ÁLGEBRA LINEAL

Apuntes elaborados por

Juan González-Meneses López.

Curso 2008/2009

Departamento de Álgebra.

Universidad de Sevilla.

Índice general

Tema 1. Matrices. Determinantes. Sistemas de ecuaciones lineales

1.1. Matrices: definición, operaciones y propiedades básicas.

En este tema estudiaremos las matrices como objeto matemático y su aplicación al estudio de los sistemas de ecuaciones lineales. Veremos sus propiedades fundamentales, las operaciones básicas, y una aplicación importante de estos conceptos: el Teorema de Rouché-Frobenius.

A partir de ahora fijaremos un **cuerpo de escalares**, que llamaremos K. La definición de cuerpo se dará en el Tema 2. Por ahora es suficiente pensar que K es el conjunto de los números racionales, reales o complejos, y que un *escalar* es uno de estos números.

Una **matriz** $m \times n$ es una tabla de m filas y n columnas de escalares. Es decir, un objeto de la forma

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

donde cada a_{ij} es un escalar.

Una vez vista la definición de matriz, fijaremos algunas notaciones:

Denotaremos $\mathcal{M}_{m\times n}(K)$ al conjunto de matrices $m\times n$, cuyo cuerpo de escalares es K. Si no nos interesa especificar el cuerpo de escalares, escribiremos simplemente $\mathcal{M}_{m\times n}$.

Normalmente usaremos una letra mayúscula para denotar una matriz, y la misma letra en minúscula, con los subíndices correspondientes, para denotar sus *elementos* o *entradas*. Por ejemplo, escribiremos una matriz $A \in \mathcal{M}_{m\times n}$ como sigue:

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

Si queremos especificar la letra que usaremos para los elementos de una matriz, escribiremos $A = (a_{ij})$.

Comencemos a estudiar las propiedades de las matrices.

Diremos que dos matrices A y B son **iguales** si ambas tienen las mismas dimensiones (es decir, $A, B \in \mathcal{M}_{m\times n}$), y además $a_{ij} = b_{ij}$ para todo $i, j \quad 1 \leq i \leq m$, $1 \leq j \leq n$.

Dadas dos matrices $A, B \in \mathcal{M}_{m\times n}$, definimos su **suma**, $A + B$, como la matriz $C \in \mathcal{M}_{m\times n}$ tal que

$$c_{ij} = a_{ij} + b_{ij}.$$

Dada una matriz $A \in \mathcal{M}_{m\times n}(K)$ y un escalar $\alpha \in K$, definimos su **producto**, αA, como la matriz $D \in \mathcal{M}_{m\times n}(K)$ tal que

$$d_{ij} = \alpha\, a_{ij}$$

Es decir, dos matrices de las mismas dimensiones se pueden sumar, término a término, dando lugar a otra matriz de la misma dimensión. Y también podemos multiplicar una matriz por un escalar, dando lugar a otra matriz de las mismas dimensiones donde cada término se ha multiplicado por el escalar.

Un ejemplo importante de matrices son los **vectores**:

Un **vector** es una matriz $m\times 1$. Las entradas de un vector se llaman **coordenadas**.

Aunque sean un caso particular de matrices, trataremos a los vectores de forma especial. Los denotaremos en negrita, y como sólo tienen una columna, no escribiremos el segundo índice de cada término. Por ejemplo, escribiremos:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}.$$

También nos referiremos como **vectores fila** a las matrices $1 \times n$. Así, un vector fila podría ser:

$$\mathbf{v} = (v_1, v_2, \dots, v_n).$$

En los vectores fila, las coordenadas se suelen escribir separadas por comas. Pero recordemos que, si no se especifica lo contrario, un vector consta de una *columna*.

Los vectores suelen resultar familiares, ya que se usan para representar los puntos de los espacios geométricos. Por ejemplo, los puntos del plano $\mathbb{R}^2$ se corresponden con los vectores de dos coordenadas: $\mathcal{M}_{2\times 1}$. Los puntos del espacio $\mathbb{R}^3$ se corresponden con los vectores de tres coordenadas: $\mathcal{M}_{3\times 1}$. Y así se puede continuar con los espacios de dimensiones superiores.

Ahora estudiaremos la operación más importante con matrices: la multiplicación. Comenzaremos con un caso particular:

Dadas dos matrices

$$A = (a_1\ a_2\ \cdots\ a_n) \in \mathcal{M}_{1\times n}, \qquad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathcal{M}_{n\times 1},$$

se define su **producto**, AB, como la matriz $C \in \mathcal{M}_{1\times 1}$ cuya única entrada es:

$$a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Nota: Si se consideran las dos matrices A y B como vectores (un vector fila y un vector columna), el producto que acabamos de definir se llama **producto escalar** de A y B. Lo estudiaremos más a fondo en temas posteriores.

Para extender esta definición a matrices con más de una fila o columna, llamaremos **fila i** de una matriz $A = (a_{ij}) \in \mathcal{M}_{m\times n}$, al vector fila $(a_{i1}\ a_{i2}\ \cdots\ a_{in}) \in \mathcal{M}_{1\times n}$, y llamaremos **columna j** al vector columna

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathcal{M}_{m\times 1}.$$

Tenemos entonces:

Dadas dos matrices $A \in \mathcal{M}_{m\times n}$ y $B \in \mathcal{M}_{n\times p}$, se define su **producto**, AB, como la matriz $C \in \mathcal{M}_{m\times p}$, donde el elemento c_{ij} es el producto de la fila i de A por la columna j de B. Es decir,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Nota: Es importante darse cuenta que **no se pueden multiplicar dos matrices de cualquier dimensión**. Sólo se pueden multiplicar A y B si el tamaño de las filas de A es igual al tamaño de las columnas de B. El resultado de la multiplicación será una matriz C con el mismo número de filas que A y el mismo número de columnas que B. Esquemáticamente:

$$\underset{m\times n}{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \hline a_{i1} & a_{i2} & \cdots & a_{in} \\ \hline \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}} \quad \underset{n\times p}{\left(\begin{array}{cc|c|cc} b_{11} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{np} \end{array}\right)} = \underset{m\times p}{\begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1p} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & \boxed{c_{ij}} & \cdots & c_{ip} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mj} & \cdots & c_{mp} \end{pmatrix}}$$

Nota: esta definición del producto de matrices puede resultar extraña. ¿Por qué no multiplicar matrices simplemente multiplicando sus entradas correspondientes? La respuesta proviene de los sistemas lineales. Arthur Cayley (1821-1895) estudiaba los sistemas de dos ecuaciones con dos incógnitas

$$\left\{ \begin{array}{l} ax + by = x' \\ cx + dy = y' \end{array} \right.$$

como transformaciones del plano, que a cada punto (x, y) le hacen corresponder el punto (x', y'). Por tanto, podemos decir que la matriz $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ transforma el plano, moviendo cada punto (x, y) a la posición (x', y'). Si consideramos ahora otra matriz $\begin{pmatrix} e & f \\ g & h \end{pmatrix}$, también transformará el plano, moviendo el punto (x', y') a la posición (x'', y''), mediante las ecuaciones:

$$\left\{ \begin{array}{l} ex' + fy' = x'' \\ gx' + hy' = y'' \end{array} \right.$$

Por tanto, si hacemos actuar estas dos transformaciones, una detrás de otra, el punto (x, y) irá a la posición (x'', y''), donde estas coordenadas verifican:

$$x'' = ex' + fy' = e(ax + by) + f(cx + dy) = (ae + cf)x + (be + df)y,$$

y por otro lado:

$$y'' = gx' + hy' = g(ax + by) + h(cx + dy) = (ag + ch)x + (bg + dh)y.$$

Por tanto, la composición de las dos transformaciones tiene por ecuación:

$$\begin{cases} (ae + cf)x + (be + df)y = x'' \\ (ag + ch)x + (bg + ch)y = y'' \end{cases}$$

Si observamos la matriz de esta transformación, vemos que es el **producto** de las matrices anteriores, ya que:

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ae + cf & be + df \\ ag + ch & bg + ch \end{pmatrix}.$$

Luego **el producto de matrices corresponde a la composición de transformaciones**. Estas definiciones de Cayley se generalizaron a cualquier dimensión. Más adelante estudiaremos las **transformaciones lineales** en general, y veremos cómo el producto de matrices corresponde a la composición de transformaciones lineales.

Hemos definido tres operaciones con matrices: la suma y el producto de matrices, y el producto de una matriz por un escalar. Veamos cuáles son las principales propiedades de estas operaciones.

Propiedades de la suma de matrices: En $\mathcal{M}_{m\times n}$ se tienen las siguientes propiedades:

1. **Propiedad conmutativa**: $A + B = B + A$.
2. **Propiedad asociativa**: $(A + B) + C = A + (B + C)$.
3. **Elemento neutro**: Existe una única matriz $\mathcal{O} \in \mathcal{M}_{m\times n}$, llamada *matriz nula*, tal que $A + \mathcal{O} = \mathcal{O} + A$, para toda matriz $A \in \mathcal{M}_{m\times n}$.
4. **Elemento opuesto**: Dada una matriz $A \in \mathcal{M}_{m\times n}$, existe otra matriz $B \in \mathcal{M}_{m\times n}$, llamada *opuesta de* A, tal que $A + B = \mathcal{O}$.

La matriz nula está formada por ceros. Por otro lado, si B es la matriz opuesta de A, se tiene $b_{ij} = -a_{ij}$.

Nota: Como $\mathcal{M}_{m\times n}$ verifica estas cuatro propiedades, se dice que $\mathcal{M}_{m\times n}$ es un **grupo abeliano** con respecto a la suma. Estudiaremos el concepto de grupo más adelante.

Propiedades del producto de matrices: Si A, B y C son matrices, de las dimensiones adecuadas para que se puedan multiplicar o sumar (en cada caso), se tiene

1. **Propiedad asociativa**: $(AB)C = A(BC)$.
2. **Propiedades distributivas**:
   a) $(A+B)C = AC + BC$.
   b) $A(B+C) = AB + AC$.
3. **Elemento neutro (a izquierda y derecha):** Existe una única matriz $I \in \mathcal{M}_{n\times n}$ tal que:
   a) $AI = A$ para toda $A \in \mathcal{M}_{m\times n}$.
   b) $IB = B$ para toda $B \in \mathcal{M}_{n\times p}$.

Nota: El producto de matrices **no es conmutativo** en general. Es decir, normalmente $AB \neq BA$, incluso cuando los dos productos estén bien definidos. Además, **no siempre existe el elemento inverso**: dada una matriz cuadrada A, no tiene por qué existir otra matriz B tal que $AB = I$.

