{31}
\end{array}$$

que en forma matricial es $\boldsymbol{b} = A\boldsymbol{x} + \boldsymbol{e}$, donde

$$\boldsymbol{b} = \begin{pmatrix} 101 \\ 105 \\ 94 \\ 85 \\ 88 \\ 32 \end{pmatrix}, A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

y $\boldsymbol{e}$ es el vector de los términos residuales. El sistema $A\boldsymbol{x} = \boldsymbol{b}$ es incompatible, y A no es de rango completo. Queremos calcular la solución de $A\boldsymbol{x} = \boldsymbol{b}$

mínimo cuadrática de norma mínima. Vamos a ver el procedimiento paso a paso. En primer lugar, calculamos la descomposición en valores singulares de $A = U\Sigma V^t$. Nos queda que

$$U = \begin{pmatrix} 0{,}4491 & 0{,}3487 & -0{,}1003 & 0{,}7500 & -0{,}0000 & -0{,}3227 \\ 0{,}4491 & 0{,}3487 & -0{,}1003 & -0{,}0956 & -0{,}0000 & 0{,}8109 \\ 0{,}4491 & 0{,}3487 & -0{,}1003 & -0{,}6545 & 0{,}0000 & -0{,}4882 \\ 0{,}3791 & -0{,}5529 & -0{,}2249 & 0{,}0000 & -0{,}7071 & 0{,}0000 \\ 0{,}3791 & -0{,}5529 & -0{,}2249 & 0{,}0000 & 0{,}7071 & 0{,}0000 \\ 0{,}3280 & -0{,}1542 & 0{,}9320 & -0{,}0000 & -0{,}0000 & 0{,}0000 \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} 2{,}9015 & 0 & 0 & 0 \\ 0 & 1{,}5449 & 0 & 0 \\ 0 & 0 & 1{,}0929 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} 0{,}8387 & -0{,}1384 & 0{,}1658 & 0{,}5000 \\ 0{,}4643 & 0{,}6772 & -0{,}2754 & -0{,}5000 \\ 0{,}2613 & -0{,}7158 & -0{,}4116 & -0{,}5000 \\ 0{,}1130 & -0{,}0998 & 0{,}8528 & -0{,}5000 \end{pmatrix},$$

y entonces

$$A^+ = V\Sigma^+ U^t = \begin{pmatrix} 0{,}0833 & 0{,}0833 & 0{,}0833 & 0{,}1250 & 0{,}1250 & 0{,}2500 \\ 0{,}2500 & 0{,}2500 & 0{,}2500 & -0{,}1250 & -0{,}1250 & -0{,}2500 \\ -0{,}0833 & -0{,}0833 & -0{,}0833 & 0{,}3750 & 0{,}3750 & -0{,}2500 \\ -0{,}0833 & -0{,}0833 & -0{,}0833 & -0{,}1250 & -0{,}1250 & 0{,}7500 \end{pmatrix},$$

$$A^+\boldsymbol{b} = \begin{pmatrix} 54{,}6250 \\ 45{,}3750 \\ 31{,}8750 \\ -22{,}6250 \end{pmatrix}.$$

Ejemplo 10.2.2. http://www.stat.umn.edu/geyer/5102/examp/dummy.html
Contrasta con Draper, Smith, "Applied Regression Analysis", 3 ed., p. 444.
http://www.stat.sc.edu/ tebbs/stat714/f10notes.pdf p. 42-43 para otro punto de vista.

10.2.1. * Aplicaciones estadísticas de Moore-Penrose

La inversa generalizada de Moore-Penrose es útil a la hora de construir formas cuadráticas sobre vectores aleatorios con distribución normal, que dan

lugar a distribuciones χ-cuadrado. Una situación habitual en inferencia estadística es aquella en la que tenemos una muestra estadística $\boldsymbol{t} \sim N_m(\boldsymbol{\theta}, \Omega)$, y queremos determinar si el vector θ de orden $m \times 1$ es igual a cero. De manera formal, queremos comprobar la hipótesis nula $H_0 : \boldsymbol{\theta} = \mathbf{0}$ contra la hipótesis alternativa $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$. Una aproximación a este problema, si Ω es definida positiva, es basar la decisión entre H_0 y H_1 sobre el estadístico

$$v_1 = \boldsymbol{t}^t \Omega^{-1} \boldsymbol{t}.$$

Si ahora T es cualquier matriz $m \times m$ tal que $TT^t = \Omega$ (Cholesky), y definimos $\boldsymbol{u} = T^{-1}\boldsymbol{t}$, entonces $E(\boldsymbol{u}) = T^{-1}\boldsymbol{\theta}$ y

$$\operatorname{var}(\boldsymbol{u}) = T^{-1} \operatorname{var}(\boldsymbol{t})(T^{-1})^t = T^{-1}(TT^t)(T^{-1})^t = I_m,$$

por lo que $\boldsymbol{u} \sim N_m(T^{-1}\boldsymbol{\theta}, I_m)$. Por tanto, $u_1, \ldots, u_m$ son variables aleatorias independientes con distribución normal, y

$$v_1 = \boldsymbol{t}^t \Omega^{-1} \boldsymbol{t} = \boldsymbol{u}^t \boldsymbol{u} = \sum_{i=1}^{m} u_i^2$$

sigue una distribución χ-cuadrado con m grados de libertad. Esta distribución χ-cuadrado es central si $\boldsymbol{\theta} = \mathbf{0}$ y no central si $\boldsymbol{\theta} \neq \mathbf{0}$, por lo que escogeríamos H_1 sobre H_0 si v_1 es suficientemente grande.

Cuando Ω es semi-definida positiva, la construcción anterior de v_1 puede generalizarse con la inversa generalizada de Moore-Penrose de Ω. En este caso, si $\operatorname{rango}(\Omega) = r$, escribamos $\Omega = X_1 \Lambda_1 X_1^t$ por el teorema espectral de matrices simétricas, con X_1 ortogonal y Λ_1 matriz diagonal con los autovalores no nulos de Ω. Entonces $\Omega^+ = X_1 \Lambda_1^{-1} X_1$, y llamemos $\boldsymbol{w} = \Lambda_1^{-1/2} X_1^t \boldsymbol{t}$, que sigue una distribución normal $N_r(\Lambda_1^{-1/2} X_1^t \boldsymbol{\theta}, I_r)$, pues

$$\begin{aligned} \operatorname{var}(\boldsymbol{w}) &= \Lambda_1^{-1/2} X_1^t \operatorname{var}(\boldsymbol{t}) X_1 \Lambda_1^{-1/2} \\ &= \Lambda_1^{-1/2} X_1^t (X_1 \Lambda_1 X_1^t) X_1 \Lambda_1^{-1/2} \\ &= I_r. \end{aligned}$$

Así, como las w_i son variables aleatorias independientes con distribución normal, entonces

$$v_2 = \boldsymbol{t}^t \Omega^+ \boldsymbol{t} = \boldsymbol{w}^t \boldsymbol{w} = \sum_{i=1}^{r} w_i^2$$

sigue una distribución χ-cuadrado, que es central si $\Lambda_1^{-1/2} X_1^t \theta = \mathbf{0}$, con r grados de libertad.

10.2.2. * Continuidad de la inversa generalizada de Moore-Penrose

Es útil establecer la continuidad de una función porque las funciones continuas gozan de buenas propiedades. En este apartado daremos condiciones bajo las que los elementos de A^+ son funciones continuas de los elementos de A. Consideremos en primer lugar el determinante de una matriz cuadrada y la inversa de una matriz no singular. Recordemos que el determinante de una matriz A de orden $m \times m$ se pude expresar como una suma de términos, con signo más o menos, cada uno de los cuales es un producto de términos de A. Por la continuidad del producto y suma de funciones, se tiene que el determinante de una matriz es una función continua de sus elementos.

Supongamos ahora que A es una matriz de orden $m \times m$ no singular, es decir, $\det(A) \neq 0$. Entonces

$$A^{-1} = \det(A)^{-1} \operatorname{Adj}(A),$$

donde $\operatorname{Adj}(A)$ es la adjunta traspuesta de A. Si $A_1, A_2, \ldots$ es una sucesión de matrices tal que $A_i \to A$ cuando $i \to \infty$ entonces, por la continuidad de la función determinante, $\det(A_i) \to \det(A)$ cuando $i \to \infty$, y entonces existe un N tal que $\det(A_i) \neq 0$ cuando $i > N$. Como cada elemento de la matriz adjunta es ± 1 veces un determinante, se sigue de la continuidad de la función determinante que si $\operatorname{Adj}(A_i)$ es la matriz adjunta de A_i, entonces $\operatorname{Adj}(A_i) \to \operatorname{Adj}(A)$ cuando $i \to \infty$. Por tanto, tenemos que para una matriz no singular A, su matriz inversa A^{-1} es una función continua de los elementos de A.

La continuidad de la inversa generalizada de Moore-Penrose no es tan directa como la continuidad de la inversa de una matriz no singular. En general, si A es una matriz $m \times n$ y $A_1, A_2, \ldots$ es una sucesión de matrices $m \times n$ con $A_i \to A$ cuando $i \to \infty$, entonces no se tiene siempre que $A_i^+ \to A^+$.

Ejemplo 10.2.3. Consideremos la sucesión A_i de matrices 2×2

$$A_i = \begin{pmatrix} 1/i & 0 \\ 0 & 1 \end{pmatrix}.$$

Es claro que $A_i \to A$, donde

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Observemos que $\operatorname{rango}(A) = 1$, mientras que $\operatorname{rango}(A_i) = 2$ para todo i. Esta va a ser la condición que necesitaremos posteriormente. Tenemos que

$$A_i^+ = \begin{pmatrix} i & 0 \\ 0 & 1 \end{pmatrix},$$

que no converge a ninguna matriz, y

$$A^+ = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

En general, si tenemos una sucesión de matrices $A_1, A_2, \ldots$, para las que $\text{rango}(A_i) = \text{rango}(A)$ a partir de un cierto i, entonces no aparecerá el problema anterior.

Teorema 10.2.4. *Sea A una matriz $m \times n$ y $A_1, A_2, \ldots$ una sucesión de matrices $m \times n$ tales que $A_i \to A$ cuando $i \to \infty$. Entonces*

$$A_i^+ \to A^* \text{ cuando } i \to \infty$$

si y solamente si existe un entero N tal que

$$\text{rango}(A_i) = \text{rango}(A) \text{ para todo } i > N.$$

Demostración. Penrose(1955), Campbell-Meyer(1979) (por verificar) □

Ejemplo 10.2.5. Las condiciones de continuidad para la inversa generalizada de Moore-Penrose tienen importantes aplicaciones en problemas de estimación y verificación de hipótesis. En particular, en este ejemplo, hablaremos de una propiedad, denominada consistencia, que algunos estimadores poseen. Un estimador t, calculado a partir de una muestra de tamaño n, se dice que es un estimador consistente de un parámetro θ si t converge en probabilidad a θ, es decir, si

$$\lim_{n\to\infty} P(|t-\theta| \geq \epsilon) = 0,$$

para todo $\epsilon > 0$. Un resultado muy importante asociado con la propiedad de consistencia es que funciones continuas de estimadores consistentes son consistentes. Así, si t es un estimador consistente de θ y $g(t)$ una función continua de t, entonces $g(t)$ es un estimador consistente de $g(\theta)$. Ahora aplicaremos algunas de estas ideas a una situación que involucre la estimación de la inversa generalizada de Moore-Penrose de una matriz de parámetros.

Por ejemplo, sea Ω una matriz de covarianza $m \times m$ semi-definida positiva, con $\text{rango}(\Omega) = r < m$. Supongamos que los elementos de la matriz Ω son desconocidos y tienen que ser estimados. Supongamos además dos cosas:

- la estimación muestral $\hat{\Omega}$ es definida positiva con probabilidad 1, esto es, $\text{rango}(\hat{\Omega}) = m$ con probabilidad 1.
- $\hat{\Omega}$ es un estimador consistente de Ω, es decir, cada elemento de $\hat{\Omega}$ es un estimador consistente del correspondiente elemento de Ω.

Sin embargo, como rango(Ω) = $r < m$, $\hat{\Omega}^+$ no es un estimador consistente de Ω^+. De manera intuitiva, el problema aquí es evidente. Si $\hat{\Omega} = X\Lambda X^t$ es la descomposición espectral de $\hat{\Omega}$ entonces $\hat{\Omega}^+ = X\Lambda^{-1}X^t$. La consistencia de $\hat{\Omega}$ implica que cuando n crece, los $m-r$ elementos menores de la diagonal de Λ convergen a cero, por lo que los $m-r$ elementos mayores de la diagonal de Λ^{-1} crecen sin límite.

La dificultad la podemos evitar si conocemos el rango r de Ω. En este caso, $\hat{\Omega}$ se puede ajustar para producir un estimador de Ω de rango r. Por ejemplo, si $\hat{\Omega}$ tiene autovalores $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, con autovectores ortonormales asociados $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$, y consideramos el proyector

$$P_r = \sum_{i=1}^{r} \boldsymbol{v}_i \boldsymbol{v}_i^t$$

entonces

$$\hat{\Omega}_* = P_r \hat{\Omega} P_r = \sum_{i=1}^{r} \lambda_i \boldsymbol{v}_i \boldsymbol{v}_i^t$$

será un estimador de Ω de rango r. Por la continuidad de los proyectores, se puede probar que $\hat{\Omega}_*$ es también un estimador consistente de Ω. Y lo que es más importante: como rango($\hat{\Omega}_*$) = rango(Ω) = r, el teorema anterior nos garantiza que $\hat{\Omega}_*^+$ es un estimador consistente de Ω^+.

10.3. * Métodos de cálculo de la inversa de Moore-Penrose

10.4. Inversas generalizadas de tipo 1

La inversa generalizada de Moore-Penrose no es más que uno de los tipos de inversas generalizadas que se han desarrollado en los últimos años. En esta sección discutiremos otras que tienen aplicaciones en estadística.

Inversa generalizada o 1-inversa

Sea A una matriz de orden $m \times n$. Una ***inversa generalizada*** de A o 1-***inversa*** es una matriz G de orden $n \times m$ tal que $AGA = A$. La notaremos por A^-.

Observemos que solamente le pedimos la primera condición de la inversa generalizada de Moore-Penrose, y de ahí el nombre de 1-inversa. La propia

pseudoinversa es una 1-inversa, pero no es la única. Por ejemplo, consideremos una matriz diagonal $D_{m\times n}$ de rango r, con los elementos no nulos en las primeras posiciones de la diagonal; podemos escribir

$$D_{m\times n} = \begin{pmatrix} D_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}, \text{ con } D_r \text{ no singular.}$$

Entonces, para cualesquiera matrices $K_{r\times(m-r)}, L_{(n-r)\times r}, M_{(n-r)\times(m-r)}$, la matriz

$$D^- = \begin{pmatrix} D_r^{-1} & K \\ L & M \end{pmatrix}$$

es una $\{1\}$-inversa de D. Se comprueba fácilmente mediante multiplicación por bloques.

Si ahora partimos de una matriz $A_{m\times n}$, por la forma normal del rango, existen $P_{m\times m}, Q_{n\times n}$ matrices no singulares tales que $A = PDQ$, con $D_{m\times n}$ diagonal de rango r de la forma que hemos considerado antes. Entonces la matriz $A^- = Q^{-1}D^-P^{-1}$ es una $\{1\}$-inversa de A para cada $\{1\}$-inversa D^- que escojamos para D.

Ejemplo 10.4.1. Consideremos la matriz

$$D = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Entonces cualquier matriz de la forma

$$G = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}$$

es una 1-inversa de D.

El ejemplo anterior nos indica que hay muchas $\{1\}$-inversas. Veamos una forma de obtenerlas todas.

Forma de las $\{1\}$-inversas

Sea A una matriz de orden $m \times n$ y A^- una inversa generalizada cualquiera. Entonces para cualquier matriz C de orden $n \times m$, la matriz

$$A^- + C - A^-ACAA^-$$

es una inversa generalizada de A, y cada inversa generalizada de A se puede expresar de esta forma para alguna matriz C.

PRUEBA: Como $AA^-A = A$, se tiene que

$$\begin{aligned} A(A^- + C - A^-ACAA^-)A &= AA^-A + ACA - AA^-ACAA^-A \\ &= A + ACA - ACA = A, \end{aligned}$$

por lo que $A^- + C - A^-ACAA^-$ es una inversa generalizada de A independientemente de la elección de A^- y C. Sea ahora B una inversa generalizada de A, y definimos

$$C = B - A^-,$$

donde A^- es una inversa generalizada particular. Como $ABA = A$, tenemos que

$$\begin{aligned} A^- + C - A^-ACAA^- &= A^- + (B - A^-) - A^-A(B - A^-)AA^- \\ &= B - A^-ABAA^- + A^-AA^-AA^- \\ &= B - A^-AA^- + A^-AA^- = B, \end{aligned}$$

y se sigue el resultado. □

Invariancia asociada a la $\{1\}$-inversa

$$A(A^tA)^-A^tA = A. \tag{10.4.1}$$

$$A(A^tA)^-A^t = AA^+, \tag{10.4.2}$$

y, por tanto, $A(A^tA)^-A^t$ es simétrica, y no depende de la elección de la inversa generalizada $(A^tA)^-$.

PRUEBA: Observemos que

$$\begin{aligned} A(A^tA)^-A^tA = AA^+A(A^tA)^-A^tA &= (AA^+)^tA(A^tA)^-A^tA \\ &= (A^+)^tA^tA(A^tA)^-A^tA = (A^+)^tA^tA \\ &= (AA^+)^tA = AA^+A = A. \end{aligned}$$

Para la segunda parte, se tiene que

$$\begin{aligned} A(A^tA)^-A^t &= A(A^tA)^-(AA^+A)^t = A(A^tA)^-A^t(A^+)^tA^t = A(A^tA)^-A^t(AA^+)^t \\ &= A(A^tA)^-A^tAA^+, \text{ porque } AA^+ \text{ es simétrica} \\ &= AA^+, \text{ por el apartado anterior.} \end{aligned}$$

□

A partir de las inversas generalizadas podemos expresar el conjunto de soluciones de un sistema compatible $A\boldsymbol{x} = \boldsymbol{b}$. Recordemos que ya habíamos expresado dicho conjunto en la forma $\boldsymbol{p} + \boldsymbol{h}$, donde $\boldsymbol{p}$ es una solución particular y $\boldsymbol{h}$ recorre el conjunto de soluciones del sistema lineal homogéneo $A\boldsymbol{x} = \boldsymbol{0}$. Lo que vamos a hacer es expresar $\boldsymbol{p}$ y $\boldsymbol{h}$ en función de A^-.

- En primer lugar, se tiene que $A^-\boldsymbol{b}$ es una solución particular. En efecto, si $\boldsymbol{u}$ es una solución del sistema, entonces $AA^-\boldsymbol{b} = AA^-A\boldsymbol{u} = A\boldsymbol{u} = \boldsymbol{b}$.
- En segundo lugar, un vector $\boldsymbol{h}$ es solución del sistema homogéneo $A\boldsymbol{x} = \boldsymbol{0}$ si y solamente si $\boldsymbol{h} = (I - A^-A)\boldsymbol{y}$ para algún vector $\boldsymbol{y}$. Por un lado, si $\boldsymbol{h} = (I - A^-A)\boldsymbol{y}$ para algún vector $\boldsymbol{y}$, entonces $A\boldsymbol{h} = A\boldsymbol{y} - AA^-A\boldsymbol{y} = \boldsymbol{0}$, y $\boldsymbol{h}$ es solución del sistema homogéneo. Recíprocamente, si $\boldsymbol{h}$ es solución, entonces $\boldsymbol{h} = \boldsymbol{h} - A^-(A\boldsymbol{h}) = (I - A^-A)\boldsymbol{h}$.

Hemos probado entonces lo siguiente.

Soluciones de un sistema compatible

Sea $A\boldsymbol{x} = \boldsymbol{b}$ un sistema compatible. Las soluciones $\boldsymbol{u}$ del sistema son de la forma $\boldsymbol{u} = A^-\boldsymbol{b} + (I - A^-A)\boldsymbol{y}$, con $\boldsymbol{y} \in \mathbb{R}^n$.

Nos queda una cuestión pendiente, relativa al cálculo de una $\{1\}$-inversa. Sabemos que, a través de la SVD, podemos obtener una de ellas, como es la A^+. Pero nos preguntamos si existe una forma más simple de obtener alguna otra.

Cálculo de una $\{1\}$-inversa

Sea $A_{m\times n}$ de rango r, y consideremos matrices $P_{m\times m}, Q_{n\times n}$ no singulares tales que

$$PAQ = \begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}.$$

Entonces para cualquier $L_{(n-r)\times(m-r)}$ la matriz

$$G = Q\begin{pmatrix} I_r & 0 \\ 0 & L \end{pmatrix}P$$

es una $\{1\}$-inversa de A.

PRUEBA: La existencia de las matrices P y Q se deduce de la forma reducida por filas de A, más un intercambio de columnas. Escribamos

$$A = P^{-1}\begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}Q^{-1}.$$

Entonces

$$\begin{aligned} AGA &= P^{-1}\begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}Q^{-1}Q\begin{pmatrix} I_r & 0 \\ 0 & L \end{pmatrix}PP^{-1}\begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} I_r & KL \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= P^{-1}\begin{pmatrix} I_r & K \\ 0 & 0 \end{pmatrix}Q^{-1} \\ &= A. \end{aligned}$$

□

Existe una especie de recíproco, que se puede consultar en [**?**, p. 38].

Nota 10.4.2. Si la matriz A es de rango r y tiene la forma

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix},$$

con A_1 no singular de orden r, entonces una $\{1\}$-inversa de A es de la forma

$$G_0 = \begin{pmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

En efecto, consideremos las transformaciones necesarias para llevar la matriz A a la forma reducida por filas. Tenemos

$$\begin{aligned} \begin{pmatrix} A_1 & A_2 & I_r & 0 \\ A_3 & A_4 & 0 & I_s \end{pmatrix} &\to \begin{pmatrix} I & A_1^{-1}A_2 & A_1^{-1} & 0 \\ A_3 & A_4 & 0 & I_s \end{pmatrix} \\ &\to \begin{pmatrix} I & A_1^{-1}A_2 & A_1^{-1} & 0 \\ 0 & A_4 - A_3A_1^{-1}A_2 & -A_3A_1^{-1} & I_s \end{pmatrix}. \end{aligned}$$

Como rango(A) = rango$(A_1) = r$, se tiene que $A_4 - A_3A_1^{-1}A_2 = 0$, por lo que la matriz P de cambio es

$$P = \begin{pmatrix} A_1^{-1} & 0 \\ -A_3A_1^{-1} & I_s \end{pmatrix}.$$

Basta entonces tomar en el teorema anterior $L = 0$ para obtener el resultado.

Ejemplo 10.4.3. Consideremos el sistema $A\boldsymbol{x} = \boldsymbol{b}$, donde

$$A = \begin{pmatrix} -6 & 2 & -2 & -3 \\ 3 & -1 & 5 & 2 \\ -3 & 1 & 3 & -1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}.$$

Entonces

$$[A \mid I] \xrightarrow{\text{rref}} \begin{bmatrix} 1 & -1/3 & 0 & \frac{11}{24} & 0 & 1/8 & -\frac{5}{24} \\ 0 & 0 & 1 & 1/8 & 0 & 1/8 & 1/8 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} = [E_A \mid P],$$

de donde tomamos

$$P = \begin{bmatrix} 0 & 1/8 & -\frac{5}{24} \\ 0 & 1/8 & 1/8 \\ 1 & 1 & -1 \end{bmatrix}.$$

Reordenamos las columnas de E_A para tener la matriz identidad en la parte superior izquierda. Entonces

$$E_A Q = E_A \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & -1/3 & \frac{11}{24} \\ 0 & 1 & 0 & 1/8 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

y tomamos como $\{1\}$-inversa a

$$A^- = Q \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P = \begin{bmatrix} 0 & 1/8 & -\frac{5}{24} \\ 0 & 0 & 0 \\ 0 & 1/8 & 1/8 \\ 0 & 0 & 0 \end{bmatrix}.$$

Las soluciones son de la forma

$$A^-\boldsymbol{b} + (I - A^- A)\boldsymbol{y} = \begin{bmatrix} -\frac{19}{24} + 1/3\, y_2 - \frac{11}{24}\, y_4 \\ y_2 \\ \frac{7}{8} - 1/8\, y_4 \\ y_4 \end{bmatrix}.$$

10.5. Inversas generalizadas mínimo cuadráticas

Inversa generalizada mínimo cuadrática

Sea A una matriz de orden $m \times n$. Una ***inversa generalizada mínimo cuadrática*** de A es una matriz G de orden $n \times m$ tal que $AGA = A$ y AG es simétrica. La notaremos por $A^{\square}$. También se la denomina $\{1,3\}$-inversa.

Ejemplo 10.5.1. Consideremos la matriz

$$D = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Entonces cualquier matriz de la forma

$$G = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha \\ \beta_1 & \beta_2 & \beta_3 \end{pmatrix}$$

es una $\{1,3\}$-inversa de D.

El motivo del nombre de esta inversa generalizada es que nos va dar una expresión de las soluciones mínimo cuadráticas de un sistema incompatible.

Sabemos que las soluciones mínimo cuadráticas son las soluciones del sistema compatible de ecuaciones normales $A^tA\boldsymbol{x} = A^t\boldsymbol{b}$. Entonces, por la sección anterior, el conjunto de soluciones mínimo cuadráticas se puede expresar como

$$(A^tA)^-A^t\boldsymbol{b} + (I - (A^tA)^-A^tA)\boldsymbol{y}.$$

Lo que vamos a probar es que la matriz $(A^tA)^-A^t$ es una inversa mínimo cuadrática.

Cálculo de inversas mínimo cuadráticas

Sea A una matriz de orden $m \times n$.

1. Para cualquier inversa mínimo cuadrática $A^{\square}$ de A se verifica que $AA^{\square} = AA^+$.
2. $(A^tA)^-A^t$ es una inversa mínimo cuadrática de A para cualquier $\{1\}$-inversa $(A^tA)^-$ de A^tA.

PRUEBA: Como $AA^{\square}A = A$ y $(AA^{\square})^t = AA^{\square}$, se tiene que

$$\begin{aligned} AA^{\square} &= AA^{+}AA^{\square} = (AA^{+})^t(AA^{\square})^t = (A^{+})^tA^t(A^{\square})^tA^t \\ &= (A^{+})^t(AA^{\square}A)^t = (A^{+})^tA^t = (AA^{+})^t = AA^{+}, \end{aligned}$$

y tenemos la primera parte. Para la segunda, recordemos que $A(A^tA)^-A^tA = A$, según vimos en (**??**), lo que nos da la primera condición de una $\{1,3\}$-inversa. Además, $A(A^tA)^-A^t$ es simétrica, tal como probamos en (**??**). □

Por tanto, el conjunto de soluciones mínimo cuadráticas de un sistema $A\boldsymbol{x} = \boldsymbol{b}$ es de la forma

$$A^{\square}\boldsymbol{b} + (I - A^{\square}A)\boldsymbol{y}, \boldsymbol{y} \in \mathbb{R}^m.$$

Ejemplo 10.5.2. Vamos a calcular una inversa mínimo cuadrática de

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 2 & 0 & 2 \end{pmatrix}.$$

En primer lugar, obtenemos

$$A^tA = \begin{bmatrix} 7 & 2 & 9 \\ 2 & 2 & 4 \\ 9 & 4 & 13 \end{bmatrix}.$$

Para calcular una $\{1\}$-inversa de A^tA, obtenemos su forma reducida por filas, y una matriz de paso.

$$[A^tA|I_3] \overset{\text{rref}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & 0 & -2/5 & 1/5 \\ 0 & 1 & 1 & 0 & \frac{9}{10} & -1/5 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix} = [E|P].$$

No hay que reordenar las columnas de E, por lo que una $\{1\}$-inversa de A^tA es

$$(A^tA)^- = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} P = \begin{bmatrix} 0 & -2/5 & 1/5 \\ 0 & \frac{9}{10} & -1/5 \\ 1 & 1 & -1 \end{bmatrix}.$$

Entonces

$$A^{\square} = (A^tA)^-A^t = \begin{bmatrix} 0 & 1/5 & 0 & 2/5 \\ 1/2 & -1/5 & 1/2 & -2/5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

es una inversa mínimo cuadrática de A. Si consideramos el vector

$$\boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix},$$

el conjunto de soluciones mínimo cuadráticas del sistema $A\boldsymbol{x} = \boldsymbol{b}$ es

$$\boldsymbol{u} = A^{\square}\boldsymbol{b} + (I - A^{\square}A)\boldsymbol{y} = \begin{bmatrix} 1/5 \\ 3/10 \\ 0 \end{bmatrix} + \left(I_3 - \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right) \boldsymbol{y}$$

$$= \begin{bmatrix} 1/5 - y_3 \\ 3/10 - y_3 \\ y_3 \end{bmatrix}.$$

Nota 10.5.3. Se puede continuar el estudio de las inversas generalizadas. Por ejemplo, podemos definir de forma análoga una $\{1,4\}$-inversa de una matriz $A_{m\times n}$ como una matriz $G_{n\times n}$ tal que $AGA = A$ y GA simétrica. Se puede probar ([**?**, 3.2]) que si un sistema $A\boldsymbol{x} = \boldsymbol{b}$ es compatible, entonces la solución de norma mínima es $\boldsymbol{u} = A^{(1,4)}\boldsymbol{b}$, donde $A^{(1,4)}$ es una $\{1,4\}$-inversa.