Por otra parte, la matriz neutra $I = (\delta_{ij})$ se llama **matriz identidad**, y es una matriz cuadrada definida por: $\delta_{ij} = 0$ si $i \neq j$, y $\delta_{ii} = 1$ para todo i. Por ejemplo, la matriz identidad de dimensión 3 es:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Propiedades del producto de matrices y escalares: Si A y B son matrices, de las dimensiones adecuadas para que se puedan sumar o multiplicar (en cada caso), y si α y β son escalares, se tiene

1. $\alpha(\beta A) = (\alpha\beta)A$.
2. $\alpha(AB) = (\alpha A)B = A(\alpha B)$.
3. $(\alpha + \beta)A = \alpha A + \beta A$.
4. $\alpha(A + B) = \alpha A + \alpha B$.

Terminaremos esta sección estudiando una última operación de matrices, llamada **trasposición**.

Dada una matriz $A \in \mathcal{M}_{m\times n}$, llamamos **traspuesta** de A a la matriz $A^t \in \mathcal{M}_{n\times m}$, definida de forma que las filas de A sean las columnas de A^t, y viceversa. Es decir, si $A^t = (b_{ij})$, se tiene $b_{ij} = a_{ji}$ para todo i, j.

Ejemplo 1.1 *Si* $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, *entonces* $A^t = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

Utilizaremos la traspuesta de una matriz en temas posteriores. Por ahora nos limitaremos a ver algunas propiedades:

Propiedades de la trasposición: Sean A y B matrices de las dimensiones adecuadas. Se tiene:

1. $(A+B)^t = A^t + B^t$.
2. $(AB)^t = B^t A^t$.
3. $(A^t)^t = A$.

Por último, hay un tipo especial de matriz que será importante más adelante:

Una matriz A es **simétrica** si $A^t = A$.

Observemos que, si A es simétrica, entonces debe ser una matriz cuadrada. Las matrices cuadradas tienen propiedades especiales, que estudiaremos en este tema. Pero ahora continuaremos con propiedades importantes de las filas y columnas de una matriz.

1.2. Transformaciones elementales de filas: matrices escalonadas y reducidas.

A la hora de aplicar las matrices al estudio de los sistemas de ecuaciones lineales, y para estudiar las propiedades de los determinantes, una herramienta esencial consiste en las llamadas *transformaciones elementales* de matrices, que se definen como sigue.

Las **transformaciones elementales de filas** que se pueden aplicar a una matriz, son las siguientes:

1. Intercambiar dos filas.

2. Multiplicar una fila por un escalar no nulo.

3. Añadir a una fila un múltiplo no nulo de otra.

A partir de esta definición, se obtiene el siguiente concepto:

Diremos que dos matrices son **equivalentes por filas** si podemos obtener una, a partir de la otra, mediante transformaciones elementales de filas.

Gracias a las transformaciones elementales de filas, podremos siempre transformar cualquier matriz en otra, equivalente por filas, que es más sencilla desde un punto de vista que veremos más adelante. Estas matrices *sencillas* vienen definidas a continuación.

Diremos que una matriz es **escalonada por filas** si cumple lo siguiente:

1. Todas las filas de ceros (si las hay) están en la parte inferior de la matriz.

2. En las filas que no sean de ceros, el primer término no nulo de una fila está *más a la izquierda* del primer término no nulo de la fila siguiente.

Ejemplo 1.2 *La siguiente matriz es escalonada por filas:*

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

Un método para transformar cualquier matriz en una escalonada por filas es el siguiente:

El **método de eliminación de Gauss** aplicado a una matriz, la transforma en una matriz equivalente que es *escalonada por filas*. Consiste en los siguientes pasos:

Paso 1: Si es necesario, intercambiar la primera fila con otra, para que la primera columna que no sea de ceros tenga un elemento no nulo en la primera posición.

Paso 2: Sumar a cada fila un múltiplo adecuado de la primera, de manera que la primera columna que no sea de ceros tenga sólo un elemento no nulo: el de la primera fila.

Paso 3: Ignorando temporalmente la primera fila, repetir todo el proceso con las restantes filas.

Como este proceso da lugar, claramente, a una matriz escalonada por filas, hemos demostrado el siguiente resultado:

Proposición 1.3 *Toda matriz $m \times n$ es equivalente por filas a otra matriz $m \times n$ escalonada por filas.*

Demostración: Sólo hay que aplicar a la matriz inicial el método de eliminación de Gauss. □

A continuación veremos cómo, usando transformaciones elementales, podemos obtener matrices aún más sencillas que las escalonada por filas: las matrices *reducidas por filas.*

Diremos que una matriz es **reducida por filas** si cumple lo siguiente:

1. Es escalonada por filas.
2. El primer elemento no nulo de cada fila, llamado **pivote**, es 1.
3. Encima (y debajo) de cada pivote sólo hay ceros.

Ejemplo 1.4 *La siguiente matriz es reducida por filas:*

$$\begin{pmatrix} 1 & 0 & 4 & 3 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Se tiene entonces:

Método de eliminación de Gauss-Jordan para transformar una matriz en otra equivalente por filas, que sea reducida por filas:

Paso 1: Aplicar a la matriz el método de Gauss.

Paso 2: Multiplicar cada fila no nula por un escalar conveniente, de manera que todos los pivotes sean 1.

Paso 3: Comenzando por el pivote más a la derecha, eliminar todos los elementos no nulos que tenga encima, sumándole a cada fila un múltiplo conveniente de la fila de este pivote. Realizar la misma operación con todos los pivotes, de derecha a izquierda.

Después de aplicar este método a una matriz, se obtiene claramente otra matriz equivalente (puesto que se han aplicado transformaciones elementales de filas) que es reducida por filas (por construcción). Hemos probado por tanto el siguiente resultado:

Teorema 1.5 *Toda matriz $m \times n$ es equivalente por filas a otra matriz $m \times n$ reducida por filas.*

Demostración: Basta con aplicar a la matriz inicial el método de eliminación de Gauss-Jordan. □

Una propiedad importante de la forma reducida por filas equivalente a una matriz dada es que es *única*. Pero aún no tenemos las herramientas suficientes para demostrar esto.

1.3. Dependencia lineal y rango.

El concepto de *dependencia lineal* de vectores es fundamental para el estudio de matrices, sistemas lineales y, como veremos en temas posteriores, espacios vectoriales.

Geométricamente, un vector de n coordenadas se representa, en el espacio de dimensión n, como una flecha que parte del origen y termina en el punto que tiene esas coordenadas. Las operaciones básicas de matrices, aplicadas a vectores, se ven geométricamente como sigue:

- Multiplicar un vector por un escalar (digamos, un número real), equivale a multiplicar la longitud del vector por ese escalar.
- Sumar dos vectores $\mathbf{v}_1$ y $\mathbf{v}_2$ corresponde al siguiente procedimiento: Si se traslada el vector $\mathbf{v}_2$, sin cambiar su dirección ni su tamaño, hasta hacer que su comienzo coincida con el final del vector $\mathbf{v}_1$, entonces vector $\mathbf{v}_1 + \mathbf{v}_2$ es el que une el origen de coordenadas con el final de este nuevo vector $\mathbf{v}_2$.

Dados r vectores $\mathbf{v}_1, \ldots, \mathbf{v}_r$ de la misma dimensión, llamamos **combinación lineal** de estos vectores a cualquier expresión de la forma:

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r,$$

donde $\alpha_1, \ldots, \alpha_r$ son escalares cualesquiera.

Es decir, una combinación lineal de r vectores es **otro vector**, que resulta de cambiar el tamaño de cada uno de los vectores iniciales, y sumar los resultados (haciendo comenzar cada vector en el final del vector precedente).

Ejemplo 1.6 *Una combinación lineal de un sólo vector, $\mathbf{v}$, tiene la forma $\alpha\mathbf{v}$, donde α es un escalar. Por tanto es otro vector con la misma dirección que $\mathbf{v}$, y cuyo tamaño es α veces el tamaño de $\mathbf{v}$. Por tanto, $\alpha\mathbf{v}$ está* **en la recta** *determinada por $\mathbf{v}$.*

Ejemplo 1.7 *Una combinación de dos vectores de $\mathbb{R}^3$ es otro vector que está* **en el plano** *determinado por estos dos vectores.*

Diremos que un vector $\mathbf{v}$ **depende linealmente** de un conjunto de vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ si $\mathbf{v}$ se puede escribir como combinación lineal de $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

Ejemplo 1.8 *El vector* $(3, -2, 2)$ *depende linealmente de los vectores* $(1, 0, 2)$ *y* $(-1, 2, 2)$*, ya que se tiene la combinación lineal:*

$$\begin{pmatrix} 3 \\ -2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + (-1) \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}.$$

Ejemplo 1.9 *El vector* $\mathbf{0}$*, con todas sus coordenadas nulas, depende linealmente de cualquier conjunto de vectores. Basta tomar todos los coeficientes* 0 *en la combinación lineal.*

Ejemplo 1.10 *Cualquier vector depende linealmente de un conjunto de vectores que lo contenga. Basta tomar su coeficiente* 1*, y todos los demás* 0*.*

Hay otra forma de ver la dependencia lineal:

Diremos que un sistema (o conjunto) de vectores de la misma dimensión $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es **linealmente dependiente**, si existen r escalares $\alpha_1, \ldots, \alpha_r$, *no todos nulos*, tales que

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r = \mathbf{0}.$$

En caso contrario, es decir, si la única forma de escribir el vector $\mathbf{0}$ como combinación lineal de estos vectores es tomando $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$, diremos que el sistema S es **linealmente independiente** o **libre**.

La relación entre esta definición de dependencia lineal y la anterior viene dada por el siguiente resultado.

Lema 1.11 *Un sistema de vectores* $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ *es linealmente dependiente si y sólo si uno de ellos es combinación lineal de los demás.*

Demostración: Directa. $\square$

Si en un sistema de vectores, uno de ellos es combinación lineal de los demás, ese vector "sobra", desde el punto de vista geométrico. Es decir, si lo quitamos del sistema, el conjunto de vectores que se puede definir como combinación lineal de los vectores del sistema sigue siendo el mismo. Podríamos, por tanto, ir eliminando vectores del sistema, hasta que no pudiéramos eliminar más; es decir, hasta que el sistema fuera linealmente independiente. En efecto, se tiene:

Teorema 1.12 *Dado un sistema de r vectores $S = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}$, no todos nulos, se verifica:*

1. *Existe al menos un sistema $S_0 \subset S$ linealmente independiente; y todos los demás vectores de S dependen linealmente de los de S_0.*
2. *Todos los sistemas S_0 que satisfacen la condición anterior tienen el mismo número de elementos. A este número lo llamamos* **rango** *de S.*

Demostración: La demostración de 1 ya está esbozada arriba. Para demostrar 2, se supone que se tienen dos subsistemas libres, S_1 y S_2, con distinto número de vectores. Si S_2 tiene más vectores que S_1, se demuestra que $\mathbf{0}$ puede escribirse como una combinación lineal no trivial de los elementos de S_2, escribiendo éstos como combinación lineal de los de S_1, y usando que un sistema homogéneo con menos ecuaciones que incógnitas tiene soluciones no triviales, como veremos en el teorema de Rouché-Frobenius. □

El rango de un sistema de vectores se puede también definir como sigue:

El **rango** de un sistema de vectores S es el tamaño del mayor sistema libre que se puede formar con los vectores de S.