Capítulo 11

Matrices no negativas

11.1. Introducción

Una matriz A de coeficientes reales se denomina **no negativa** si $a_{ij} \geq 0$, y la notaremos por $A \geq 0$. En general, $A \geq B$ significa que $a_{ij} \geq b_{ij}$. De forma semejante, una matriz A es **positiva** si $a_{ij} > 0$, y escribiremos $A > 0$. En general, $A > B$ significa que $a_{ij} > b_{ij}$.

Existen múltiples aplicaciones de las matrices no negativas, y en este capítulo investigaremos sus propiedades. La primera se refiere en qué medida las propiedades $A > 0$ o $A \geq 0$ se traslada a los valores de sus autovalores y autovectores. El estudio de estas propiedades se denomina teoría de Perron-Frobenius, debido a las contribuciones de Oskar Perron y Ferdinand G. Frobenius. Perron dedicó su atención a las matrices positivas, y Frobenius hizo extensiones importantes a las matrices no negativas.

Seguimos los textos de [**?**, **?**, **?**].

Figura 11.1: O. Perron (1880-1975), F.G. Frobenius (1849-1917)

11.2. Matrices irreducibles

Matriz irreducible

Sea $n \geq 2$. Una matriz $A_{n\times n}$ es **reducible** si existe una matriz de permutación P tal que

$$P^t AP = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

donde A_{11} y A_{22} son cuadradas. En otro caso, se dice que A es **irreducible**.

Nota 11.2.1.

- Toda matriz positiva es irreducible.
- Una expresión de la forma $P^t AP$ tiene el efecto de intercambiar las filas de la misma forma que se intercambian las columnas.
- Si una matriz tiene una fila o columna nula, entonces es reducible.

Potencias positivas

Sea $A_{n\times n}$ una matriz no negativa e irreducible. Entonces

$$(I_n + A)^{n-1}\boldsymbol{v} > 0 \tag{11.2.1}$$

para todo $\boldsymbol{v} \geq 0$ no nulo. En particular, $(I_n + A)^{n-1} > 0$.

Demostración. Consideremos un vector $\boldsymbol{v}$ no nulo, con $\boldsymbol{v} \geq 0$, y definimos

$$\boldsymbol{w} = (I_n + A)\boldsymbol{v} = \boldsymbol{v} + A\boldsymbol{v}.$$

Como $A \geq 0$, el producto $A\boldsymbol{v}$ es un vector no negativo, por lo que $\boldsymbol{w}$ tiene, a lo más, tantos elementos no nulos como $\boldsymbol{v}$, y, al menos, tantos elementos positivos como $\boldsymbol{v}$. Vamos a probar que si $\boldsymbol{v}$ tiene alguna componente nula, entonces $\boldsymbol{w}$ tiene, al menos, un elemento no nulo más que $\boldsymbol{v}$. Sea P una matriz de permutación tal que

$$P\boldsymbol{v} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{0} \end{pmatrix}, \text{ y } \boldsymbol{u} > 0.$$

Entonces

$$P\boldsymbol{w} = P(I_n + A)\boldsymbol{v} = P(I_n + A)P^t \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{0} \end{pmatrix} + PAP^t \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{0} \end{pmatrix},$$

ya que $PP^t = I_n$. Si agrupamos los elementos de $P\boldsymbol{w}$ y de PAP^t de forma consistente con la de $P\boldsymbol{v}$, es decir,

$$P\boldsymbol{w} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \text{ y } PAP^t = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

entonces se sigue que

$$\boldsymbol{x} = \boldsymbol{u} + A_{11}\boldsymbol{u}, \boldsymbol{y} = A_{21}\boldsymbol{u}.$$

Como A es no negativa e irreducible, se tiene que $A_{11} \geq 0$, $A_{21} \geq 0$ y $A_{21} \neq 0$, por lo que $\boldsymbol{x} > 0, \boldsymbol{y} \geq 0$, y como $\boldsymbol{u} > 0$, tenemos garantía de que $\boldsymbol{y} \neq \boldsymbol{0}$. Por tanto, $\boldsymbol{w}$ tiene al menos una componente no nula más que $\boldsymbol{w}$.

Si $\boldsymbol{w} = (I_n + A)\boldsymbol{v}$ no es ya un vector positivo, repetimos el argumento anterior con $\boldsymbol{w}$, y entonces $(I_n + A)^2\boldsymbol{v}$ tendrá, al menos, dos componentes positivas más que $\boldsymbol{v}$. De este modo, tras, a lo más, $n-1$ pasos encontramos que $(I_n + A)^{n-1}\boldsymbol{v} > 0$ para cualquier vector no nulo $v \geq 0$.

Finalmente, si tomamos $\boldsymbol{v} = \boldsymbol{e}_i, i = 1, 2, \ldots, n$, donde $\boldsymbol{e}_i$ es el vector i-ésimo de la base estándar de $\mathbb{R}^n$, concluimos que $(I_n + A)^{n-1} > 0$. □

El concepto de matriz irreducible no está asociado con las magnitudes o los signos de los elementos de la matriz, sino con la disposición de los elementos nulos y no nulos en la matriz. Para estudiar si una matriz dada $A_{n\times n} \geq 0$ es irreducible, consideramos el grafo dirigido $\mathcal{G}(A)$ con n vértices, en donde hay una arista del vértice i al vértice j si y solamente si $a_{ij} \neq 0$.

Reducibilidad y grafos

Un grafo $\mathcal{G}$ es **fuertemente conexo** si para cada par de vértices i, j existe un camino dirigido $(i, i_1), (i_i, i_2), \ldots, (i_s, j)$ de aristas que conecta i con j.

1. Si P es una matriz de permutación, $\mathcal{G}(A) = \mathcal{G}(P^t AP)$.
2. A es una matriz irreducible si y solamente si $\mathcal{G}(A)$ es fuertemente conexo.

Demostración.

1. El grafo dirigido asociado a la matriz $P^t AP$ se obtiene del grafo de la matriz A mediante una reordenación de los vértices, y esto no afecta al carácter fuertemente conexo.

2. Si la matriz A es reducible, entonces existe una matriz de permutación P tal que

$$B = P^t AP = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

con A_{11}, A_{22} de orden $r < n$. Consideremos el grafo $\mathcal{G}(B)$ de esta matriz. Observemos que no se puede conectar el vértice $r+1$ con el vértice r, ya que cualquier camino dirigido que comience en $r+1$ solamente conecta vértices mayores o iguales que $r+1$, y cualquier camino dirigido que finalice en r solamente conecta vértices menores o iguales que r. Para que existiese un camino dirigido de $r+1$ a r tendría que haber una flecha (i, j) con $i \geq r+1$ y $j \leq r$, lo cual no es posible porque $b_{ij} = 0$ si $i \geq r+1$ y $j \leq r$. Entonces el grafo $\mathcal{G}(P^t AP)$ no es fuertemente conexo para alguna matriz de permutación, y por el apartado anterior, $\mathcal{G}(A)$ no es fuertemente conexo.

El recíproco se deduce de manera similar. Si el grafo $\mathcal{G}(A)$ no es fuertemente conexo, existe un par de vértices i, j que no se pueden conectar mediante un camino dirigido. Asignando nuevos nombres a los vértices, supongamos que desde el vértice n no se puede alcanzar el vértice 1. Si hay otros vértices, aparte de n, inaccesibles desde n, los etiquetamos como $2, \ldots, r$, para que todos los nodos inaccesibles desde n, con la posible excepción del propio n, sean $\bar{n} = \{1, 2, \ldots, r\}$. Etiquetamos los restantes vértices, accesibles desde n como $\underline{n} = \{r+1, \ldots, n-1\}$. Ningún vértice en $\bar{n}$ puede ser accesible desde un vértice de $\underline{n}$, porque en otro caso los vértices de $\bar{n}$ serían accesibles desde n a través de vértices de $\underline{n}$. En otras palabras, si $r+k \in \underline{n}$ y $r+k \to i \in \bar{n}$, entonces $n \to r+k \to i$, que es imposible. Entonces $a_{ij} = 0$ para $i = r+1, r+2, \ldots, n-1$ y $j = 1, 2, \ldots, r$. Esto significa que, tras un etiquetado, la matriz A sería de la forma $A = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}$, con X de orden $r \times r$ y Z de orden $(n-r) \times (n-r)$. Por tanto, A es reducible.

□

Ejemplo 11.2.2. Consideremos el grafo de los vuelos entre las islas Canarias gestionados por una compañía aérea. Es claro que desde cualquier vértice se puede acceder a otro mediante alguna de las aristas, por lo que el grafo es fuertemente conexo. Sin embargo, es fácil modificar el grafo para que ya no sea fuertemente conexo. Por ejemplo, si se elimina el vuelo de Tenerife Sur a Gran Canaria, desde Tenerife Sur no es posible acceder a otro punto. En cambio, si eliminamos los vuelos entre Tenerife Norte y Gran Canaria, el grafo sigue siendo fuertemente conexo. Esto se traduce en el carácter irreducible o no de la matriz de adyacencia del grafo correspondiente.

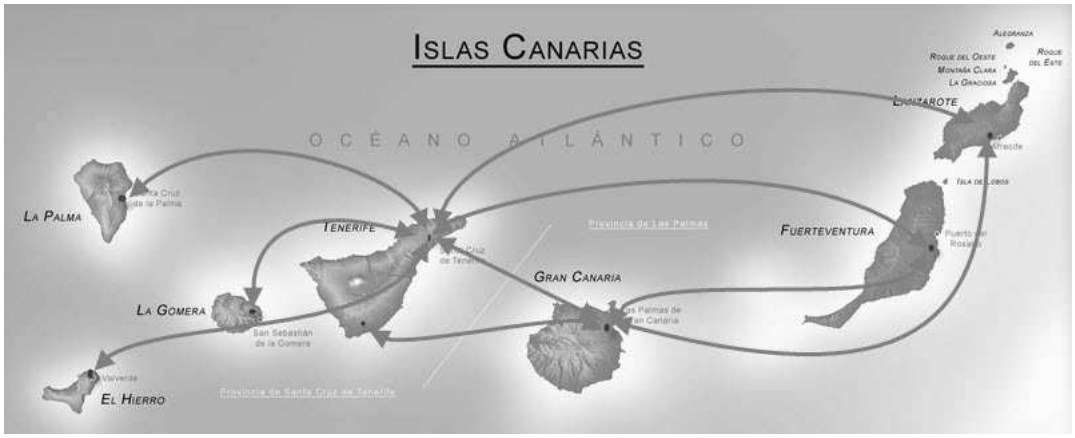

Figura 11.2: Vuelos internos en las islas Canarias

11.3. Teorema de Perron-Frobenius

Par autovalor/autovector no negativo

Sea $A_{n\times n} \geq 0$ una matriz no negativa, $\mathscr{L} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \geq 0, \boldsymbol{x} \neq \boldsymbol{0}\}$, y $\rho : \mathscr{L} \subset \mathbb{R}^n \to \mathbb{R}$, la función definida por

$$\rho(\boldsymbol{x}) = \min_{1\leq i\leq n} \left\{ \frac{[A\boldsymbol{x}]_i}{x_i} \mid x_i \neq 0 \right\}.$$

Para todo $\boldsymbol{x} \in \mathscr{L}$ se verifica que

1. $\rho(\boldsymbol{x}) \geq 0$.
2. $\rho(\boldsymbol{x})x_i \leq [A\boldsymbol{x}]_i = \sum_{j=1}^n a_{ij}x_j$, para todo $i = 1,\ldots,n$.
3. $A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x} \geq 0$, y además $\rho(\boldsymbol{x})$ es el mayor número con esta propiedad.
4. Si A es irreducible, $\boldsymbol{x} \in \mathscr{L}$ y $\boldsymbol{y} = (I_n + A)^{n-1}\boldsymbol{x}$, entonces $\rho(\boldsymbol{y}) \geq \rho(\boldsymbol{x})$.
5. Si A es irreducible, existe $\boldsymbol{v} > 0$ tal que $\rho(\boldsymbol{v}) = \max\{\rho(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathscr{L}\}$.

Demostración. 1. Como $\boldsymbol{x} \in \mathscr{L}$, se tiene que $[A\boldsymbol{x}]_i \geq 0, x_i \geq 0$ para todo $i = 1,\ldots,n$, y tenemos que $\rho(\boldsymbol{x}) \geq 0$.

2. Si $x_i \neq 0$, se sigue que $\rho(\boldsymbol{x})x_i \leq [A\boldsymbol{x}]_i$, pues es el mínimo, y si $x_i = 0$ es trivial.

3. Como $[A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x}]_i \geq 0$, el vector $A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x}$ es no negativo. Si M es un número real tal que $A\boldsymbol{x} - M\boldsymbol{x} \geq 0$, entonces para todo $i = 1, \ldots, n$ se tiene que $Mx_i \leq [A\boldsymbol{x}]_i$. Para todas las componentes $x_i \neq 0$ se tiene entonces que
$$M \leq \frac{[A\boldsymbol{x}]_i}{x_i},$$
esto es, M es una cota inferior. Entonces $M \leq \rho(\boldsymbol{x})$, lo que prueba el resultado.