Ahora relacionaremos, de forma muy sencilla, los sistemas de vectores con las matrices. Simplemente, a un sistema de m vectores de dimensión n, le asociamos una matriz $m \times n$, donde cada fila es un vector del sistema. Así, podemos definir:

El **rango** de una matriz es el rango del sistema de vectores formado por sus filas.

Al rango de una matriz A lo denotaremos $\mathrm{rg}(A)$.

Si ahora modificamos la matriz, usando transformaciones elementales de filas, estaremos modificando el sistema de vectores asociado. Podemos, por tanto, intercambiar la posición de los vectores, multiplicar un vector por un escalar no nulo, o sumar a un vector un múltiplo no nulo de otro. Pero en cualquier caso, es tiene:

Lema 1.13 *Las transformaciones elementales de filas no alteran del rango de una matriz.*

Demostración: Directa, usando la definición de rango de un sistema de vectores. $\square$

Gracias a este resultado, podremos calcular fácilmente el rango de una matriz:

Teorema 1.14 *Consideremos una matriz $A \in \mathcal{M}_{m\times n}$, y sea A' una matriz reducida equivalente por filas a A. Entonces, el rango de A es igual al número de filas no nulas de A'.*

Demostración: Sólo hay que ver que las filas no nulas de A' forman un sistema libre. Se forma una combinación lineal igualada a cero, y se ve que las coordenadas de los pivotes sólo se pueden anular si el coeficiente de esa fila es nulo. $\square$

Nota: Acabamos de probar que el número de filas no nulas de la forma reducida por filas de una matriz, está determinado por la matriz. Además, cualquier forma escalonada de la misma matriz debe también tener el mismo número de filas no nulas.

1.4. Matrices elementales.

Una vez estudiadas las transformaciones elementales de filas de una matriz, y cómo se pueden utilizar para calcular el rango, veamos la relación entre estas transformaciones y la multiplicación de matrices.

Comenzamos definiendo tres tipos de matrices, que llamaremos **matrices elementales**, y que son el resultado de aplicar a la matriz identidad los tres tipos de transformaciones elementales. Definiremos matrices cuadradas $n\times n$, luego $I \in \mathcal{M}_{n\times n}$ será la matriz identidad de dimensión n.

En primer lugar, dados $i, j,\ 1 \leq i, j \leq n$, definimos T_{ij} como la matriz que se obtiene de

I al intercambiar sus filas i y j.

$$T_{ij} = \left(\begin{array}{ccccccccccc}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & 0 & \cdots & \cdots & \cdots & 1 & & & \\
 & & & \vdots & 1 & & & \vdots & & & \\
 & & & \vdots & & \ddots & & \vdots & & & \\
 & & & \vdots & & & 1 & \vdots & & & \\
 & & & 1 & \cdots & \cdots & \cdots & 0 & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{array}\right)
\begin{array}{l}
\\ \\ \\ \text{fila } i \\ \\ \\ \\ \text{fila } j \\ \\ \\ \\
\end{array}$$

A continuación, dado $i,\ 1 \leq i \leq n$, y un escalar $\alpha \in K$, definimos $M_i(\alpha)$ como la matriz que se obtiene de I al multiplicar su fila i por α.

$$M_i(\alpha) = \left(\begin{array}{ccccccc}
1 & & & & & & \\
 & \ddots & & & & & \\
 & & 1 & & & & \\
 & & & \alpha & & & \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{array}\right)
\begin{array}{l}
\\ \\ \\ \text{fila } i \\ \\ \\ \\
\end{array}$$

Finalmente, dados $i, j \quad (1 \leq i, j \leq n, \quad i \neq j)$, y un escalar $\alpha \in K$, definimos $P_{ij}(\alpha)$ como la matriz que se obtiene de I al sumarle a la fila i la fila j multiplicada por α.

$$P_{ij}(\alpha) = \left(\begin{array}{ccccccccccc}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & 1 & \cdots & \cdots & \cdots & \alpha & & & \\
 & & & \vdots & 1 & & & \vdots & & & \\
 & & & \vdots & & \ddots & & \vdots & & & \\
 & & & \vdots & & & 1 & \vdots & & & \\
 & & & 0 & \cdots & \cdots & \cdots & 1 & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{array}\right)
\begin{array}{l}
\\ \\ \\ \text{fila } i \\ \\ \\ \\ \text{fila } j \\ \\ \\ \\
\end{array}$$

Podemos describir estos tres tipos de matrices de otra manera:

- T_{ij} coincide con I, salvo en los términos: $t_{ii} = t_{jj} = 0,\ t_{ij} = t_{ji} = 1$.
- $M_i(\alpha)$ coincide con I salvo el el término: $m_{ii} = \alpha$.
- $P_{ij}(\alpha)$ coincide con I salvo en el término: $p_{ij} = \alpha$.

La relación entre las transformaciones elementales de filas y el producto de matrices viene dada por el siguiente resultado:

Lema 1.15 *Sea $A \in \mathcal{M}_{n\times p}$. Se tiene:*

1. *$T_{ij}A$ es la matriz que resulta al intercambiar las filas i y j de A.*
2. *$M_i(\alpha)A$ es la matriz que resulta al multiplicar por α la fila i de A.*
3. *$P_{ij}(\alpha)A$ es la matriz que resulta al sumar a la fila i de A, la fila j multiplicada por α.*

Es decir, aplicar una transformación elemental de filas a una matriz equivale a multiplicarla, a la izquierda, por la matriz elemental correspondiente.

Si seguimos aplicando transformaciones elementales, estaremos multiplicando más matrices elementales a la izquierda. Así podremos llegar hasta una forma reducida, equivalente por filas a la matriz A. Por tanto, se tiene:

Proposición 1.16 *Sea $A \in \mathcal{M}_{m\times n}$ y sea A' una forma reducida por filas de A. Entonces existe una matriz $P \in \mathcal{M}_{m\times m}$, producto de matrices elementales, tal que $A' = PA$.*

Este resultado tiene varias aplicaciones. En primer lugar, podemos ya probar que la forma reducida por filas de una matriz es única.

Lema 1.17 *Si $A, B \in \mathcal{M}_{m\times n}$ son dos matrices reducidas por filas, que son equivalentes por filas, entonces $A = B$.*

Demostración: Ya sabemos que las transformaciones elementales por filas no varían el rango de una matriz, y que si una matriz es reducida por filas, entonces su rango es el número de filas distintas de cero que tiene. Por tanto, el número de filas distintas de cero de A y B es el mismo. Se demuestra entonces el resultado por inducción en n, el número de columnas. Si $n = 1$, entonces o bien $A = B = 0$, o bien $a_{11} = b_{11} = 1$ y todas las demás entradas son cero. En cualquier caso, $A = B$.

Supongamos el resultado cierto para menos de n columnas, con $n > 1$. Sean A' y B' las matrices formadas por las $n-1$ primeras columnas de A y B respectivamente. Ambas son reducidas por filas, pero además son equivalentes por filas, usando las mismas transformaciones que convierten A en B. Por tanto, por hipótesis de inducción, $A' = B'$.

Sólo queda demostrar que la última columna de A y de B son iguales. Sea $r = \mathrm{rg}(A')$. Hay dos posiblidades: si la última columna de A contiene un pivote, entonces $a_{r+1,n} = 1$ y todas las demás entradas de la última columna son ceros. Pero en este caso $\mathrm{rg}(A) = \mathrm{rg}(B) = r+1$, luego la última columna de B también tiene un pivote en la misma posición, y por tanto $A = B$.

Si, por contra, $\mathrm{rg}(A) = \mathrm{rg}(B) = r$, entonces la última columna de A y de B podrá tener sus r primeras entradas no nulas, y el resto deberán ser nulas. Llamemos A_n y B_n a la última columna de A y B, respectivamente. Como A y B son equivalentes por filas, se tiene $B = PA$, donde P es producto de matrices elementales. Más aún, como $A' = B'$, las columnas de los r pivotes de A y B coinciden. Pero al multiplicar P por la columna del primer pivote de A, obtenemos la columna del primer pivote de B. Es decir:

$$\begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{m1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Lo mismo ocurre con la segunda columna de P (usando el segundo pivote), y así sucesivamente, hasta usar los r pivotes. Por tanto, las r primeras columnas de P son iguales a las de la matriz identidad. Pero entonces, como $PA_n = B_n$, donde A_n y B_n sólo tienen r entradas no nulas, un cálculo directo muestra que $A_n = B_n$, y por tanto $A = B$. $\square$

Teorema 1.18 *La forma reducida por filas de una matriz es única.*

Demostración: Su hubiera dos formas reducidas, A' y A'', de una matriz A, ambas serían equivalentes por filas a A, luego serían equivalentes por filas entre ellas. Por tanto, según el resultado anterior, $A' = A''$. $\square$

1.5. Matrices invertibles.

Existe un tipo importante de matrices *cuadradas*: aquellas que admiten una matriz inversa. La definición es la siguiente.

Sea $A \in \mathcal{M}_{n\times n}$. Se dice que A es **invertible** si existe otra matriz $A^{-1} \in \mathcal{M}_{n\times n}$ tal que $AA^{-1} = A^{-1}A = I$. En este caso, A^{-1} se llama la **inversa** de A.

Algunas propiedades de las matrices invertibles son las siguientes:

Teorema 1.19 *Sean $A, B \in \mathcal{M}_{n\times n}$. Se verifica:*

1. *La inversa de A, si existe, es única.*
2. *Si A y B son invertibles, entonces $(AB)^{-1} = B^{-1}A^{-1}$.*
3. *Si A es invertible, entonces A^t también es invertible, y se tiene: $(A^t)^{-1} = (A^{-1})^t$.*
4. *Si A tiene una fila o una columna de ceros, entonces no es invertible.*

Demostración:

1. Si A' y A'' son dos inversas de A, se tiene $A' = A'I = A'(AA'') = (A'A)A'' = IA'' = A''$.
2. Si multiplicamos AB, ya sea a la izquierda o a la derecha, por $B^{-1}A^{-1}$, se obtiene I, luego esta matriz es la inversa de AB.
3. Se tiene $(A^{-1})^tA^t = (A\,A^{-1})^t = I^t = I$. La multiplicación por la derecha es análoga.
4. Si la fila i de A es de ceros, al multiplicarla a la derecha por cualquier matriz, ésta tendrá la fila i de ceros. Lo mismo ocurre con las columnas, multiplicando a la izquierda.

□

Corolario 1.20 *Se tiene:*

1. *Si $A_1, A_2, \cdots, A_r \in \mathcal{M}_{n\times n}$ son invertibles, entonces su producto es invertible, y la inversa es: $(A_1A_2\cdots A_r)^{-1} = A_r^{-1}\cdots A_2^{-1}A_1^{-1}$.*

2. *Si una matriz P es producto de matrices elementales, entonces P es invertible.*

Demostración: La primera propiedad se demuestra igual que la propiedad 2 del teorema anterior. La segunda, demostrando que las matrices elementales son invertibles, y aplicando la propiedad 1. De hecho, se tiene:

$$(T_{i,j})^{-1} = T_{i,j}, \qquad (M_i(\alpha))^{-1} = M_i(\alpha^{-1}), \qquad (P_{i,j}(\alpha))^{-1} = P_{i,j}(-\alpha).$$

$\square$

Veamos ahora cómo es la forma reducida por filas de una matriz invertible:

Teorema 1.21 *Si $A \in \mathcal{M}_{n\times n}$ es una matriz invertible, su forma reducida por filas es la matriz identidad I.*

Demostración: Si usamos el método de Gauss-Jordan para hallar A', la forma reducida por filas de A, tenemos que $A' = PA$, donde P es producto de matrices elementales. Por el resultado anterior, P es invertible, pero A también lo es, por tanto A' es invertible. Ahora bien, A' no puede tener una fila de ceros, ya que en ese caso no sería invertible. Por tanto, en A' hay n pivotes, y la única matriz $n \times n$ reducida por filas que puede tener n pivotes es I. Es decir, $A' = I$. $\square$

Corolario 1.22 *Una matriz $A \in \mathcal{M}_{n\times n}$ es invertible si y sólo si $rg(A) = n$.*

Demostración: Si A es invertible, el teorema anterior nos dice que su forma reducida por filas es I, que tiene n filas no nulas, luego $\mathrm{rg}(A) = n$.

Si $\mathrm{rg}(A) < n$, entonces A', la forma reducida por filas de A, tiene una fila de ceros, luego no es invertible. Pero sabemos que $A' = PA$, por lo que, si A fuera invertible, A' también lo sería. $\square$

Estos resultados nos dan un método sencillo para calcular la inversa de una matriz invertible: Dada $A \in \mathcal{M}_{n\times n}$ invertible, le aplicamos el método de Gauss-Jordan, para calcular su reducida por filas (es decir, I), recordando a cada paso la matriz elemental utilizada. El producto de todas estas matrices, en orden inverso, forma la matriz P, tal que $PA = I$. Es decir, $A^{-1} = P$. Para calcular P (es decir, A^{-1}), podemos multiplicar todas las matrices elementales utilizadas, o mejor aún, ir aplicando a la matriz identidad las mismas operaciones elementales que le apliquemos a A. Por tanto tenemos:

Método para calcular la inversa de una matriz, usando matrices elementales:

A^{-1} es la matriz resultante de aplicar a I las mismas operaciones elementales que se le apliquen a A, para hallar su forma reducida por filas (usando el método de Gauss-Jordan).

Una forma sencilla de aplicar este método es el siguiente. Dada la matriz $A \in \mathcal{M}_{n\times n}$, se considera la matriz $(A|I) \in \mathcal{M}_{n\times 2n}$ que consiste en yuxtaponer la matriz A y la matriz identidad $I \in \mathcal{M}_{n\times n}$. A continuación, se le aplican a esta matriz las transformaciones elementales que transforman A en I, y obtendremos, en las últimas n columnas, la matriz A^{-1}. Es decir, habremos transformado $(A|I)$ en $(I|A^{-1})$.

A continuación mostraremos dos caracterizaciones más de las matrices invertibles, con ayuda de las transformaciones elementales:

Teorema 1.23 *Una matriz $A \in \mathcal{M}_{n\times n}$ es invertible si y sólo si existe una matriz $B \in \mathcal{M}_{n\times n}$ tal que $AB = I$.*

Demostración: Si A es invertible, basta tomar $B = A^{-1}$.

Supongamos que existe B tal que $AB = I$. Si A no es invertible, entonces su forma reducida por filas, A', tiene una fila de ceros. Además, $A' = PA$, donde P es producto de matrices elementales, y por tanto invertible. Pero entonces tendríamos:

$$A'B = (PA)B = P(AB) = PI = P,$$

donde $A'B$ tiene una fila de ceros (al tenerla A'), y P no tiene una fila de ceros (por ser invertible). Contradicción. $\square$

Teorema 1.24 *Una matriz $A \in \mathcal{M}_{n\times n}$ es invertible si y sólo si es producto de matrices elementales.*

Demostración: Si A es invertible, entonces A^{-1} también lo es. Por lo tanto existe una matriz P, producto de matrices elementales, tal que $PA^{-1} = I$ (ya que I es la forma reducida por filas de A^{-1}). Pero entonces P es la inversa de A^{-1}, es decir, $P = A$. □

Corolario 1.25 *Si $A \in \mathcal{M}_{m\times n}$, y $P \in \mathcal{M}_{n\times n}$ es una matriz invertible, entonces $rg(A) = rg(PA)$.*

Demostración: Como P es invertible, es producto de matrices elementales. Por tanto, PA se obtiene de A al aplicarle una serie de transformaciones elementales, y por tanto deben tener el mismo rango. □

1.6. Transformaciones elementales de columnas.

En esta sección veremos que todas las propiedades que hemos estudiado sobre las filas de una matriz, son también ciertas para sus columnas. Basta trasponer todas las matrices que encontremos. Así, se definen las transformaciones elementales de columnas de forma análoga a las de filas, y se definen las matrices escalonadas o reducidas por columnas, como las traspuestas de las escalonadas o reducidas por filas.

También se tienen las matrices elementales por columnas que, curiosamente, son las mismas que las de filas, ya que la traspuesta de una matriz elemental es otra matriz elemental. La correspondencia de transformaciones y matrices es la siguiente:

1. Matriz que resulta de I al intercambiar las columnas i y j: $T_{i,j}$.
2. Matriz que resulta de I al multiplicar por α la columna i: $M_i(\alpha)$.
3. Matriz que resulta de I al sumarle a la columna i la columna j multiplicada por α: $P_{j,i}(\alpha)$.

Hay que tener cuidado con la última matriz, que es la única que cambia al hablar de columnas en vez de filas. Esto es debido a que $(P_{i,j}(\alpha))^t = P_{j,i}(\alpha)$, mientras que las traspuestas de las demás no cambian.

Un cambio importante al tratar con columnas es el siguiente: Aplicar una transformación elemental por columnas a una matriz equivale a multiplicarla **a la derecha** por la matriz elemental correspondiente. Esto es debido a la propiedad $(AB)^t = B^t A^t$, con lo que, cuando antes multiplicábamos a izquierda, ahora hay que hacerlo a derecha.

Por lo demás, todas las propiedades anteriores se verifican, cambiando filas por columnas. El único problema que tenemos es que hemos definido el **rango** de una matriz usando filas. Veamos que, si lo definimos usando columnas, el rango sigue siendo el mismo.

Lema 1.26 *Si $A \in \mathcal{M}_{m\times n}$, y $Q \in \mathcal{M}_{n\times n}$ es una matriz invertible, entonces rg(A) = rg(AQ).*

DEMOSTRACIÓN: Sea r el rango de A, y sea A' la forma reducida por filas de A. Existe entonces una matriz invertible P tal que $A' = PA$. Por otra parte, se tiene $\mathrm{rg}(A') \geq \mathrm{rg}(A'Q)$, ya que las últimas $m - r$ filas de A' son nulas, y por tanto también lo son las de $A'Q$. Pero entonces:

$$\mathrm{rg}(A) = \mathrm{rg}(A') \geq \mathrm{rg}(A'Q) = \mathrm{rg}(PAQ) = \mathrm{rg}(AQ).$$

La última igualdad se tiene por el corolario 1.25. Tenemos entonces $\mathrm{rg}(A) \geq \mathrm{rg}(AQ)$. La desigualdad opuesta se obtiene fácilmente, aplicando el mismo razonamiento a las matrices AQ y Q^{-1}. Es decir, se tiene $\mathrm{rg}(AQ) \geq \mathrm{rg}(AQQ^{-1}) = \mathrm{rg}(A)$. □

Corolario 1.27 *Si dos matrices A y B son equivalentes por columnas, entonces rg(A) = rg(B).*

Teorema 1.28 *El rango de una matriz es el número de columnas de su forma reducida por columnas.*

DEMOSTRACIÓN: Sea $A \in \mathcal{M}_{m\times n}$, y A' su forma reducida por columnas. Sabemos que existe una matriz invertible Q tal que $A' = AQ$, y por el corolario anterior: $\mathrm{rg}(A) = \mathrm{rg}(A')$. Tenemos que probar entonces que el rango de A' es igual al número de columnas no nulas que tiene, digamos r. Para ello, hallaremos la forma reducida por filas de A'. Cada columna no nula de A' contiene un pivote. Mediante transformaciones de filas, llevamos estos pivotes a las posiciones $(1,1), (2,2), \ldots, (r,r)$. Encima de estos pivotes sólo hay ceros, por tanto, las transformaciones de filas que anulan las entradas inferiores, no alteran estos pivotes. En conclusión, la forma reducida por filas de A' es exactamente:

$$\left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right),$$

donde I_r es la matriz identidad de tamaño r. Por tanto, $\text{rg}(A) = \text{rg}(A') = r$. $\square$

Ahora ya podemos enunciar, usando columnas, todos los resultados que vimos por filas. Las demostraciones son totalmente análogas al caso de filas.

Teorema 1.29 *El rango de una matriz es el rango del sistema de vectores formado por sus columnas.*

Teorema 1.30 *La forma reducida por columnas de una matriz es única.*

Teorema 1.31 *Si $A \in \mathcal{M}_{n\times n}$ es una matriz invertible, su forma reducida por columnas es la matriz identidad I.*

En definitiva, da igual usar filas o columnas para estudiar el rango o la invertibilidad de una matriz. Una última consecuencia de esto es el siguiente resultado:

Teorema 1.32 *Dada $A \in \mathcal{M}_{m\times n}$, se tiene* $\quad rg(A^t) = rg(A)$.