4. Supongamos que A es irreducible. Tenemos que
$$A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x} \geq 0.$$
Multiplicamos ambos lados de la desigualdad por $(I_n + A)^{n-1}$, y obtenemos
$$A(I_n + A)^{n-1}\boldsymbol{x} - \rho(\boldsymbol{x})(I_n + A)^{n-1}\boldsymbol{x} \geq 0,$$
pues A y $(I_n + A)^{n-1}$ conmutan. Entonces
$$A\boldsymbol{y} - \rho(\boldsymbol{x})\boldsymbol{y} \geq 0,$$
pero el apartado anterior nos decía que $\rho(\boldsymbol{y})$ es el mayor número que verifica la desigualdad anterior. Entonces
$$\rho(\boldsymbol{y}) \geq \rho(\boldsymbol{x}).$$

5. Supongamos que A es irreducible. En primer lugar, observamos que $\rho(\alpha\boldsymbol{x}) = \rho(\boldsymbol{x})$ para todo $\boldsymbol{x} \in \mathscr{L}$ y $\alpha > 0$. Por tanto, a la hora de calcular el supremo de $\{\rho(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathscr{L}\}$, podemos restringirnos al conjunto
$$\mathscr{M} = \{\boldsymbol{x} \in \mathscr{L} \mid \sum_{i=1}^{n} x_i^2 = 1\} = \mathscr{L} \cap S^n,$$
que es un compacto de $\mathbb{R}^n$. La función ρ no es, en general, continua en $\mathscr{M}$. Por ejemplo, consideremos la matriz
$$A = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}.$$

Esta matriz es irreducible, y tomemos $\boldsymbol{x}(\epsilon) = \frac{1}{\sqrt{1+\epsilon^2}}(1,0,\epsilon)^t$, para $\epsilon > 0$. Entonces $\boldsymbol{x}(\epsilon) \in \mathscr{M}$, y

$$\rho(\boldsymbol{x}(\epsilon)) = \text{mín}\{\frac{2+\epsilon}{1}, \frac{\epsilon}{\epsilon}\} = 1.$$

Sin embargo,

$$\rho(\boldsymbol{x}(0)) = 2 \neq 1 = \lim_{\epsilon \to 0} \rho(\boldsymbol{x}(\epsilon)).$$

Por tanto, necesitamos restringirnos a un compacto donde ρ sea continua. Consideremos entonces el conjunto $\mathscr{N} = \{(I_n + A)^{n-1}\boldsymbol{x} \mid \boldsymbol{x} \in \mathscr{M}\}$. Por la ecuación **??**, todo elemento de $\mathscr{N}$ es un vector positivo, por lo que $\mathscr{N} \subset \mathscr{L}$. Además, $\mathscr{N}$ es una imagen continua de $\mathscr{M}$, por lo que es un compacto, y ρ es continua en $\mathscr{N}$, pues no hay denominadores nulos. Por tanto, ρ alcanza un máximo $\boldsymbol{y}^0$ en $\mathscr{N}$. Sea $\boldsymbol{x}^0 = \frac{1}{\|\boldsymbol{y}^0\|_2}\boldsymbol{y}^0 \in \mathscr{M}$. Sea $\boldsymbol{x}$ cualquier vector de $\mathscr{M}$. Si $\boldsymbol{y} = (I_n + A)^{n-1}\boldsymbol{x}$, entonces

$$\begin{aligned}\rho(\boldsymbol{x}) &\leq \rho(\boldsymbol{y}) \text{ por el apartado anterior} \\ &\leq \rho(\boldsymbol{y}^0) \text{ por maximalidad de } \boldsymbol{y}^0 \text{ en } \mathscr{N} \\ &= \rho(\boldsymbol{x}^0).\end{aligned}$$

Como $\boldsymbol{x}$ era un vector arbitrario de $\mathscr{M}$, se sigue que ρ tiene un máximo absoluto en $\boldsymbol{x}^0$.

□

Nota 11.3.1. La función ρ se denomina función de Collatz-Wielandt asociada a la matriz A.

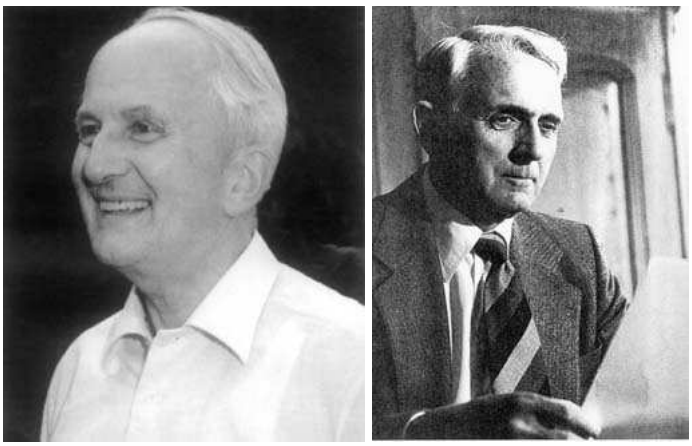

Figura 11.3: L. Collatz (1910-1990), H. Wielandt (1910-2001)

Nota 11.3.2. Puede existir más de un vector positivo en $\mathscr{L}$ donde la función ρ alcance su valor máximo. A tales vectores se les denomina **vectores extremales** de A.

Teorema de Perron-Frobenius

Sea $A_{n\times n}$ no negativa e irreducible, y $r = \max\{\rho(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathscr{L}\}$.

1. r es un número real positivo.
2. Si $\boldsymbol{x} \geq 0$ es un vector no nulo tal que $A\boldsymbol{x} - r\boldsymbol{x} \geq 0$, entonces $\boldsymbol{x}$ es autovector de r, $\boldsymbol{x} > 0$, y r es autovalor de A.
3. A tiene, al menos, un autovalor r real y positivo, con un autovector asociado $\boldsymbol{v} > 0$.
4. Si $\boldsymbol{x}$ es autovector de A asociado a r, entonces todas sus componentes son no nulas.
5. $r \geq |\lambda|$ para cualquier autovalor λ de A.
6. El autovalor r tiene multiplicidad algebraica igual a 1, y por tanto el espacio de autovectores asociado es de dimensión 1.

El único vector definido por

$$A\boldsymbol{v} = r\boldsymbol{v}, \boldsymbol{v} > 0 \text{ y } \|\boldsymbol{v}\|_1 = 1$$

se denomina vector de Perron, y a r se le llama autovalor de Perron.

Demostración. 1. Por la definición de r, tenemos que $r \geq \rho(\boldsymbol{e})$, donde $\boldsymbol{e} = (1,\ldots,1)^t$. Entonces

$$r \geq \rho(\boldsymbol{e}) = \min_i \frac{[A\boldsymbol{e}]_i}{1} = \min_i \sum_{j=1}^{n} a_{ij} > 0,$$

porque A no puede tener una fila de ceros, ya que es irreducible.

2. Supongamos que $A\boldsymbol{x} - r\boldsymbol{x} \neq \boldsymbol{x}$. Entonces

$$(I_n + A)^{n-1}(A\boldsymbol{x} - r\boldsymbol{x}) > 0, \text{ por } \textbf{??}.$$

Esto significa que $A\boldsymbol{y} - r\boldsymbol{y} > 0$ para $\boldsymbol{y} = (I_n + A)^{n-1}$, pues A y $(I_n + A)^{n-1}$ conmutan. Como tenemos una desigualdad estricta, existe $\epsilon > 0$ tal que

$$A\boldsymbol{y} - (r+\epsilon)\boldsymbol{y} \geq 0.$$

Pero entonces $\rho(\boldsymbol{y}) \geq r+\epsilon$, o bien que $\rho(\boldsymbol{y}) > r$, lo que contradice el carácter máximo de r. Por tanto, $\boldsymbol{x}$ es autovalor no negativo asociado a r.

Nos falta comprobar que $\boldsymbol{x}$ es positivo. Si $\boldsymbol{x}$ tuviera k coordenadas nulas, con $1 \leq k < n$, entonces $(1+r)\boldsymbol{x}$ tendría k coordenadas nulas también. Pero

$$(1+r)\boldsymbol{x} = (I_n + A)\boldsymbol{x},$$

que tiene menos de k componentes nulas. Por tanto, $\boldsymbol{x}$ es un vector positivo.

3. Consideremos $\boldsymbol{v} \in \mathscr{L}$ tal que $\rho(\boldsymbol{v}) = r$, pues en algún vector se alcanza el máximo. Entonces $A\boldsymbol{v} - r\boldsymbol{v} \geq 0$, y por el apartado anterior, $\boldsymbol{v}$ es un autovector positivo asociado a r.

4. Sea $\boldsymbol{x}$ autovector de A asociado a r, y llamemos $|\boldsymbol{x}|$ al vector que se obtiene tomando el módulo de cada componente. Entonces $|\boldsymbol{x}| \geq 0$ y es no nulo. Además,

$$r|\boldsymbol{x}| = |r\boldsymbol{x}| = |A\boldsymbol{x}| \leq A|\boldsymbol{x}|,$$

es decir, $A|\boldsymbol{x}| - r|\boldsymbol{x}| \geq 0$. Por el primer apartado, $|\boldsymbol{x}|$ es autovector asociado a r, y $|\boldsymbol{x}| > 0$, por lo que todas las componentes de $\boldsymbol{x}$ son no nulas.

5. Sea ahora λ un autovalor de A, y $\boldsymbol{u}$ un autovector asociado. Entonces

$$\lambda u_i = \sum_{j=1}^{n} a_{ij} u_j, i = 1, \ldots, n,$$

de donde

$$|\lambda||u_i| \leq \sum_{j=1}^{n} a_{ij}|u_j|.$$

En notación vectorial, podemos escribir $|\lambda||\boldsymbol{u}| \leq A|\boldsymbol{u}|$, donde $|\boldsymbol{u}|$ es el vector de componentes iguales a los módulos de las componentes de $\boldsymbol{u}$. Por la propiedad de maximalidad de ρ, se sigue que

$$|\lambda| \leq \rho(|\boldsymbol{u}|) \leq r.$$

6. Nos falta ver que r es un autovalor simple de A. En primer lugar, probaremos que el espacio de autovectores $\text{null}(rI_n - A)$ asociado a r es de dimensión 1, y a continuación veremos que $\text{null}(rI_n - A)^2 = \text{null}(rI_n - A)$, por lo que la multiplicidad algebraica del autovalor r es igual a 1. Supongamos que $\boldsymbol{x} = (x_1, \ldots, x_n)^t, \boldsymbol{y} = (y_1, \ldots, y_n)^t$ son autovectores de A asociados a r. Entonces tienen todas sus componentes no nulas, como hemos

visto antes. El vector $y_1\boldsymbol{x} - x_1\boldsymbol{y}$ está en $\text{null}(rI_n - A)$, es decir, es autovector asociado a r pero su primera coordenada es nula, por lo que la única posibilidad es que sea el vector nulo, esto es, $\boldsymbol{y}$ depende linealmente de $\boldsymbol{x}$. Sea entonces $\text{null}(rI_n - A) = \langle \boldsymbol{v} \rangle$, con $\boldsymbol{v} > 0$.

Veamos ahora que $\text{null}(rI_n - A) = \text{null}(rI_n - A)^2$. La inclusión $\subset$ la tenemos siempre. Sea $\boldsymbol{u} \in \text{null}(rI_n - A)^2$. Entonces $(rI_n - A)\boldsymbol{u} \in \text{null}(rI_n - A)$, por lo que existe $\alpha \in \mathbb{R}$ tal que

$$(rI_n - A)\boldsymbol{u} = \alpha \boldsymbol{v}.$$

Consideremos un autovector $\boldsymbol{w}$ de A^t asociado a r. Por los mismos argumentos anteriores, podemos suponer que $\boldsymbol{w} > 0$. Entonces

$$0 = \boldsymbol{w}^t(rI_n - A)\boldsymbol{u} = \boldsymbol{w}^t \alpha \boldsymbol{v} = \alpha(\boldsymbol{w}^t \boldsymbol{v}),$$

y por el carácter positivo de $\boldsymbol{v}$ y $\boldsymbol{w}$ se sigue que $\alpha = 0$, con lo que tenemos el resultado.

□

Nota 11.3.3. Necesitamos un resultado sobre números complejos, que necesitaremos en las siguientes pruebas. Sean z_1, z_2 números complejos no nulos tales que $|z_1 + z_2| = |z_1| + |z_2|$. entonces existe $\alpha > 0$ tal que $z_2 = \alpha z_1$. Sean $z_1 = a_1 + ia_2, z_2 = b_1 + ib_2$. De la hipótesis, se tiene que

$$((a_1 + b_1)^2 + (a_2 + b_2)^2)^{1/2} = (a_1^2 + a_2^2)^{1/2} + (b_1^2 + b_2^2)^{1/2},$$

que tras elevar al cuadrado y desarrollar nos lleva a

$$a_1 b_1 + a_2 b_2 = (a_1^2 + a_2^2)^{1/2}(b_1^2 + b_2^2)^{1/2}.$$

Sean $\boldsymbol{x}_1 = (a_1, a_2)^t, \boldsymbol{x}_2 = (b_1, b_2)^t$ vectores de $\mathbb{R}^2$. Lo anterior nos dice que $\boldsymbol{x}_1 \bullet \boldsymbol{x}_2 = \|\boldsymbol{x}_1\| \|\boldsymbol{x}_2\|$, y por la desigualdad CBS se deduce que $\boldsymbol{x}_2 = \alpha \boldsymbol{x}_1$, con $\alpha = \frac{\boldsymbol{x}_1 \bullet \boldsymbol{x}_2}{\|\boldsymbol{x}_1\|} = \frac{\|\boldsymbol{y}\|}{\|\boldsymbol{x}\|} > 0$. Entonces $\boldsymbol{z}_2 = \alpha \boldsymbol{z}_1$. En general, por inducción, si $\left|\sum_{i=1}^n z_i\right| = \sum_{i=1}^n |z_i|$ para $z_1, \ldots, z_n$ números complejos no nulos, entonces $z_i = \alpha_i z_1, i = 2, \ldots, n$, con $\alpha_i > 0$.

De lo anterior se tiene que $\alpha_i = \frac{|z_i|}{|z_1|}$, por lo que $z_i = \frac{z_1}{|z_1|}|z_i| = \theta |z_i|$, con θ un número complejo de módulo 1.

Máximo autovalor

Sea $A_{n\times n}$ una matriz no negativa e irreducible, y r su autovalor de Perron. Si A tiene una fila de elementos no nulos, entonces $|\lambda| < r$ para todo autovalor de A distinto de r.