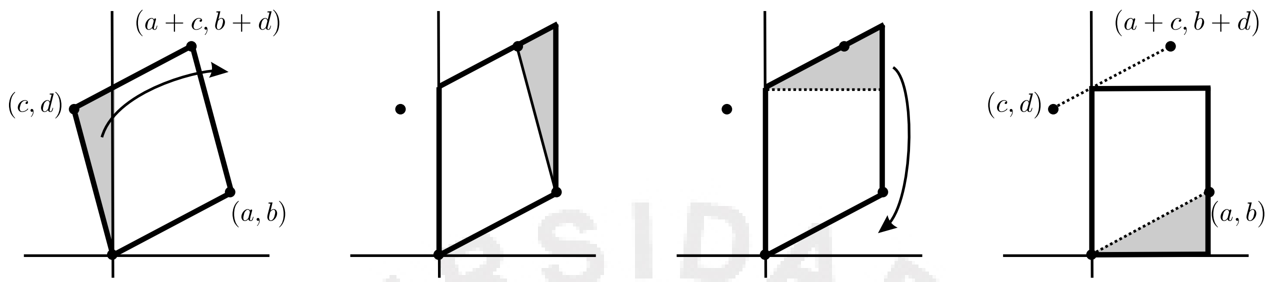
Demostración: La forma reducida por columnas de A^t es la traspuesta de la forma reducida por filas de A. Por tanto, el número de columnas no nulas una (el rango de A^t) es igual al número de filas no nulas de la otra (el rango de A). $\square$

1.7. Determinantes: definición y propiedades. Teorema de Cauchy-Binet.

Para saber lo que son los determinantes, volvamos a estudiar vectores en el plano. Supongamos que tenemos dos vectores $\mathbf{v}_1 = (a, b)$ y $\mathbf{v}_2 = (c, d)$. Estos vectores definen un paralelogramo, cuyos vértices son los puntos $(0, 0)$, (a, b), (c, d) y $(a + c, b + d)$. Pues bien, el **área** de este paralelogramo es:

$$A = ad - bc.$$

En efecto, si dibujamos el paralelogramo, podemos ir transformándolo (como en el dibujo), manteniendo siempre su área, hasta obtener un rectángulo.



La base de este rectángulo es a. Por tanto, para hallar su área sólo hay que conocer su altura. Pero la altura nos la da el punto de corte, con el eje y, de la recta que une (c,d) con $(a+c,b+d)$. O más fácilmente, de la recta que pasa por (c,d) con dirección (a,b). La ecuación de esta recta es:

$$y-d=\frac{b}{a}(x-c).$$

Como buscamos el punto de corte con el eje y, imponemos que $x=0$, y obtenemos la altura:

$$y=d-\frac{bc}{a}.$$

Por tanto, el área del paralelepípedo original es:

$$A=a(d-\frac{bc}{a})=ad-bc.$$

Podemos entonces definir el **determinante** de una matriz 2×2, como el área del paralelogramo definido por sus vectores fila. El determinante de una matriz A se denota $\det A$, o bien cambiando los paréntesis que delimitan la matriz por segmentos verticales. Es decir:

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

Esta definición se puede extender a matrices de tamaño mayor. Por ejemplo, el determinante de una matriz 3×3 es el volumen del paralelepípedo determinado por sus vectores filas. En este caso, se tiene la conocida fórmula:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}=a_1b_2c_3+a_2b_3c_1+a_3b_1c_2-a_3b_2c_1-a_2b_1c_3-a_1b_3c_2.$$

Si agrupamos estos sumandos, sacando factor común las variables a_1, b_1, c_1, obtenemos lo siguiente:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}=a_1\begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}-b_1\begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix}+c_1\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}.$$

Es decir, podemos definir los determinantes de matrices 3×3 usando los determinantes de matrices 2×2. Este proceso se puede generalizar, dando lugar a la definición del determinante de una matriz $n \times n$. Primero hay que definir lo siguiente:

Dada una matriz cuadrada $A \in \mathcal{M}_{n\times n}$, llamamos **submatriz complementaria de** a_{ij}, y la denotamos M_{ij}, a la matriz que se obtiene de A al eliminar su fila i y su columna j.

Llamamos **menor-(i,j)** de A al determinante $\det(M_{ij})$.

Usando estos menores, podemos definir el determinante de una matriz 3×3 como:

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

Para ahorrarnos notación y problemas de signos, definimos lo siguiente:

Dada una matriz cuadrada $A \in \mathcal{M}_{n\times n}$, llamamos **adjunto** o **cofactor** del elemento a_{ij} al escalar $A_{ij} = (-1)^{i+j}\det(M_{ij})$.

El factor $(-1)^{i+j}$ simplemente nos da un signo, que varía si se aumenta i o j en una unidad. Por tanto, podemos volver a definir el determinante de una matriz 3×3 como:

$$\det(A) = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}.$$

Recordemos que, aunque usamos letras mayúsculas por ser la notación clásica, **los adjuntos son escalares**.

Observemos que el adjunto no está bien definido, porque sólo sabemos la definición de los determinantes de matrices 2×2 o 3×3. Pero ahora ya podemos generalizar sin problemas el concepto de determinante:

Dada una matriz $A = (a_{11}) \in \mathcal{M}_{1\times 1}$, se define el **determinante** de A como $\det(A) = \det(a_{11}) = a_{11}$.

Dada una matriz cuadrada $A \in \mathcal{M}_{n\times n}$, con $n > 1$, se llama **determinante** de A, y se denota $\det(A)$ o $|A|$, al escalar definido por:

$$\det(A) = a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n}.$$

Esta forma de definir el determinante se llama **desarrollo por la primera fila**. Observemos que, ahora sí, tanto los determinantes como los adjuntos están bien definidos, ya que para definir el determinante de una matriz de orden n (es decir, $n\times n$), se necesitan adjuntos de orden $n-1$. Para éstos, se necesitan determinantes de orden $n-1$, y así sucesivamente, hasta llegar a los determinantes de orden 1, que están bien definidos por sí mismos. Esto es lo que se llama una definición **recurrente**.

En este tema veremos que los determinantes tienen muchas aplicaciones. Ya hemos visto, por ejemplo, que sirven para calcular áreas de trapecios y volúmenes de paralelepípedos. Pero también se pueden usar para resolver sistemas lineales, comprobar si una matriz es invertible, e incluso calcular su inversa. Comencemos viendo algunas propiedades importantes sobre las columnas de una matriz y su determinante.

Proposición 1.33 *Sea A una matriz cuadrada $n \times n$. Se tiene:*

1. *Si en A se intercambian dos columnas, el determinante cambia de signo.*
2. *Si en A se multiplica una columna por un escalar α, el determinante queda multiplicado por α.*
3. *Si A tiene una columna de ceros, entonces* $\det(A) = 0$.
4. *Si descomponemos la columna j de A en suma de dos vectores, $\mathbf{v}$ y $\mathbf{w}$, y si llamamos A' y A'' a las matrices que resultan de A al sustituir la columna j por $\mathbf{v}$ y $\mathbf{w}$, respectivamente, entonces* $\det(A) = \det(A') + \det(A'')$.
5. *Si A tiene dos columnas iguales, entonces* $\det(A) = 0$.
6. *Si a una columna de A le sumamos otra multiplicada por un escalar, su determinante no cambia.*

Demostración:

1. Esta propiedad se demuestra por inducción en n. Si $n = 1$ la propiedad no tiene sentido. Si $n = 2$, se verifica claramente. Supongamos que es cierta para $n-1$ y probémosla para $n > 2$. Supongamos, en primer lugar, que las columnas que se intercambian son consecutivas: j y $j+1$, y sea A' la matriz resultante de intercambiar estas dos columnas. En ese caso, las submatrices complementarias M_{1k}, con $k \neq j, j+1$, se transforman en las submatrices complementarias M'_{1k} de la matriz A', donde se han intercambiado dos columnas. Por tanto, por hipótesis de inducción, $\det(M_{1k}) = -\det(M'_{1k})$ para $k \neq j, j+1$, es decir $A_{1k} = -A'_{1k}$.

Por otra parte, M_{1j} resulta de eliminar la fila 1 y la columna j de A, que es lo mismo que eliminar la fila 1 y la columna $j+1$ de A'. Es decir, $M_{1j} = M'_{1j+1}$. Análogamente, $M_{1j+1} = M'_{1j}$. Pero entonces, como los índices varían en una unidad, se tiene: $A_{1j} = -A'_{1j+1}$, y $A_{1j+1} = -A'_{1j}$. Además, $a_{1j} = a'_{1j+1}$ y $a_{1j+1} = a'_{1j}$. Por tanto,

$$\det(A) = \left(\sum_{k\neq j,j+1} a_{1k}A_{1k}\right) + a_{1j}A_{1j} + a_{1j+1}A_{1j+1} =$$

$$-\left(\sum_{k\neq j,j+1} a'_{1k}A'_{1k}\right) - a'_{1j+1}A'_{1j+1} - a'_{1j}A'_{1j} = -\det(A').$$

Si, por último, las dos columnas intercambiadas no son consecutivas, observemos que podemos intercambiarlas mediante una sucesión de intercambios de columnas consecutivas (que llamaremos trasposiciones). Sólo hay que ver que el número de estos intercambios es impar. Sean i y j, con $i < j$, las columnas intercambiadas. En primer lugar, llevamos la columna i a la posición j mediante $j-i$ trasposiciones. La columna j habrá quedado en la posición $j-1$, luego harán falta $j-1-i$ trasposiciones para llevarla a la posición i. Una vez hecho esto, todas las columnas están en su lugar, salvo la i y la j que están intercambiadas. Hemos usado, $2i+2j-1$ trasposiciones, luego hemos cambiado el signo de la matriz un número impar de veces. Por tanto, $\det(A) = -\det(A')$.

2. El resultado es evidente para $n = 1$. Supondremos que es cierto para $n-1$, y lo probaremos para n, con $n > 1$. Sea A' la matriz que resulta al multiplicar por α la columna j de A. Se tiene $a'_{1j} = \alpha a_{1j}$, mientras que $M_{1j} = M'_{1j}$, donde esta última matriz es la submatriz complementaria de $a_{1,j}$ en A. Por otra parte, si $k \neq j$, tenemos $a'_{1k} = a_{1k}$, mientras que M'_{1k} se obtiene de M_{1k} al multiplicar una de sus columnas por α. Por hipótesis de inducción, tenemos $\det(M'_{1k}) = \alpha\det(M_{1k})$, es decir, $A'_{1k} = \alpha A_{1k}$. Por tanto,

$$\det(A') = a'_{1j}A'_{1j} + \sum_{k\neq j} a'_{1k}A'_{1k} = \alpha a_{1j}A_{1j} + \sum_{k\neq j} a_{1k}\alpha A_{1k} = \alpha\det(A).$$

3. Sea A' la matriz que resulta al multiplicar por 0 la columna de ceros de A. Obviamente $A' = A$, pero además, al haber multiplicado por 0 una columna, tenemos $\det(A') = 0\det(A) = 0$. Es decir, $\det(A) = 0$.