Demostración. Supongamos que todos los elementos de la primera fila de A son no nulos, y sea λ un autovalor de A con $|\lambda| = r$, con $\boldsymbol{u}$ autovector asociado. Entonces

$$r|\boldsymbol{u}| = |\lambda \boldsymbol{u}| = |A\boldsymbol{u}| \leq A|\boldsymbol{u}|,$$

por la desigualdad triangular. Entonces el vector $|\boldsymbol{u}|$ es autovector asociado a r, $|\boldsymbol{u}| > 0$, y

$$|A\boldsymbol{u}| = |\lambda \boldsymbol{u}| = |\lambda||\boldsymbol{u}| = r|\boldsymbol{u}| = A|\boldsymbol{u}|.$$

Si nos fijamos en la primera fila de A en la igualdad anterior, nos queda que

$$\left| \sum_{j=1}^{n} a_{1j} u_j \right| = \sum_{j=1}^{n} a_{1j} |u_j|.$$

Como $a_{1j} \neq 0$ para todo $j = 1, \ldots, n$, se sigue que existe j_0 tal que el vector $\boldsymbol{u}$ es de la forma $\boldsymbol{u} = u_{j_0}(s_1, s_2, \ldots, s_n)^t$, con $u_{j_0} \neq 0, s_i > 0$. Entonces $\boldsymbol{u} = \alpha \boldsymbol{v}$, con $\boldsymbol{v} \geq 0$, y $|\boldsymbol{u}| = |\alpha|\boldsymbol{v}$, de donde $\boldsymbol{v}$ es autovector asociado a r. Por tanto, $\boldsymbol{u}$ también lo es, por lo que $\lambda = r$. □

Observemos que lo anterior se aplica a las matrices positivas.

11.4. Matrices primitivas

Matrices primitivas

Una matriz A no negativa irreducible que tenga un único autovalor $r = \rho(A)$ en su circunferencia espectral, se denomina **matriz primitiva**.
Una matriz A no negativa e irreducible con $r = \rho(A)$ es primitiva si y solamente si existe el límite

$$\lim_{k \to \infty} \left(\frac{A}{r} \right)^k > \mathbf{0},$$

y en tal caso $\lim_{k\to\infty} \left(\frac{A}{r}\right)^k = \frac{\boldsymbol{v}\boldsymbol{w}^t}{\boldsymbol{w}^t\boldsymbol{v}}$, donde $\boldsymbol{v}$ es vector de Perron de A y $\boldsymbol{w}$ es vector de Perron de A^t.

Demostración. El teorema de Perron-Frobenius asegura que $1 = \rho(A/r)$ es un autovalor simple de A/r, y es claro que A es primitiva si y solamente si A/r es primitiva. En otras palabras, A es primitiva si y solamente si $1 = \rho(A/r)$ es el único autovalor en el círculo unidad, que es equivalente a decir que existe el límite $\lim(A/r)^k$, y es positivo.

En este caso, si aplicamos lo que probamos en 8.3.2 a la matriz $B = \frac{1}{r}A$, tenemos que

$$\lim B^k = \boldsymbol{v}_1 \boldsymbol{w}_1^t,$$

donde $\boldsymbol{v}_1$ es autovector de B asociado a 1, $\boldsymbol{w}_1$ es autovector de B^t asociado a 1, y $\boldsymbol{w}_1^t \boldsymbol{v}_1 = 1$. Entonces

$$B\boldsymbol{v}_1 = \boldsymbol{v}_1, B^t \boldsymbol{w}_1 = \boldsymbol{w}_1 \text{ implica que } A\boldsymbol{v}_1 = r\boldsymbol{v}_1, A\boldsymbol{w}_1 = r\boldsymbol{w}_1.$$

Podemos suponer $\boldsymbol{v}_1, \boldsymbol{w}_1 > 0$. Entonces los vectores de Perron asociados son

$$\boldsymbol{v} = \frac{1}{\|\boldsymbol{v}_1\|_1}\boldsymbol{v}_1, \boldsymbol{w} = \frac{1}{\|\boldsymbol{w}_1\|_1}\boldsymbol{w}_1.$$

De $\boldsymbol{w}_1^t \boldsymbol{v}_1 = 1$ se deduce que $\|\boldsymbol{v}_1\|_1 \|\boldsymbol{w}_1\| \boldsymbol{w}^t \boldsymbol{v} = 1$, y

$$\lim \left(\frac{A}{r}\right)^k = \boldsymbol{v}_1 \boldsymbol{w}_1^t = \|\boldsymbol{v}_1\|_1 \boldsymbol{v} \|\boldsymbol{w}_1\|_1 \boldsymbol{w} = \frac{\boldsymbol{v}\boldsymbol{w}^t}{\boldsymbol{w}^t \boldsymbol{v}}.$$

□

Nota 11.4.1. Si A es no negativa e irreducible, con una fila de elementos no nulos, entonces A es primitiva. De aquí se deduce que toda matriz positiva es primitiva.

Test de Frobenius para matrices primitivas

Una matriz $A \geq 0$ es primitiva si y solamente si $A^m > 0$ para algún $m > 0$.

Demostración. Supongamos que $A^m > 0$ para algún $m > 0$. Esto implica que A es irreducible. En otro caso, existiría una matriz de permutación tal que

$$A = P \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix} P^t, \text{ de donde } A^m = P \begin{pmatrix} X^m & * \\ 0 & Z^m \end{pmatrix} P^t,$$

y A^m tendría entradas nulas.

Supongamos ahora que A tiene h autovalores en su circunferencia espectral, de tal forma que $r = \rho(A) = |\lambda_1| = \cdots = |\lambda_h| > |\lambda_{h+1}| \geq \cdots \geq |\lambda_n|$. Si λ es un autovalor de A de multiplicidad algebraica k, entonces λ^m es autovalor de A^m, con la misma multiplicidad algebraica. Entonces $\lambda_k^m, 1 \leq k \leq h$ es un autovalor de A^m que está en su círculo espectral, con multiplicidad algebraica igual a la de λ_k en A. Como A^m es irreducible, el teorema de Perron-Frobenius garantiza

que A^m tiene un único autovalor, que debe ser r^m, en su círculo espectral, por lo que $r^m = \lambda_1^m = \cdots = \lambda_h^m$. Pero esto significa que la multiplicidad algebraica de r^m, que es la de r, es igual a h. Por tanto, $h = 1$.

Recíprocamente, si A es primitiva, con $r = \rho(A)$, entonces

$$\lim_{k\to\infty}\left(\frac{A}{r}\right)^k > \mathbf{0},$$

por lo que existe un $m > 0$ tal que $(A/r)^m > 0$, y entonces $A^m > 0$. □

El cálculo de $A^2, A^3, \ldots$ puede ser muy laborioso. Existe un teorema de Wielandt que nos dice que $A_{n\times n}$ es primitiva si y solamente si $A^{n^2-2n+2} > 0$. Además, $n^2 - 2n + 2$ es el menor exponente que funciona para la clase de matrices primitivas $n \times n$ que tienen todo ceros en la diagonal.

Ejemplo 11.4.2. Queremos determinar si la matriz

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 3 & 4 & 0 \end{pmatrix}$$

es primitiva. Para aliviar en lo posible los cálculos, consideramos la matriz $B = \beta(A)$ definida como

$$b_{ij} = \begin{cases} 1 & \text{si } a_{ij} > 0, \\ 0 & \text{si } a_{ij} = 0, \end{cases}$$

Entonces $[B^k]_{ij} > 0$ si y solamente si $[A^k]_{ij} > 0$, para todo $k > 0$. Esto significa que en lugar de usar $A^2, A^3, \ldots$, para determinar el carácter primitivo, basta considerar

$$B_1 = \beta(A), B_2 = \beta(B_1B_1), B_3 = \beta(B_1B_2), \ldots,$$

sin ir más lejos de $n^2 - 2n + 2$. Todos estos cálculos solamente necesitan operaciones lógicas de tipo 'AND' y 'OR'. En nuestro caso,

$$B_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$
$$B_4 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad B_5 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Si calculamos los autovalores de A, tenemos $\lambda_1 = 3{,}1474, \lambda_2 = -2{,}3289, \lambda_3 = -0{,}8186$. El espacio de autovectores asociado a λ_1 es

$$\text{null}(A - \lambda_1 I) = \langle \boldsymbol{v}_1 = \begin{bmatrix} -0{,}167978 \\ -0{,}528699 \\ -0{,}832022 \end{bmatrix} \rangle,$$

y el vector de Perron es

$$\boldsymbol{v} = \frac{1}{\|\boldsymbol{v}_1\|_1}(-\boldsymbol{v}_1) = \begin{bmatrix} 0{,}109883 \\ 0{,}345849 \\ 0{,}544268 \end{bmatrix}.$$

Para el cálculo del límite $\lim\left(\frac{A}{\lambda_1}\right)^k$ necesitamos el autovector de Perron de A^t. En este caso,

$$\text{null}(A^t - \lambda_1 I) = \langle \boldsymbol{w}_1 = \begin{bmatrix} 0{,}455172 \\ 0{,}751514 \\ 0{,}477541 \end{bmatrix} \rangle,$$

y el vector de Perron es

$$\boldsymbol{w} = \frac{1}{\|\boldsymbol{w}_1\|_1}\boldsymbol{w}_1 = \begin{bmatrix} 0{,}270256 \\ 0{,}446207 \\ 0{,}283537 \end{bmatrix}.$$

Entonces

$$\lim\left(\frac{A}{\lambda_1}\right)^k = \frac{\boldsymbol{v}\boldsymbol{w}^t}{\boldsymbol{w}^t\boldsymbol{v}} = \begin{bmatrix} 0{,}087772 & 0{,}144916 & 0{,}092085 \\ 0{,}276256 & 0{,}456114 & 0{,}289833 \\ 0{,}434749 & 0{,}717793 & 0{,}456114 \end{bmatrix}.$$

Nota 11.4.3. Hemos visto en la prueba del test de Frobenius que toda matriz primitiva es irreducible. Sin embargo, no toda matriz irreducible es primitiva, como ocurre con la matriz

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Ahora damos una serie de resultados para obtener el número de autovalores de módulo máximo, y, como corolario, un criterio para determinar si una matriz es primitiva.

Teorema de Wielandt

Si $|C| \leq A$, con $A_{n\times n}$ irreducible, y r autovalor de Perron de A, entonces para cualquier autovalor s de C

$$|s| \leq r.$$

La igualdad se alcanza si y solamente si

$$C = \exp(i\varphi)DAD^{-1},$$

donde $s = \exp(i\varphi)r$ y $|D| = I_n$.

Demostración. Sea $C\boldsymbol{y} = s\boldsymbol{y}$, con $\boldsymbol{y} \neq \boldsymbol{0}$ un autovector asociado a s. Entonces

$$|C||\boldsymbol{y}| \geq |s||\boldsymbol{y}|,$$

por la desigualdad triangular. Como $A \geq |C|$, tenemos que

$$A|\boldsymbol{y}| \geq |s||\boldsymbol{y}|.$$

Mediante la función de Collatz-Wielandt ρ_A, sabemos que $|s| \leq \rho_A(|\boldsymbol{y}|) \leq r$, y tenemos la primera parte.

Supongamos que $C = \exp(i\varphi)DAD^{-1}$, con $|D| = I_n$. Entonces las matrices C y $\exp(i\varphi)A$ son semejantes, y si r es el autovalor maximal de A, entonces $r\exp(i\varphi)$ es autovalor de C.

Para la necesidad de la condición, sea $s = \exp(i\varphi)r$, y $\boldsymbol{y}$ un autovector asociado. Entonces $|\boldsymbol{y}|$ es autovector de r, $|\boldsymbol{y}| > 0$, y

$$r|\boldsymbol{y}| = A|\boldsymbol{y}| \geq |C||\boldsymbol{y}| \geq |s||\boldsymbol{y}| = r|\boldsymbol{y}|,$$

de donde $(A - |C|)\boldsymbol{y} = \boldsymbol{0}$. Como $A - |C| \geq 0$ y $|\boldsymbol{y}|$ es un vector positivo, se tiene que $A = |C|$. Definamos

$$D = \operatorname{diag}(\frac{y_1}{|y_1|}, \frac{y_2}{|y_2|}, \ldots, \frac{y_n}{|y_n|}),$$

y $G = (g_{ij}) = \exp(-i\varphi)D^{-1}CD$. Entonces de la igualdad $C\boldsymbol{y} = s\boldsymbol{y}$ se deduce

$$CD|\boldsymbol{y}| = sD|\boldsymbol{y}| = r\exp(i\varphi)D|\boldsymbol{y}|.$$

Entonces $G|\boldsymbol{y}| = r|\boldsymbol{y}|$, de donde $G|\boldsymbol{y}| = A|\boldsymbol{y}|$. Por la definición de G, $|G| = |C|$, y entonces $|G| = A$. En conclusión, tenemos que $|G||\boldsymbol{y}| = G|\boldsymbol{y}|$, que podemos escribir como

$$\sum_{j=1}^{n}(|g_{ij}| - g_{ij})|y_j| = 0, i = 1, 2, \ldots, n,$$

que implica que $|g_{ij}| - g_{ij} = 0$, para todo i, j, ya que $|y_{ij}| > 0$. En conclusión, $G = |G| = A$, y de la definición de G nos queda que $C = \exp(i\varphi)DAD^{-1}$. □

Autovalores maximales

Sea $A_{n\times n} \geq 0$ una matriz irreducible con autovalor de Perron r, y $\lambda_1, \ldots, \lambda_h$ los autovalores de A de módulo r. Entonces $\lambda_1, \ldots, \lambda_h$ son las raíces de $\lambda^h - r = 0$.

Demostración. Sea $\lambda_t = r\exp(i\varphi_t)$, $t = 1,2,\ldots,h$. Como $|\lambda_t| = r$, la condición de igualdad del teorema anterior, con $C = A$ y $s = \lambda_t$, tenemos que

$$A = \exp(i\varphi_t)D_t A D_t^{-1}, t = 1,2,\ldots,h.$$

Por tanto, A y $\exp(i\varphi_t)A$ son semejantes. Como r es un autovalor simple de A, se sigue que, para cada t, $\exp(i\varphi t)r = \lambda_t$ es un autovalor simple de $\exp(i\varphi_t)A$, y por tanto de A.