4. Sean $\mathbf{v} = (v_1, \dots, v_n)$ y $\mathbf{w} = (w_1, \dots, w_n)$. La propiedad es cierta para $n = 1$. Como de costumbre usaremos la inducción, suponiendo que el resultado es cierto para $n-1$, con $n > 1$. Al descomponer la columna j, tenemos: $a_{1j} = v_1 + w_1 = a'_{1j} + a''_{1j}$, y además $M_{1j} = M'_{1j} = M''_{1j}$, donde estas dos últimas matrices son las correspondientes matrices complementarias de A' y A'', respectivamente. Pero también, para $k \neq j$, se tiene $a_{1k} = a'_{1k} = a''_{1k}$, y además M'_{1k} y M''_{1k} son las matrices que se obtienen

al descomponer en dos sumandos una columna de M_{1k}. Por hipótesis de inducción: $\det(M_{1k}) = \det(M'_{1k}) + \det(M''_{1k})$, luego $A_{1k} = A'_{1k} + A''_{1k}$. En resumen:

$$\det(A) = a_{1j}A_{1j} + \sum_{k\neq j} a_{1k}A_{1k} = (a'_{1j} + a''_{1j})A_{1j} + \sum_{k\neq j} a_{1k}(A'_{1k} + A''_{1k})$$

$$= \left(a'_{1j}A'_{1j} + \sum_{k\neq j} a'_{1k}A'_{1k} \right) + \left(a''_{1j}A''_{1j} + \sum_{k\neq j} a''_{1k}A''_{1k} \right) = \det(A') + \det(A'').$$

5. Según la propiedad 1, si intercambiamos las dos columnas iguales, obtenemos una matriz A' tal que $\det(A') = -\det(A)$. Pero claramente $A' = A$, por tanto $\det(A) = -\det(A)$, luego $\det(A) = 0$.

6. Sea B la matriz que resulta de A al sumarle, a su columna i, la columna j multiplicada por α. Según la propiedad 4, $\det(B) = \det(A) + \det(A')$, donde la columna i de A' es igual a la columna j multiplicada por α. Pero entonces, por la propiedad 2, $\det(A') = \alpha \det(A'')$, donde A'' tiene dos columnas iguales, es decir, por la propiedad 5, $\det(A'') = 0$. Uniendo todo esto, se tiene:

$$\det(B) = \det(A) + \det(A') = \det(A) + \alpha \det(A'') = \det(A) + 0 = \det(A).$$

□

Gracias al resultado anterior, hemos visto cómo se comporta el determinante de una matriz si le aplicamos transformaciones elementales de columnas (propiedades 1, 2 y 6). Esto nos va a ayudar a obtener fácilmente muchas más propiedades de los determinantes.

Lema 1.34 *Consideremos la matriz identidad $I \in \mathcal{M}_{n\times n}$. Se tiene:* $\det(I) = 1$.

Demostración: Directa, por inducción en n, a partir de la definición. □

Una matriz $A \in \mathcal{M}_{n\times n}$ se dice **singular** si $\det(A) = 0$. En caso contrario se dice **no singular**.

Teorema 1.35 *Una matriz $A \in \mathcal{M}_{n\times n}$ es no singular si y sólo si $rg(A) = n$, es decir, si y sólo si es invertible.*

Demostración: Si A es no singular, es decir, $\det(A) \neq 0$, aplicar transformaciones elementales de columnas nunca puede anular el determinante, ya que, o bien cambia de signo, o bien se multiplica por un escalar no nulo, o bien se mantiene. Por tanto, la reducida por columnas de A tiene determinante no nulo. Pero esta reducida, o bien es la identidad, con lo que $\mathrm{rg}(A) = n$ y se tiene el resultado, o bien tiene una columna de ceros, con lo que su determinante sería cero, y llegaríamos a una contradicción.

Si, por otra parte, A tiene rango n, entonces su forma reducida por columnas es I. Por tanto, aplicando una serie de transformaciones elementales de columnas a A, obtenemos una matriz, I, cuyo determinante vale 1. Ahora bien, si A fuera singular, es decir, si $\det(A) = 0$, al aplicar cualquier transformación elemental el determinante seguiría siendo cero, luego es imposible. $\square$

Ahora veamos cómo se comporta el determinante con respecto al producto de matrices. Primero estudiaremos las matrices elementales:

Proposición 1.36 *Los determinantes de las matrices elementales son los siguientes:*

1. $\det(T_{ij}) = -1$.
2. $\det(M_i(\alpha)) = \alpha$.
3. $\det(P_{ij}(\alpha)) = 1$.

Demostración: La matriz T_{ij} se obtiene al permutar dos columnas de I, luego su determinante es el opuesto al de I, es decir, -1. La matriz $M_i(\alpha)$ se obtiene al multiplicar la columna i de I por α, luego su determinante es $\alpha \det(I) = \alpha$. Por último, la matriz $P_{ij}(\alpha)$ resulta de sumarle, a la columna j de I, la columna i multiplicada por α, luego su determinante es igual al de I, es decir, 1. $\square$

Proposición 1.37 *Si $A \in \mathcal{M}_{n\times n}$ es una matriz cualquiera, y $P_1, \cdots, P_r \in \mathcal{M}_{n\times n}$ son matrices elementales, entonces* $\det(AP_1 \cdots P_r) = \det(A)\det(P_1)\cdots\det(P_r)$.

Demostración: Lo haremos por inducción en r. Si $r = 1$, la matriz AP_1 es el resultado de aplicar a A la transformación elemental de columnas correspondiente a P_1. Por tanto, el resultado se obtiene de las proposiciones 1.33 y 1.36.

Si $r > 2$ y suponemos el resultado cierto para menos de r matrices elementales, sea $P' = P_1 \cdots P_{r-1}$. Por hipótesis de inducción, tenemos $\det(A) = \det(AP'P_r) = \det(AP')\det(P_r)$.

Pero, de nuevo por hipótesis de inducción, $\det(AP') = \det(A)\det(P_1)\cdots\det(P_{r-1})$, de donde se sigue el resultado. $\square$

Corolario 1.38 *Si $P \in \mathcal{M}_{n\times n}$ es producto de matrices elementales: $P = P_1 \cdots P_r$, entonces* $\det(P) = \det(P_1)\cdots\det(P_r)$.

Demostración: Este es un caso particular del resultado anterior, tomando $A = I$, y recordando que $\det(I) = 1$. $\square$

Teorema 1.39 (Teorema de Cauchy-Binet) *Dadas $A, B \in \mathcal{M}_{n\times n}$, se tiene:*

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

Demostración: Supongamos primero que B es singular. En ese caso $\det(B) = 0$, y B', la forma reducida por columnas de B, tiene una columna de ceros. Pero $B' = BP$, donde P es producto de matrices elementales, luego $B = B'P^{-1}$, donde P^{-1} también es producto de matrices elementales (recordemos que la inversa de una matriz elemental también es una matriz elemental). Por tanto, $AB = AB'P^{-1}$. Como B' tiene una columna de ceros, AB' también la tiene, por tanto $\det(AB') = 0$. Pero sabemos que, al ser P^{-1} producto de matrices elementales, $\det(AB) = \det(AB'P^{-1}) = \det(AB')\det(P^{-1}) = 0$. Por tanto, $\det(AB) = 0$, y el resultado es cierto en este caso.

Supongamos entonces que B es no singular. Entonces tiene rango n, luego es producto de matrices elementales: $B = P_1 \cdots P_r$. Pero en este caso, la proposición 1.37 y el corolario 1.38 nos dicen que $\det(AB) = \det(A)\det(P_1)\cdots\det(P_r) = \det(A)\det(B)$. $\square$

1.8. Desarrollo por filas y columnas. Adjunta e inversa.

Hasta ahora hemos visto una única definición del determinante de una matriz: su desarrollo por la primera fila. En esta sección veremos otras definiciones alternativas, desarrollando por cualquier fila o cualquier columna, y mostraremos que todas las propiedades que hemos visto para columnas se verifican también para filas. Para ello, vamos a empezar estudiando la trasposición de matrices.

Proposición 1.40 *Si $P \in \mathcal{M}_{n\times n}$ es una matriz elemental, entonces* $\det(P) = \det(P^t)$.

Demostración: Recordemos que $(T_{ij})^t = T_{ij}$ y $(M_i(\alpha))^t = M_i(\alpha)$, luego para estos tipos de matrices, el resultado es evidente. Por otra parte, $(P_{ij}(\alpha))^t = P_{ji}(\alpha)$, pero $\det(P_{ij}(\alpha)) = \det(P_{ji}(\alpha)) = 1$, luego el resultado es cierto. $\square$

Teorema 1.41 *Dada $A \in \mathcal{M}_{n\times n}$, se tiene $\det(A^t) = \det(A)$.*

Demostración: Si A es singular, entonces $\mathrm{rg}(A) = \mathrm{rg}(A^t) < n$, por lo que A^t también es singular, es decir, $\det(A) = \det(A^t) = 0$.

Si A es no singular, entonces es producto de matrices elementales: $A = P_1 \cdots P_r$. Pero entonces

$$\det(A^t) = \det((P_1 \cdots P_r)^t) = \det(P_r^t \cdots P_1^t) = \det(P_r^t) \cdots \det(P_1^t)$$

$$= \det(P_r) \cdots \det(P_1) = \det(P_1) \cdots \det(P_r) = \det(A).$$

$\square$

Este teorema nos permite volver a enunciar, para filas, todas las propiedades que vimos sobre columnas de una matriz. Sólo necesitamos darnos cuenta que, las propiedades de las columnas de A son las propiedades de las filas de A^t. Así, se demuestran de forma directa las siguientes propiedades:

Proposición 1.42 *Sea A una matriz cuadrada $n \times n$. Se tiene:*

1. *Si en A se intercambian dos filas, el determinante cambia de signo.*
2. *Si en A se multiplica una fila por un escalar α, el determinante queda multiplicado por α.*
3. *Si A tiene una fila de ceros, entonces $\det(A) = 0$.*
4. *Si descomponemos la fila i de A en suma de dos vectores, $\mathbf{v}$ y $\mathbf{w}$, y si llamamos A' y A'' a las matrices que resultan de A al sustituir la fila i por $\mathbf{v}$ y $\mathbf{w}$, respectivamente, entonces $\det(A) = \det(A') + \det(A'')$.*
5. *Si A tiene dos filas iguales, entonces $\det(A) = 0$.*
6. *Si a una fila de A le sumamos otra multiplicada por un escalar, su determinante no cambia.*

Por tanto, las transformaciones elementales de filas de una matriz actúan sobre el determinante de forma análoga a las transformaciones de columnas. Ya podemos entonces definir el determinante de una matriz usando el desarrollo por cualquier fila o columna.

Teorema 1.43 *Dada $A \in \mathcal{M}_{n\times n}$, se tiene, para cualesquiera i, j, $(1 \leq i, j \leq n)$:*

1. $\det(A) = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}$ *(desarrollo por la fila i).*
2. $\det(A) = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj}$ *(desarrollo por la columna j).*

Demostración: Demostremos primero el desarrollo por la fila i. Sea A' la matriz que se obtiene de A al trasladar su fila i hasta la primera posición. Para ello, hay que usar $i-1$ trasposiciones de filas, por tanto: $\det(A) = (-1)^{i-1}\det(A')$. Ahora bien, $a'_{1j} = a_{ij}$ para todo j. Además, $M'_{1j} = M_{ij}$, donde M'_{ij} es la matriz complementaria de $a'_{i,j}$ en A'. Pero entonces

$$A'_{1j} = (-1)^{1+j}\det(M'_{1j}) = (-1)^{1+j}\det(M_{ij}) = (-1)^{1+j}(-1)^{-i-j}A_{ij} = (-1)^{1-i}A_{ij},$$

es decir, $A_{ij} = (-1)^{i-1}A'_{1j}$. De todos estos resultados, se obtiene:

$$\begin{aligned}
\det(A) &= (-1)^{i-1}\det(A') \\
&= (-1)^{i-1}(a'_{11}A'_{11} + a'_{12}A'_{12} + \cdots + a'_{1n}A'_{1n}) \\
&= a'_{11}(-1)^{i-1}A'_{11} + a'_{12}(-1)^{i-1}A'_{12} + \cdots + a'_{1n}(-1)^{i-1}A'_{1n} \\
&= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}.
\end{aligned}$$

El desarrollo por columnas se demuestra simplemente usando traspuestas. Como se tiene $a^t_{ij} = a_{ji}$, también $A^t_{ij} = A_{ji}$, y además $\det(A^t) = \det(A)$, el desarrollo por la columna j de A es equivalente al desarrollo por la fila j de A^t. $\square$

Veamos ahora cómo estas nuevas definiciones del determinante nos pueden dar otra forma de construir la matriz inversa.

Dada $A \in \mathcal{M}_{n\times n}$, se define la matriz **adjunta** de A, $\mathrm{adj}(A)$, como la matriz cuya entrada (i, j) es el adjunto $A_{i,j}$.

Proposición 1.44 *Dada $A \in \mathcal{M}_{n\times n}$, se tiene $A^{-1} = \dfrac{1}{\det(A)}\,\mathrm{adj}(A)^t$.*

DEMOSTRACIÓN: Para ver que el resultado es cierto, calcularemos la matriz $B = A\ \mathrm{adj}(A)^t$. Primero, para $i = 1\dots,n$, el elemento b_{ii}, de la diagonal principal de B, es el siguiente:

$$b_{ii} = (\text{fila } i \text{ de } A)(\text{columna } i \text{ de } \mathrm{adj}(A)^t) = a_{i1}A_{i1} + \cdots + a_{in}A_{in}.$$

Pero esto es el desarrollo, por la fila i, del determinante de A. Por tanto, $b_{ii} = \det(A)$, para $i = 1,\dots,n$.

Ahora, si $i \neq j$, tenemos:

$$b_{ij} = (\text{fila } i \text{ de } A)(\text{columna } j \text{ de } \mathrm{adj}(A)^t) = a_{i1}A_{j1} + \cdots + a_{in}A_{jn}.$$

Ahora bien, sea A' la matriz que se obtiene de A al sustituir su fila j por la fila i. Es decir, A' tiene dos filas repetidas, la i y la j, por tanto $\det(A') = 0$. Pero el desarrollo de este determinante por la fila j es precisamente el que acabamos de obtener. Es decir, $b_{ij} = 0$. Por tanto, acabamos de demostrar que

$$B = A\ \mathrm{adj}(A)^t = \begin{pmatrix} \det(A) & & & 0 \\ & \det(A) & & \\ & & \ddots & \\ 0 & & & \det(A) \end{pmatrix}$$

Si dividimos esta matriz por $\det(A)$, obtenemos la matriz identidad. Por tanto,

$$\frac{1}{\det(A)} A\ \mathrm{adj}(A)^t = I \quad \Rightarrow \quad A^{-1} = \frac{1}{\det(A)}\ \mathrm{adj}(A)^t.$$

□

1.9. Cálculo de determinantes.

Hasta ahora, las única manera que conocemos de calcular un determinante, consiste en desarrollarlo por una fila o una columna de la matriz. Sin embargo, este procedimiento de cálculo no es nada eficaz, ya que, para calcular el determinante de una matriz $n \times n$, hay que calcular n determinantes de matrices $(n-1)\times(n-1)$, y para cada uno de estos, hay que calcular $(n-1)$ determinantes de matrices $(n-2)\times(n-2)$, y así sucesivamente. Por tanto, el número de operaciones que hay que efectuar es del orden de $n!$.

Hay un método mucho más rápido y simple para calcular un determinante, en el que se usan, una vez más, las transformaciones y las matrices elementales. Comenzaremos por ver dos tipos de matrices cuyo determinante es muy sencillo:

Se dice que $A \in \mathcal{M}_{n\times n}$ es **triangular inferior** si $a_{ij} = 0$ para todo $i < j$.
Se dice que $A \in \mathcal{M}_{n\times n}$ es **triangular superior** si $a_{ij} = 0$ para todo $i > j$.

El siguiente resultado es evidente a partir de las definiciones:

Proposición 1.45 *Se tiene:*

- *Una matriz cuadrada escalonada por filas es triangular superior.*
- *Una matriz cuadrada escalonada por columnas es triangular inferior.*
- *La traspuesta de una matriz triangular superior es triangular inferior, y viceversa.*

Calculemos ahora el determinante de las matrices triangulares:

Proposición 1.46 *Si $A \in \mathcal{M}_{n\times n}$ es triangular inferior o superior, entonces su determinante es el producto de los elementos de su diagonal principal. Es decir,* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Demostración: Procedemos por inducción en n. El resultado es claramente cierto si $n = 1$ o $n = 2$. Supongamos entonces que $n > 2$, y que el resultado es cierto para $n - 1$.

Supongamos primero que A es triangular inferior. Entonces, todos los elementos de su primera fila son nulos salvo, a lo sumo, a_{11}. Por tanto, $\det(A) = a_{11}A_{11} = a_{11}\det(M_{11})$. Pero M_{11} es también triangular inferior, y los elementos de su diagonal principal son $a_{22}, \cdots, a_{nn}$. Por tanto, por hipótesis de inducción, $\det(M_{11}) = a_{22}\cdots a_{nn}$, y el resultado es cierto.

Por último, si A es triangular superior, la primera columna de M_{1j} es una columna de ceros, para todo $j = 2, \ldots, n$. Por tanto, $A_{1j} = 0$ si $j > 1$. Luego $\det(A) = a_{11}A_{11} = a_{11}\det(M_{11})$. Pero M_{11} es triangular superior, así que podemos aplicar, igual que antes, la hipótesis de inducción para obtener el resultado. $\square$

Ya tenemos por tanto un método rápido para el cálculo de determinantes:

Método para calcular determinantes: Dada $A \in \mathcal{M}_{n\times n}$, usamos el método de eliminación de Gauss para hallar una forma escalonada A' de A. Vamos recordando, durante el proceso, las transformaciones elementales utilizadas. El determinante de A es el producto de los determinantes de las matrices elementales correspondientes, multiplicado por los elementos de la diagonal principal de A'.

Es decir, si $A = P_1 \cdots P_r A'$, donde $P_1, \ldots, P_r$ son las matrices elementales que se emplean en el método de Gauss, y A' es escalonada por filas, se tiene:

$$\det(A) = \det(P_1) \cdots \det(P_r) \det(A'),$$

pero los determinantes de cada P_i son conocidos y, como A' es triangular superior, su determinante es muy fácil de calcular. Así, tenemos:

$$\det(A) = \det(P_1) \cdots \det(P_r) a'_{11} \cdots a'_{nn}.$$

Nota: En la práctica, el cálculo del determinante de $A \in \mathcal{M}_{n\times n}$ se efectúa eligiendo una fila o columna, preferiblemente que tenga algún 0 o algún 1, y consiguiendo mediante transformaciones elementales que todas las entradas de esa fila o esa columna sean nulas, excepto una (como máximo). Entonces se desarrolla el determinante por esa fila o columna, con lo que el cálculo queda reducido a una matriz más pequeña que la anterior. Continuando este proceso, obteniendo a cada paso una matriz más pequeña, se termina simplemente calculando el determinante de una matriz 2×2.

1.10. Rango y menores. Método del orlado.

En esta sección daremos una nueva caracterización del rango de una matriz, utilizando los determinantes. Ya sabemos que, dada una matriz $A \in \mathcal{M}_{n\times n}$, $\det(A) \neq 0$ si y sólo si $\mathrm{rg}(A) = n$. Pero no sabemos nada sobre el rango de A si $\det(A) = 0$, o si la matriz no es cuadrada. Para poder precisar más, definiremos los *menores* de una matriz, de los que ya vimos algunos ejemplos en secciones precedentes.

Dada $A \in \mathcal{M}_{m\times n}$, y dadas p filas $1 \leq i_1 < i_2 < \cdots < i_p \leq m$ y p columnas $1 \leq j_1 < j_2 < \cdots < j_p \leq n$, se llama **submatriz cuadrada** de orden p de A, determinada por estas p filas y p columnas, a la matriz M cuyas entradas son los elementos de A que pertenecen, a la vez, a una de estas filas y a una de estas columnas.

Se llama **menor** de orden p de A, correspondiente a estas filas y estas columnas, al **determinante** de M.

Aunque A no sea cuadrada, notemos que las submatrices cuadradas de orden p sí lo son, y por tanto se puede calcular su determinante. Podemos entonces definir el rango de una matriz en función de sus menores.

Teorema 1.47 *Dada $A \in \mathcal{M}_{m\times n}$, entonces $rg(A) = r$ si y sólo si A tiene algún menor no nulo de orden r, y todos los menores de A de orden mayor que r son nulos.*

Demostración: Supongamos que $\mathrm{rg}(A) = r$. Entonces sabemos que tiene r filas linealmente independientes. Sean $i_1, \ldots, i_r$ dichas filas. La matriz A' formada por estas r filas tiene, por tanto rango r. Pero eso quiere decir que A' tiene r columnas linealmente independientes, digamos $j_1, \ldots, j_r$. Por tanto, la matriz M formada por estas columnas de A' tiene rango r. Pero además, M es una submatriz cuadrada de A, de orden r, asociada a estas filas y estas columnas; y como tiene rango r, su determinante es no nulo. Por tanto, existe un menor no nulo de orden r.

Si hubiera un menor no nulo de orden $p > r$, las filas correspondientes a ese menor formarían una matriz $A' \in \mathcal{M}_{p\times n}$, que tendría una submatriz $p\times p$ de determinante no nulo. Es decir, A' tendría p columnas linealmente independientes. En ese caso, A' tendría rango p, luego sus p filas serían linealmente independientes, y por tanto, habría p filas de A linealmente independientes. Esto contradice el hecho de que $\mathrm{rg}(A) = r$.

Supongamos ahora que A tiene algún menor no nulo de orden r, y todos los menores de A de orden mayor que r son nulos. Según hemos demostrado antes, si $\mathrm{rg}(A) = p > r$, entonces A tendría un menor no nulo de orden p, lo cual es imposible. Y si $\mathrm{rg}(A) = q < r$, entonces todos los menores de A de orden mayor que q serían nulos. Pero esto también es imposible, ya que sabemos que tiene un menor no nulo de orden r. $\square$

Terminemos este tema dando un método para calcular el rango de una matriz, usando menores. Hay que decir que este método no es el más eficaz, ya que usando el método de eliminación de Gauss, que es más rápido, obtenemos una matriz escalonada, en la que el número de filas no nulas es el rango de la matriz. Sin embargo, el método que vamos a dar puede servir para estudiar los vectores fila o vectores columna de una matriz, ya que, a diferencia del método de Gauss, éste no los va a modificar.