Ahora

$$\begin{aligned} A &= \exp(i\varphi_k)D_k A D_k^{-1} \\ &= \exp(i\varphi_k)D_k(\exp(i\varphi_s)D_s A D_s^{-1})D_k^{-1} \\ &= \exp(i\varphi_k + i\varphi_s)(D_k D_s)A(D_k D_s)^{-1}, \end{aligned}$$

y entonces $r\exp(i\varphi_k + i\varphi_s)$ también es un autovalor de A. Esto significa que $\exp(i\varphi_k + i\varphi_s)$ es uno de los números $\exp(i\varphi_t)$, y que el conjunto

$$\mathcal{G} = \{1, \exp(i\varphi_1), \ldots, \exp(i\varphi_{h-1})\}$$

es cerrado por la multiplicación, por lo que es un grupo de orden h. La potencia h-ésima de cualquier elemento es igual a 1, por lo que $\mathcal{G}$ son las raíces h-ésima de la unidad, y se sigue el resultado. □

Invariancia rotacional

El espectro de una matriz irreducible de índice h es invariante por una rotación de ángulo $2\pi/h$, pero no por un ángulo menor.

Demostración. Un autovalor λ pertenece al espectro de A si y solamente si $\lambda\exp(i2\pi/h)$ pertenece al espectro de $\exp(i2\pi/h)A$. Entonces el espectro de $\exp(i2\pi/h)A$ es el espectro de A rotado un ángulo $2\pi/h$. Como A y $\exp(i2\pi/h)A$ son semejantes, tienen el mismo espectro. ninguna rotación menor que $2\pi/h$ deja invariante al espectro de A, porque los autovalores de mayor módulo no se mantendrían invariantes. □

Matrices primitivas y polinomio característico

Sea $A_{n\times n}$ una matriz no negativa, irreducible, y escribamos

$$c(\lambda) = \det(\lambda I_n - A) = \lambda^n + c_{n_1}\lambda^{n-n_1} + c_{n_2}\lambda^{n-n_2} + \cdots + c_{n_t}\lambda^{n-n_t}$$

su polinomio característico, con $c_{n_1}, c_{n_2}, \ldots, c_{n_t}$ no nulos, y

$$n > n-n_1 > n-n_2 > \cdots > n-n_t \geq 0.$$

Entonces el índice h de A es igual al máximo común divisor de las diferencias

$$n_1, n_2, \ldots, n_t.$$

Demostración. Si $\lambda_1, \ldots, \lambda_n$ son los autovalores de A (incluidas las multiplicidades), entonces $\{\omega\lambda_1, \ldots, \omega\lambda_n\}$ son también autovalores de A, donde $\omega = \exp(i2\pi/h)$. Por el desarrollo de los coeficientes del polinomio característico en función de sus raíces, tenemos que

$$c_{k_j} = (-1)^{k_j} \sum_{1\leq i_1<\cdots<i_{k_j}\leq n} \lambda_{i_1}\ldots\lambda_{i_{k_j}} = (-1)^{k_j} \sum_{1\leq i_1<\cdots<i_{k_j}\leq n} \omega\lambda_{i_1}\ldots\omega\lambda_{i_{k_j}} = \omega^{k_j} c_{k_j},$$

de donde $\omega^{k_j} = 1$. Entonces, h divide a cada k_j. Si d divide a cada k_j, para $d > h$, entonces $\gamma^{-k_j} = 1$ para $\gamma = \exp(i2\pi/d)$. Entonces $\gamma\lambda$ es un autovalor de A si λ es autovalor de A, porque $c(\gamma\lambda) = 0$. Pero esto significa que el espectro de A es invariante por la rotación de un ángulo $2\pi/d < 2\pi/h$, que contradice el resultado anterior. □

Nota 11.4.4. Es importante disponer de criterios que permitan decidir si una matriz es o no primitiva, sin necesidad de calcular los autovalores. Existen métodos numéricos eficientes que permiten obtener el autovalor de mayor módulo de una matriz primitiva, así como un autovector asociado, sin necesidad de calcular los restantes autovalores. Son procedimientos iterados, como el método de la potencia y sus variantes, que no trataremos aquí. Una referencia es [**?**, p. 314] o el clásico [**?**, cap. 7].

11.5. Modelo de población de Leslie

Dividamos una población de hembras de una misma especie en distintos grupos de edad $G_1, G_2, \ldots, G_n$, donde cada grupo cubre el mismo número de

años. Así, si la vida más larga se estima en L años, la amplitud de cada grupo es de L/n años. El grupo G_1 está formado por los individuos cuya edad está en el intervalo $[0, L/n)$, es decir, los recién nacidos y los que tengan edad menor que L/n. En general, el grupo G_k está formado por los individuos de edad comprendida entre $(k-1)L/n$ y kL/n. Supongamos que los censos de población se realizan en intervalos de tiempo iguales a la amplitud de los grupos de edad.

Sea f_i el número promedio de hijas de cada hembra del grupo G_i (tasa de fecundidad de G_i), y s_i la fracción de individuos del grupo i que sobreviven al intervalo entre censos y pasan a formar parte del grupo G_{i+1} (tasa de supervivencia). Sea $p_i(j)$ el número de hembras del grupo G_i en el instante j. Entonces se verifican las siguientes relaciones:

$$\begin{aligned} p_1(j+1) &= p_1(j)f_1 + p_2(j)f_2 + \cdots + p_n(j)f_n, \\ p_2(j+1) &= p_1(j)s_1, \\ &\vdots \\ p_n(j+1) &= p_{n-1}(j)s_{n-1}. \end{aligned}$$

Además, el cociente

$$P_i(j) = \frac{p_i(j)}{p_1(j) + \cdots + p_n(j)}$$

es la proporción de miembros del grupo G_i en la población total, en el instante j.

El vector $\boldsymbol{P}(j) = (P_1(j), P_2(j), \ldots, P_n(j))^t$ representa la distribución de edades de la población en el instante j, y, suponiendo que existe, $\boldsymbol{P}^* = \lim_{j\to\infty} \boldsymbol{P}(j)$ es la distribución de edades de la población a largo plazo.

Las ecuaciones anteriores se pueden expresar en forma matricial como

$$\begin{pmatrix} p_1(j+1) \\ p_2(j+1) \\ \vdots \\ p_n(j+1) \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & \cdots & f_{n-1} & f_n \\ s_1 & 0 & \cdots & 0 & 0 \\ 0 & s_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-1} & 0 \end{pmatrix} \begin{pmatrix} p_1(j) \\ p_2(j) \\ \vdots \\ p_n(j) \end{pmatrix},$$

que en forma matricial es

$$\boldsymbol{p}(j+1) = L\boldsymbol{p}(j), \text{ donde } L = \begin{pmatrix} f_1 & f_2 & \cdots & f_{n-1} & f_n \\ s_1 & 0 & \cdots & 0 & 0 \\ 0 & s_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{n-1} & 0 \end{pmatrix}.$$

La matriz L se denomina matriz de Leslie, en honor de P.H. Leslie, que introdujo este modelo en 1945. La relación de recurrencia se puede resolver, y obtenemos

$$\boldsymbol{p}(j) = L^j \boldsymbol{p}(0) \text{ para todo } j > 0.$$

La matriz L es no negativa, pues $s_i > 0, i = 1, \ldots, n-1$, y $f_i \geq 0, i = 1, \ldots, n$. Si $n > 2$, el grafo de la matriz L es fuertemente conexo si y solamente si $f_n > 0$.

El polinomio característico de la matriz L es

$$\det(\lambda I_n - L) = \lambda^n - f_1\lambda^{n-1} - f_2 s_1 \lambda^{n-2} - \cdots - f_{n-1} s_1 \ldots s_{n-2}\lambda - f_n s_1 \ldots s_{n-1}.$$

Basta que haya dos términos f_i, f_{i+1} consecutivos no nulos para garantizar el carácter primitivo de L. De ahora en adelante supondremos que L es irreducible y primitiva.

Tenemos garantizada entonces la existencia de un autovalor real positivo r de L de multiplicidad algebraica igual a 1, con un autovector asociado $\boldsymbol{v} > 0$. Además, $|\lambda| < r$ para cualquier otro autovalor. Así, la matriz $\frac{1}{r}L$ tiene a 1 como autovalor máximo, y se le puede aplicar el resultado 8.3.2:

$$\lim_{j\to\infty} \frac{L^j}{r^j} = \boldsymbol{v}\boldsymbol{w}^t,$$

para algún $\boldsymbol{w} \in \mathbb{R}^n$, esto es, una matriz cuyas columnas son proporcionales a $\boldsymbol{v}$. Por otra parte, sea $\mathbf{1}$ el vector con todas sus componentes iguales a 1. La población total en un instante j es $\mathbf{1}^t\boldsymbol{p}(j)$, y podemos escribir

$$\begin{aligned}
\boldsymbol{P}^* &= \lim_{j\to\infty} P(j) = \lim_{j\to\infty} \frac{\boldsymbol{p}(j)}{\mathbf{1}^t\boldsymbol{p}(j)} = \lim_{j\to\infty} \frac{L^j\boldsymbol{p}(0)}{\mathbf{1}^t L^j \boldsymbol{p}(0)} \\
&= \lim_{j\to\infty} \frac{(\frac{L}{r})^j \boldsymbol{p}(0)}{\mathbf{1}^t(\frac{L}{r})^j\boldsymbol{p}(0)} \\
&= \frac{\boldsymbol{v}\boldsymbol{w}^t\boldsymbol{p}(0)}{\mathbf{1}^t\boldsymbol{v}\boldsymbol{w}^t\boldsymbol{p}(0)} = \frac{\boldsymbol{v}(\boldsymbol{w}^t\boldsymbol{p}(0))}{\mathbf{1}^t\boldsymbol{v}(\boldsymbol{w}^t\boldsymbol{p}(0))} \\
&= \frac{\boldsymbol{v}}{v_1 + \cdots + v_n},
\end{aligned}$$

que es el vector de Perron de L.

Ejemplo 11.5.1. Consideremos una población de salmones, dividida en tres clases de un año cada una. La clase 1 contiene los salmones en su primer año de vida, la clase 2 a los salmones entre 1 y 2 años, y la clase 3 a los salmones de más de dos años.

Supongamos que hay 1 000 hembras en cada una de las tres clases. Entonces

$$\boldsymbol{p}(0) = \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}.$$

Supongamos que la tasa de supervivencia del salmón en la primera clase es de 0,5 %, la tasa de supervivencia del salmón en la segunda clase es 10 %, y que cada hembra de la tercera clase produce 2000 hembras en su puesta. Entonces $s_2 = 0,005$, $s_3 = 0,10$ y $f_3 = 2000$. La matriz de Leslie es entonces

$$L = \begin{pmatrix} 0 & 0 & 2000 \\ 0,005 & 0 & 0 \\ 0 & 0,10 & 0 \end{pmatrix}.$$

Para calcular el vector de distribución por edad después de un año, usamos la ecuación $\boldsymbol{p}(1) = L\boldsymbol{p}(0)$. Vamos a emplear MATLAB para dicho cálculo. Primero, introducimos el vector de distribución de edad inicial y la matriz de Leslie.

```
>> p0=[1000;1000;1000];
>> L=[0,0,2000;0.005, 0,0;0,0.1,0]

L =

  1.0e+003 *

         0         0    2.0000
    0.0000         0         0
         0    0.0001         0
```

Notemos que MATLAB usa notación científica. El valor `1.0e+003` que precede a la matriz indica que debemos multiplicar cada entrada de la matriz por 1×10^3, es decir, hay que mover la coma decimal tres lugares a la derecha. Vamos a probar un nuevo formato para la salida (con `help format` se obtiene una lista completa de todas las posibilidades).

```
>> format short g
>> L=[0,0,2000;0.005, 0,0;0,0.1,0]

L =

            0             0          2000
        0.005             0             0
            0           0.1             0
```

El comando `format short g` indica a MATLAB que use el mejor entre formato fijo o en coma flotante, según cada entrada de la matriz. Ahora calculamos $\boldsymbol{p}(1)$ como sigue.

```
>> p1=L*p0

p1 =

     2000000
           5
         100
```

El vector de distribución de edad $\boldsymbol{p}(1)$ muestra que tras el primer año hay 2 000 000 de salmones en la primera clase, 5 en la segunda clase y 100 en la tercera clase. Procedemos ahora a calcular $\boldsymbol{p}(2)$, el vector de distribución por edad después de 2 años.

```
>> p2=L*p1

p2 =

       2e+005
        10000
          0.5
```

El mismo resultado lo tendríamos con

```
>> p2=L^2 *x0

p2 =

       2e+005
        10000
          0.5
```

El vector de distribución por edad $\boldsymbol{p}(2)$ indica que después de 2 años hay 200 000 salmones en la primera clase de edad, 10 000 en la segunda clase de edad y 0,5 en la tercera clase. En la realidad, es imposible tener medio salmón. Sin embargo, apartemos de momento esta cuestión y calculemos la población tras 3 años.

```
>> p3=L*p2
```

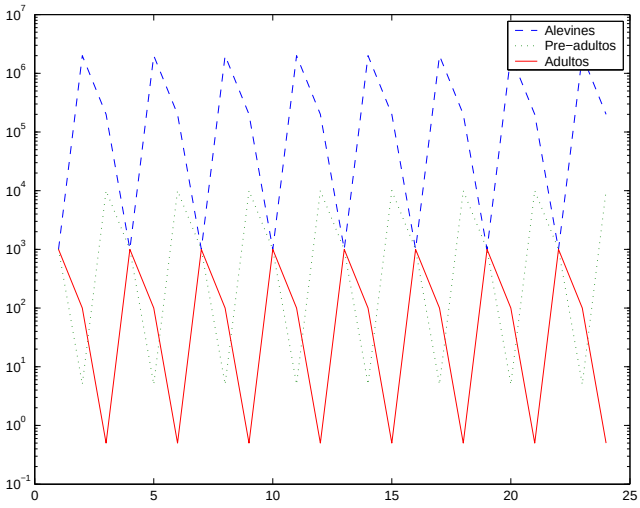

Figura 11.4: Población de salmones a lo largo del tiempo.

```
p3 =

        1000
        1000
        1000
```

Observemos que la población de salmones ha vuelto a su configuración original, con 1 000 peces en cada categoría. ¿Qué ha ocurrido? En este caso parece que hay un problema con el límite. Es fácil ver que L es irreducible. Sin embargo, su polinomio característico es

$$\det(\lambda I_3 - L) = \lambda^3 - 1,$$

por lo que no es primitiva. Los autovalores son las raíces cúbicas de la unidad, todas con norma 1. Este es el motivo de su comportamiento cíclico, de periodo tres.

Ejemplo 11.5.2. Consideremos ahora una población con tres clases de edad. Supongamos que cada hembra de la segunda y tercera clases producen una descendencia femenina de 4 y 3 miembros, respectivamente, en cada iteración. Supongamos además que el 50 % de las hembras de la primera clase sobreviven a la segunda clase, y que el 25 % de las hembras de la segunda clase llegan vivas a la tercera clase. La matriz de Leslie de esta población es

$$L = \begin{pmatrix} 0 & 4 & 3 \\ 0{,}5 & 0 & 0 \\ 0 & 0{,}25 & 0 \end{pmatrix}.$$

Supongamos que el vector inicial de población es

$$\boldsymbol{p}^{(0)} = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}.$$

Haremos los cálculos con MATLAB.

```
>> L=[0,4,3;0.5,0,0;0,0.25,0];
>> x0=[10;10;10];
```

Vamos a seguir los cambios en la población sobre un periodo de 10 años. Empezamos en el año cero y acabamos en el año 11. Hay tres clases que calcular en cada iteración. Empezamos creando una matriz que contendrá los datos de la población. La matriz tendrá tres filas, y cada fila contendrá los datos de una clase de edad. La matriz tendrá 11 columnas, y la primera de ellas tendrá el vector inicial de distribución por edad. Las diez restantes columnas almacenarán los vectores de distribución por edad en cada paso de la iteración (desde el año 1 hasta el año 10).

```
>> X=zeros(3,11);
```

Ponemos el vector inicial en la primera columna de la matriz X.

```
>> X(:,1)=x0;
```

Ahora usaremos la ecuación $\boldsymbol{p}(k) = L^k\boldsymbol{p}(0)$ para calcular el vector de distribución por edad en los siguientes 10 años. Estos diez vectores se pondrán en las columnas 2 a la 11 de la matriz X. En el paso k-ésimo, calculamos el vector de distribución por edad número k multiplicando el correspondiente $k-1$ por la matriz L. Esto admite el siguiente bucle `for`.

```
>> for k=2:11, X(:,k)=L*X(:,k-1);end
```

Podemos ver el resultado introduciendo la variable que contiene los datos.

```
>> X

X =

  1.0e+003 *

    0.0100    0.0700    0.0275    0.1437    0.0813    0.2978
    0.0100    0.0050    0.0350    0.0138    0.0719    0.0406
    0.0100    0.0025    0.0013    0.0088    0.0034    0.0180
```

```
    0.2164    0.6261    0.5445    1.3333    1.3238
    0.1489    0.1082    0.3130    0.2722    0.6667
    0.0102    0.0372    0.0271    0.0783    0.0681
```

Recordemos que el prefijo `1.0e+003` significa que cada número en la salida debe multiplicarse por 10^3. Para el resto de la actividad, usaremos otro formato.

```
>> format short g
>> X

X =

  Columns 1 through 7

    10       70        27.5      143.75     81.25     297.81      216.41
    10        5        35         13.75     71.875     40.625     148.91
    10        2.5       1.25       8.75      3.4375    17.969     10.156

  Columns 8 through 11

  626.09      544.49      1333.3        1323.8
  108.2       313.05       272.25        666.67
  37.227       27.051       78.262        68.062
```

La distribución de población en cada año aparece como un vector columna de la matriz X. La gráfica de la evolución de la población a lo largo del tiempo, que aparece en la figura **??**, se puede obtener como sigue.

```
>> t=0:10;
>> plot(t,X')
>> xlabel('Tiempo')
>> ylabel('Poblaci\'{o}n')
```

El gráfico se aclara si añadimos una leyenda a cada color.

```
>> legend('Primera clase de edad','Segunda clase de edad', ...
'Tercera clase de edad')
```

Observemos que el número de hembras en cada grupo de edad en la figura **??** se incrementa con el tiempo, con cierto comportamiento oscilatorio. Podemos dibujar el logaritmo de la población a lo largo del tiempo, tal como aparece en la figura **??**, con la siguiente secuencia de comandos.

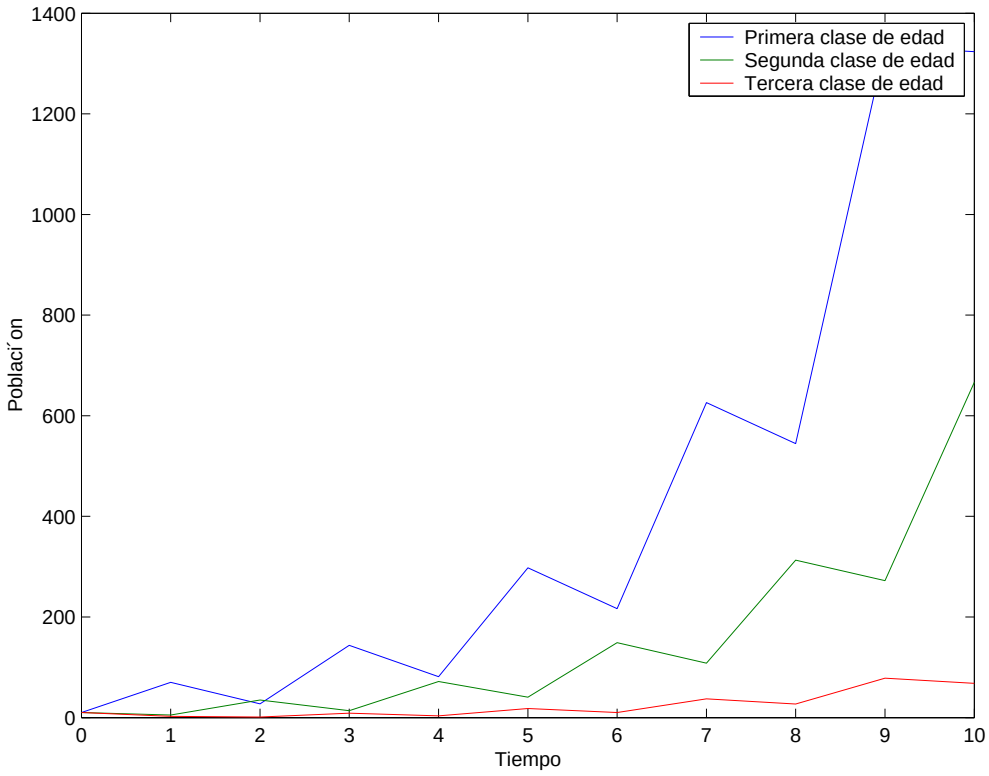

Figura 11.5: Evolución de la población.

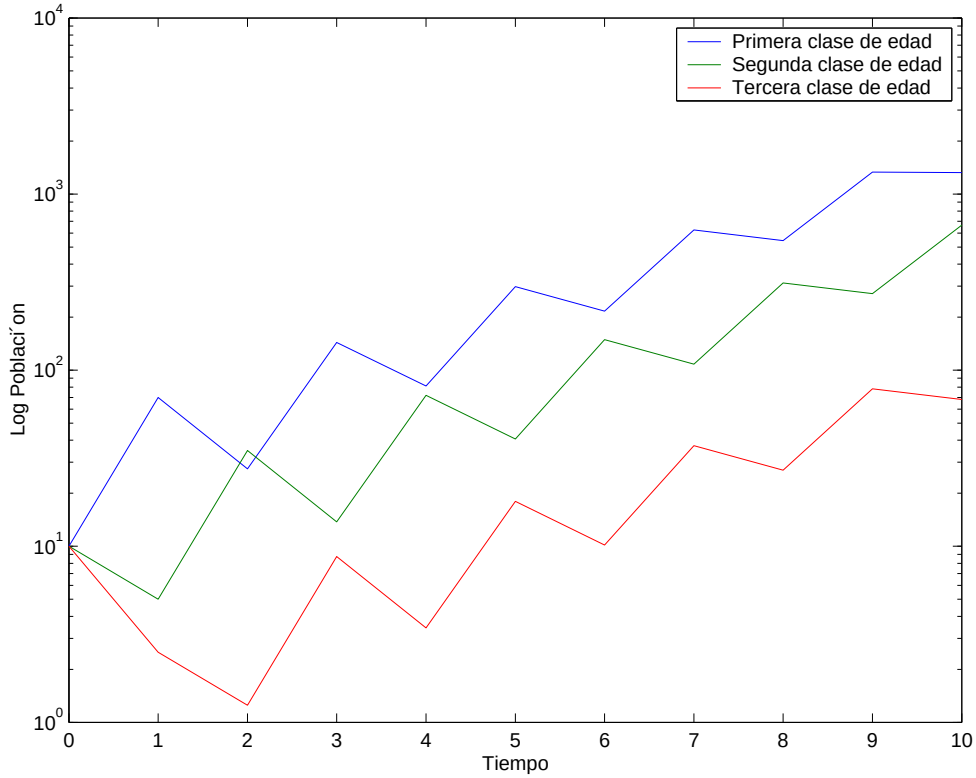

Figura 11.6: Evolución de log de la población.

```
>> t=(0:10);
>> semilogy(t,X')
>> xlabel('Tiempo')
>> ylabel('Log Poblaci´on')
>> legend('Primera clase de edad','Segunda clase de edad', ...
'Tercera clase de edad')
```

Para comprender el comportamiento a largo plazo de la población, calculamos los autovalores y autovectores de L.

```
>> L=[0,4,3;.5,0,0;0,.25,0]

L =

         0         4         3
       0.5         0         0
         0      0.25         0
```

```
>> [V,D]=eig(L)

V =

     -0.94737      0.93201      0.22588
     -0.31579       -0.356     -0.59137
    -0.052632     0.067989      0.77412


D =

          1.5            0            0
            0       -1.309            0
            0            0     -0.19098
```

En este caso, vemos que $\lambda_1 = 1{,}5$ es el autovalor dominante, y un autovector asociado es

$$\boldsymbol{v}_1 = \begin{pmatrix} -0{,}94737 \\ -0{,}31579 \\ -0{,}052632 \end{pmatrix},$$

que es la primera columna de la matriz V. Análogamente, $\lambda_2 = -1{,}309, \lambda_3 = -0{,}19098$ y sus autovectores asociados son la segunda y tercera columna de la matriz V.

$$\boldsymbol{v}_2 = \begin{pmatrix} 0{,}93201 \\ -0{,}356 \\ -0{,}067989 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0{,}22588 \\ -0{,}59137 \\ 0{,}77412 \end{pmatrix}.$$

Por lo que sabemos, el límite de las proporciones de cada clase de edad sobre la población total es igual a $\boldsymbol{v}_1 / \sum_{j=1}^{n} v_{1j}$. En este caso podemos calcular

```
>> v1=V(:,1)

v1 =

   -0.9474
   -0.3158
   -0.0526

>> v1/sum(v1)

ans =
```

```
    0.7200
    0.2400
    0.0400
```

Por tanto, la primera clase de edad compondrá el 72 % de la población, la segunda clase el 24 % y la tercera clase el 4 % de la población total. Vamos a comprobar que, en efecto, el comportamiento a largo plazo de la población sigue este esquema.

```
>> L=[0,4,3;.5,0,0;0,.25,0]

L =

         0    4.0000    3.0000
    0.5000         0         0
         0    0.2500         0

>> x0=[10;10;10]

x0 =

    10
    10
    10

>> x100=L^100*x0

x100 =

  1.0e+019 *

    1.1555
    0.3852
    0.0642

>> x=x100/sum(x100)

x =

    0.7200
    0.2400
    0.0400
```

Lo anterior ha calculado el porcentaje de población de cada clase de edad tras 100 años. Vemos que coincide con lo que habíamos deducido a partir de v_1.

Vamos a dibujar la evolución de los porcentajes de cada clase de edad en los primeros 100 años. Primero almacenamos los vectores de distribución por edad.

```
>> X=zeros(3,101);
>> X(:,1)=[10;10;10];
>> for k=2:101,X(:,k)=L*X(:,k-1);end
```

Ahora podemos obtener los porcentajes de cada clase de edad sobre la población total dividiendo cada columna por su suma.

```
>> X=zeros(3,101);
>> X(:,1)=[10;10;10];
>> for k=2:101,X(:,k)=L*X(:,k-1);end
>> G=zeros(3,101);
>> for k=1:101, G(:,k)=X(:,k)/sum(X(:,k));end
```

La gráfica de estas poblaciones normalizadas es interesante.

```
>> t=0:100;
>> plot(t,G')
>> xlabel('Tiempo')
>> ylabel('Porcentajes')
>> legend('Primera clase de edad','Segunda clase de edad',...
'Tercera clase de edad')
```

La salida aparece en la figura **??**. Después de un número suficiente de años, el porcentaje de organismos en cada clase se aproxima a 74 %, 24 % y 4 %.

El autovalor dominante $r = 1{,}5$ nos dice que la población tiende a crecer sin límite. El caso $r < 1$ significa extinción.

11.6. Cadenas de Markov homogéneas y finitas

Una matriz $P_{n\times n} = (p_{ij})$ con coeficientes reales y no negativa se dice que es una **matriz estocástica** si sus filas o columnas suman 1. Se dice que es **doblemente estocástica** si sus filas y columnas suman 1.

Nos centraremos en el caso en que las columnas suman 1. No es raro encontrar textos donde esta condición se supone sobre las filas.

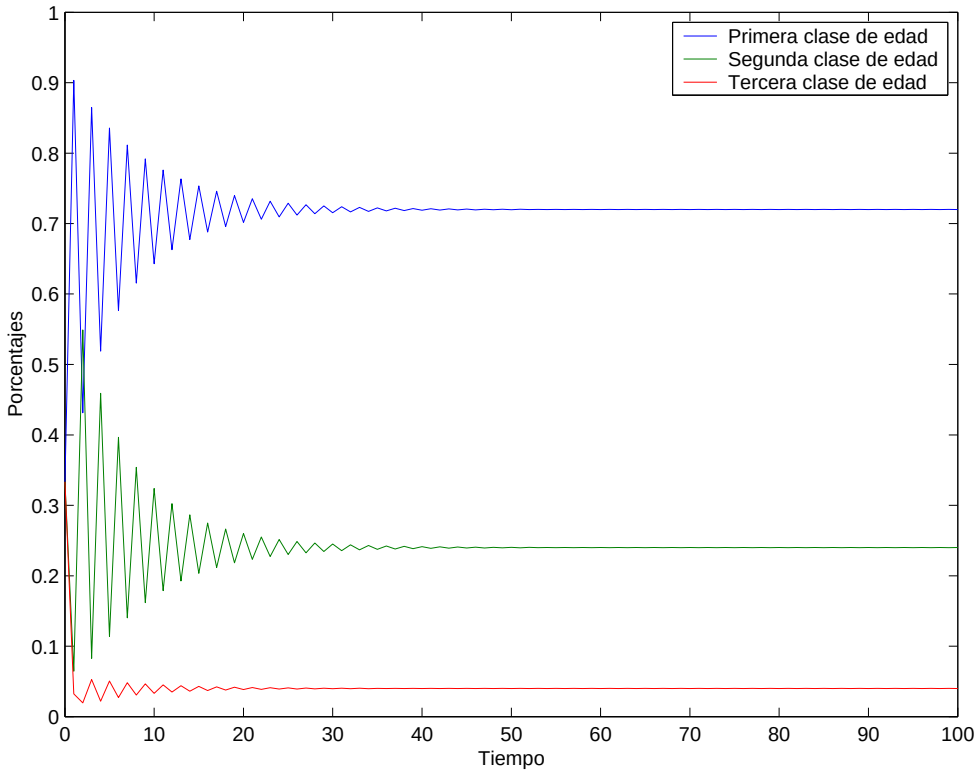

Figura 11.7: Evolución de los porcentajes de cada clase.

Figura 11.8: A.A. Markov (1856-1922)

Consideremos un conjunto de variables aleatorias X_τ, en las que todas tienen el mismo rango $\{S_1,\ldots,S_n\}$, denominado espacio de estados. Una **cadena de Markov** es un proceso estocástico que verifica

$$P(X_{\tau+1}=S_j \mid X_\tau = S_{i_\tau}, X_{\tau-1}=S_{i_{\tau-1}},\ldots,X_0=S_{i_0}) = P(X_{\tau+1}=S_j \mid X_\tau = S_{i_\tau}),$$

para cada $\tau = 0,1,2,\ldots$ En otras palabras, la probabilidad de que $X_{\tau+1}$ se encuentre en el estado S_j solamente depende del estado en que se hallase X_τ, y no en los estados de periodos anteriores. Si la probabilidad tampoco depende del valor de τ, se dice que la cadena de Markov es **homogénea**, y si el número de estados es finito, la llamaremos **finita**.

Cada cadena de Markov define una matriz estocástica, y recíprocamente. El valor $p_{ij}(\tau) = P(X_\tau = S_i \mid X_{\tau-1} = S_j)$ es la probabilidad de encontrarse en el estado S_i en el instante τ supuesto que estaba en el estado S_j en el instante $\tau-1$, y $p_{ij}(\tau)$ es la probabilidad de transición del estado S_j al estado S_i en el instante τ. La matriz $P(\tau) = (p_{ij}(\tau))$ es una matriz no negativa, y cada columna suma 1. Por tanto, $P(\tau)$ es una matriz estocástica. En las cadenas de Markov homogéneas, las probabilidades de transición no dependen de τ, y tenemos la matriz de transición P. De manera clara, toda matriz estocástica $P_{n\times n}$ define una cadena de Markov de n estados, en donde sus entradas representan las probabilidades de transición.

Un **vector de distribución de probabilidad** es un vector no negativo $\boldsymbol{p} = (p_1,\ldots,p_n)^t$ tal que $\sum_k p_k = 1$. Para una cadena de Markov de n estados, el vector de distribución de probabilidad del k-ésimo paso se define como

$$\boldsymbol{p}(k) = \begin{pmatrix} p_1(k) \\ \vdots \\ p_n(k) \end{pmatrix}, k = 1,2,\ldots \text{ donde } p_j(k) = P(X_k = S_j).$$

En otras palabras, $p_j(k)$ es la probabilidad de estar en el estado j-ésimo tras k pasos, pero antes del $(k+1)$-ésimo. El vector de distribución inicial es

$$\boldsymbol{p}(0) = \begin{pmatrix} p_1(0) \\ \vdots \\ p_n(0) \end{pmatrix}, \text{ donde } p_j(0) = P(X_0 = S_j).$$

El vector de distribución del k-ésimo paso se puede describir a través del teo-

rema de probabilidad total. Así,

$$\begin{aligned} p_j(1) &= P(X_1 = S_j) = P(X_1 = S_j \wedge (X_0 = S_1 \vee X_0 = S_2 \vee \cdots \vee X_0 = S_n)) \\ &= P((X_1 = S_j \wedge X_0 = S_1) \vee (X_1 = S_j \wedge X_0 = S_2) \vee \cdots \vee (X_1 = S_j \wedge X_0 = S_n)) \\ &= \sum_{i=1}^{n} P(X_1 = S_j \wedge X_0 = S_i) = \sum_{i=1}^{n} P(X_0 = S_i) P(X_1 = S_j \mid X_0 = S_i) \\ &= \sum_{i=1}^{n} p_i(0) p_{ji} \text{ para } j = 1, 2, \ldots, n. \end{aligned}$$

Por tanto, $\boldsymbol{p}(1) = P\boldsymbol{p}(0)$. Pero la propiedad de no memoria de la cadena de Markov nos dice que también se tiene que $\boldsymbol{p}(2) = P\boldsymbol{p}(1)$, y, en general, que $\boldsymbol{p}(k) = P\boldsymbol{p}(k-1)$. Por tanto,

$$\boldsymbol{p}(k) = P^k \boldsymbol{p}(0),$$

y tenemos que la entrada (i, j) en P^k representa la probabilidad de transición de S_j a S_i en exactamente k pasos. Por esta razón, a P^k se la denomina matriz de transición del k-ésimo paso.

El problema fundamental de las cadenas de Markov es el comportamiento asintótico de la cadena, que está relacionado con la existencia o no de $\lim P^k$, y aquí es donde usaremos la teoría de Perron-Frobenius.

Tenemos que P es una matriz no negativa. Además, su radio espectral es 1. En efecto, si λ es autovalor de P, también lo es de su matriz traspuesta P^t. Sea $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ un autovector de P^t asociado a λ, y $|x_j| = \max\{|x_i| \mid i = 1, \ldots, n\}$. Entonces

$$|\lambda| = \frac{|\sum_{i=1}^{n} p_{ij} x_i|}{|x_j|} \leq \frac{\sum_{i=1}^{n} |p_{ij}||x_i|}{|x_j|} \leq \sum_{i=1}^{n} p_{ij} = 1.$$

En consecuencia, si P es primitiva existe un único vector $\boldsymbol{p} > 0$ asociado al autovalor $\rho = 1$ tal que $\sum_{i=1}^{n} p_i = 1$. Entonces

$$\lim_{k\to\infty} (\rho^{-1} P)^k = \lim_{k\to\infty} P^k = \boldsymbol{v}\boldsymbol{w}^t,$$

por 8.3.2. Además, $\boldsymbol{v}$ es autovector de P asociado a 1, $\boldsymbol{w}$ es autovector de P^t asociado a 1, y $\boldsymbol{w}^t\boldsymbol{v} = 1$. Tomemos $\boldsymbol{v} = \boldsymbol{p} > 0$ el vector de Perron asociado, por lo que $\|\boldsymbol{v}\|_1 = 1$. Sabemos que P^t tiene como autovector asociado a 1 el vector $\mathbf{1} = (1, 1, \ldots, 1)^t$. Entonces existe $\alpha \neq 0$ tal que $\boldsymbol{w} = \alpha\mathbf{1}$, y se tiene que verificar que $\alpha\mathbf{1}^t\boldsymbol{v} = 1$. Por tanto, $\alpha = 1$, y el vector $\boldsymbol{w}$ correspondiente a $\boldsymbol{p}$ es $\mathbf{1}$. En conclusión,

$$\lim_{k\to\infty} P^k = \boldsymbol{p}\mathbf{1}^t.$$

Además,

$$\lim_{k\to\infty} \boldsymbol{p}(k) = \lim P^k \boldsymbol{p}(0) = \boldsymbol{p}\mathbf{1}^t\boldsymbol{p}(0) = \boldsymbol{p},$$

pues $\mathbf{1}^t \boldsymbol{p}(0) = 1$, al ser $\boldsymbol{p}(0)$ un vector de probabilidad. Por tanto, el sistema se aproxima a un punto de equilibrio en que las proporciones de los distintos estados vienen dados por las entradas de $\boldsymbol{p}$. Además, el comportamiento límite no depende de las proporciones iniciales.

La teoría se puede desarrollar para el tratamiento de cadenas con matrices no primitivas o reducibles, pero escapa al objetivo del curso.

Ejemplo 11.6.1. Consideremos la cadena de Markov que se obtiene al colocar un ratón en una caja con tres módulos, y conexiones tal como aparecen en la figura **??**. Supongamos que el ratón se mueve de una habitación a otra eligiendo

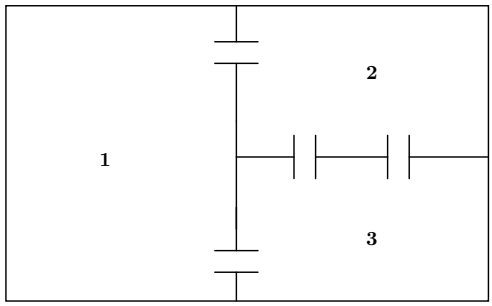

Figura 11.9:

al azar. Por ejemplo, cada minuto se abren las puertas, y se fuerza al ratón que se mueva mediante una corriente en la habitación en la que se encuentre. Si el ratón se coloca inicialmente en la habitación número 2, entonces el vector de probabilidad inicial es

$$\boldsymbol{p}(0) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Pero si el proceso se inicia mediante una moneda al aire para que caiga sobre uno de los módulos, entonces una distribución inicial razonable es

$$\boldsymbol{p}(0) = \begin{pmatrix} 0{,}5 \\ 0{,}25 \\ 0{,}25 \end{pmatrix},$$

porque el área del módulo 1 es el 50 % del área de la caja, y las áreas de los módulos 2 y 3 son, cada una, el 25 %. La matriz de transición para esta cadena de Markov es

$$M = \begin{bmatrix} 0 & 1/3 & 1/3 \\ 1/2 & 0 & 2/3 \\ 1/2 & 2/3 & 1\ 0 \end{bmatrix}.$$

Si la distribución inicial es la aleatoria, el vector de probabilidad tras tres movimientos es

$$M^3\boldsymbol{p}(0) = \begin{bmatrix} 2/9 & \frac{7}{27} & \frac{7}{27} \\ \frac{7}{18} & 2/9 & \frac{14}{27} \\ \frac{7}{18} & \frac{14}{27} & 2/9 \end{bmatrix} \begin{pmatrix} 0{,}5 \\ 0{,}25 \\ 0{,}25 \end{pmatrix} = \begin{bmatrix} \frac{13}{54} \\ \frac{41}{108} \\ \frac{41}{108} \end{bmatrix}.$$

Esto significa que la probabilidad de encontrar al ratón en la habitación número 1 tras tres movimientos es $13/54$, la de encontrarlo en la habitación número 2 es $41/108$, y la de encontrarlo en la habitación número 3 es $41/108$.

La matriz M es irreducible, pues su grafo asociado es fuertemente conexo, y es primitiva, pues si $B_1 = \beta(M)$, entonces

$$B_2 = \beta(B_1 \cdot B_1) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} > 0.$$

Por tanto, existe $\lim M^k$ y también $\lim \boldsymbol{p}(k)$. En primer lugar, el autovalor de Perron de M es igual a 1, y su autovector de Perron se obtiene a partir de

$$\text{null}(M-I) \equiv \begin{bmatrix} -1 & 1/3 & 1/3 \\ 1/2 & -1 & 2/3 \\ 1/2 & 2/3 & -1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \Rightarrow \begin{bmatrix} 1 & 0 & -2/3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0} \Rightarrow \begin{cases} x_1 &= \frac{2}{3}x_3, \\ x_2 &= x_3, \\ x_3 &= x_3. \end{cases}$$

Por tanto, $\text{null}(M-I) = \langle \boldsymbol{v} \rangle$, donde

$$\boldsymbol{v} = \begin{pmatrix} 2/3 \\ 1 \\ 1 \end{pmatrix} \text{ y el vector de Perron es } \boldsymbol{p} = \frac{1}{\|\boldsymbol{v}\|_1}\boldsymbol{v} = \begin{pmatrix} 2/8 \\ 3/8 \\ 3/8 \end{pmatrix}.$$

La distribución límite se puede interpretar como que, a largo plazo, el ratón estará en la cámara 1 el 25 % del tiempo, en la cámara 2 el 35,5 %, y en la cámara 3 el 37,5 %. Y recordemos que es independiente de cómo comenzó el proceso.

Bibliografía

[ABI02] T.N.E. Greville A. Ben-Israel. *Generalized Inverses: Theory and Applications (2nd ed.)*. Addison-Wesley, 2002.

[BR97] R.B. Bapat and T.E.S. Raghavan. *Nonnegative Matrices and Applications*. Encyclopedia of Mathematics and its applications. Cambridge University Press, Cambridge, 1997.

[Gol91] D. Goldberg. What every computer scientist should know about floating point arithmetic. *ACM Computing Surveys*, 23(1):5–48, March 1991.

[GV96] G.H. Golub and C.F. VanLoan. *Matrix Computations*. John Hopkins University Press, Baltimore, 1996.

[Hof01] Joe D. Hoffman. *Numerical Methods for Engineers and Scientists*. Marcel Dekker, 2nd edition, 2001.

[LT85] P. Lancaster and M. Tismenetsky. *The theory of matrices : with applications*. Academic Press, 1985.

[Mey98] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphi, PA, 1998.

[Min88] H. Minc. *Nonnegative Matrices*. Wiley, New York, 1988.

[RB09] G.B. Costa R. Bronson. *Matrix Methods: Applied Linear Algebra, 3rd ed.* Academic Press, MA USA, 2009.

[TB97] L.N. Trefethen and D. Bau. *Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.

[Wat02] D. Watkins. *Fundamentals of Matrix Computations, 2nd edition*. Wiley, New York, 2002.