Método del orlado, para calcular el rango de una matriz $A \in \mathcal{M}_{m\times n}$.

1. Si A es una matriz de ceros, entonces $\text{rg}(A) = 0$.
2. Si no, elegimos un elemento $a_{i_1 j_1} \neq 0$.
3. Buscamos otra fila i_2, y otra columna j_2, tal que el menor de orden 2 correspondiente a las filas i_1, i_2 y a las columnas j_1, j_2 sea no nulo. Si no existe, entonces $\text{rg}(A) = 1$. Si existe, recordamos los datos $(i_1, i_2;\ j_1, j_2)$.
4. Continuamos con el mismo proceso: si conocemos los índices $(i_1, \cdots, i_p;\ \ j_1, \cdots, j_p)$ tales que el menor correspondiente es no nulo, buscamos una fila i_{p+1}, y una columna j_{p+1}, tales que el menor asociado a $(i_1, \cdots, i_{p+1};\ j_1, \cdots, j_{p+1})$ sea no nulo. Si no existe, entonces $\text{rg}(A) = p$. Si existe, repetimos este paso, para un orden mayor.
5. En algún momento no podremos seguir aumentando el orden, y habremos obtenido el rango de A.

Proposición 1.48 *El método del orlado funciona.*

Demostración: No es evidente que este método funciona: Hay que demostrar que, dada una matriz $A \in \mathcal{M}_{m\times n}$, si tenemos un menor no nulo de orden p, y el rango de A es mayor que p, entonces existe un menor no nulo de orden $p+1$ *que contiene al anterior.*

Supongamos entonces que $\text{rg}(A) > p$, y que tenemos un menor no nulo de orden p. Las p filas correspondientes a ese menor, digamos $i_1, \ldots, i_p$, son entonces linealmente independientes, y también lo son las p columnas, $j_1, \ldots, j_p$. Sea $i \notin \{i_1, \ldots, i_p\}$. Supongamos que la fila i depende linealmente de las filas $i_1, \ldots, i_p$. Es decir, si llamamos $\mathbf{f}_i$ al vector determinado por la fila i, tendremos:

$$\mathbf{f}_i = \alpha_1 \mathbf{f}_{i_1} + \cdots + \alpha_p \mathbf{f}_{i_p}.$$

En ese caso, podemos transformar la fila i, mediante transformaciones elementales de filas (restándole cada fila $\mathbf{f}_{i_k}$ multiplicada por α_k), hasta convertirla en una fila de ceros. Si esto ocurriera para todo $i \notin \{i_1, \ldots, i_p\}$, obtendríamos una matriz A', equivalente por filas a A (luego $\text{rg}(A') = \text{rg}(A)$), que sólo tendría p filas distintas de cero. En ese caso tendríamos $\text{rg}(A) = p$, lo que no es posible.

Por tanto, debe existir una fila, i_{p+1}, que no dependa linealmente de las filas $i_1, \ldots, i_p$. En ese caso, las filas $i_1, \ldots, i_{p+1}$ de A son linealmente independientes. Sea $A'' \in \mathcal{M}_{(p+1)\times n}$ la matriz formada por las filas $i_1, \ldots, i_{p+1}$ de A. Sabemos que $\text{rg}(A'') = p+1$, y también conocemos

p columnas, $j_1, \dots, j_p$ que son linealmente independientes. Ahora podemos proceder como antes: si una columna $j \notin \{j_1, \dots, j_p\}$ depende linealmente de estas p columnas, podremos hacerla nula mediante transformaciones elementales por columnas. Si esto pasara para todo $j \notin \{j_1, \dots, j_p\}$, obtendríamos una matriz A''' equivalente por columnas a A'', con rango p. Como esto es imposible, existirá una columna j_{p+1} que no dependa linealmente de la columnas $j_1, \dots, j_p$, y por tanto el determinante de la submatriz cuadrada formada por las filas $i_1, \dots, i_{p+1}$, y las columnas $j_1, \dots, j_{p+1}$ de A, es no nulo. $\square$

1.11. Sistemas de ecuaciones lineales.

Comenzaremos viendo un ejemplo del tipo de ecuaciones que vamos a estudiar:

$$\begin{cases} 2x + y = 5 \\ x - y = 1 \end{cases}$$

Se trata de un sistema lineal de dos ecuaciones con dos incógnitas. Este sistema se puede ver desde varias perspectivas:

- Desde el punto de vista *geométrico*, cada una de las dos ecuaciones representa una recta en el plano. Resolver el sistema consiste en hallar (si los hay) los puntos de corte de las dos rectas. Esa es la razón de que estos sistema se llamen *lineales*.
- Desde el punto de vista *algebraico*, el problema consiste simplemente en hallar dos números, x e y, que satisfagan las dos igualdades. Las ecuaciones son *lineales* porque cada término (excepto los términos independientes) tiene grado 1.

Si nos quedamos en el marco algebraico, nada nos impide generalizar el concepto de *ecuación lineal* a más de dos incógnitas, y el de *sistema lineal* a más de dos ecuaciones. Así, tenemos las siguientes definiciones:

Ecuación lineal: Es una expresión de la forma

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \tag{1}$$

donde $a_1, a_2, \dots, a_n$ y b son números conocidos, y $x_1, x_2, \dots, x_n$ son incógnitas.

Una **solución** de la ecuación lineal (1) es una serie de números $\alpha_1, \dots, \alpha_n$, que la satisfagan, es decir, que verifiquen:

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_n\alpha_n = b.$$

Sistema lineal: Un sistema lineal de m ecuaciones con n incógnitas es una expresión de la forma:

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

donde cada fila es una ecuación lineal diferente, aunque las n incógnitas, $x_1, \dots, x_n$, son las mismas para todas ellas.

Una **solución** del sistema lineal (2) es una serie de números $\alpha_1, \dots, \alpha_n$, que satisfagan las m ecuaciones, es decir, tales que

$$\left\{\begin{array}{ccccccccc} a_{11}\alpha_1 & + & a_{12}\alpha_2 & + & \cdots & + & a_{1n}\alpha_n & = & b_1 \\ a_{21}\alpha_1 & + & a_{22}\alpha_2 & + & \cdots & + & a_{2n}\alpha_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}\alpha_1 & + & a_{m2}\alpha_2 & + & \cdots & + & a_{mn}\alpha_n & = & b_m. \end{array}\right.$$

Diremos que un sistema lineal es:

- **compatible**: si admite alguna solución,
- **incompatible**: si no la admite.

Dado un sistema compatible, diremos que es

- **compatible determinado**: si admite una única solución,
- **compatible indeterminado**: si admite más de una.

En este último caso veremos que admite infinitas soluciones.

Ejemplos:

- Intersección de dos rectas en el plano. Tres casos posibles, según las rectas sean
  1. Secantes $\rightsquigarrow$ Solución única.
  2. Paralelas $\rightsquigarrow$ Ninguna solución.
  3. Coincidentes $\rightsquigarrow$ Infinitas soluciones.
- Intersección de dos o tres planos en el espacio (de dimensión 3).
- Algunas aplicaciones:
  - En Física: Cálculo de los voltajes de nudos en un circuito de corriente continua.
  - En Arquitectura: Cálculo de estructuras de edificios.
  - En Economía: Modelo de Leontiev de entradas y salidas.

Un caso especial importante de sistemas lineales es el siguiente:

Un sistema lineal se dice **homogéneo** si todos sus términos independientes son nulos. Es decir, si es de la forma:

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

Nota: Hasta ahora hemos hablado de **números**, sin especificar de qué tipo de números se trata. En esta asignatura usaremos, salvo que se especifique lo contrario, los números racionales ($\mathbb{Q}$), reales ($\mathbb{R}$) o complejos ($\mathbb{C}$). Aunque también se puede utilizar cualquier otro tipo de "números", siempre que verifiquen una serie de condiciones, que veremos más adelante. Por tanto, a partir de ahora, en vez de *números* diremos **escalares**, y al conjunto de números que estemos utilizando lo llamaremos **cuerpo de escalares** o simplemente **cuerpo**. La definición de cuerpo se verá más adelante. Por ahora basta con pensar que un escalar es un número racional, real o complejo.

1.12. Método de eliminación de Gauss.

Para resolver sistemas de ecuaciones de cualquier tipo, una de las estrategias más utilizadas consiste en ir *simplificando* el sistema, de manera que sea cada vez más fácil de resolver, pero que siga teniendo *las mismas soluciones* que el sistema original. Por tanto, debemos usar el siguiente concepto:

Dos sistemas de ecuaciones lineales son **equivalentes** si toda solución de uno es también solución del otro.

Algunas de las operaciones que se pueden aplicar a un sistema, de manera que se obtenga un sistema equivalente, son las siguientes:

1. Intercambiar dos ecuaciones.
2. Multiplicar una ecuación por un escalar distinto de cero.
3. Añadir a una ecuación un múltiplo no nulo de otra.

Observemos cómo estas tres operaciones recuerdan mucho a las transformaciones elementales que definimos para las filas de una matriz. En efecto, se trata esencialmente de las mismas operaciones, puesto que a todo sistema lineal de m ecuaciones y n incógnitas, de la forma

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

le podemos asociar su **matriz de coeficientes**:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

y añadiéndole la columna de términos independientes, obtendremos la llamada **matriz ampliada**:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

Esta matriz contiene toda la información del sistema: cada fila corresponde a una ecuación. Así, aplicar las operaciones arriba indicadas a un sistema, equivale a aplicar las correspondientes *transformaciones elementales de filas* a su matriz ampliada.

Curiosamente, si recordamos el *método de Gauss* para transformar una matriz en otra escalonada por filas, el mismo método aplicado a sistemas lineales nos servirá para resolver el sistema.

El **Método de eliminación de Gauss**, para resolver un sistema lineal, consiste en aplicar al sistema las tres operaciones básicas anteriores, de la siguiente forma:

> **Paso 1:** Si es necesario, intercambiar la primera ecuación con otra, para que x_1 aparezca en la primera ecuación.
>
> **Paso 2:** Eliminar x_1 de cada ecuación (salvo la primera), sumándole un múltiplo adecuado de la primera ecuación.
>
> **Paso 3:** Ignorando temporalmente la primera ecuación, repetir todo el proceso con las restantes ecuaciones, que forman un sistema de $m-1$ ecuaciones con menos de n incógnitas.

Al terminar de aplicar el método de eliminación de Gauss, habremos transformado el sistema en otro equivalente, pero que va a ser muy fácil de resolver, puesto que su matriz ampliada es escalonada. Como veremos, podremos encontrarnos tres casos diferentes, así que vamos a estudiar tres ejemplos, uno de cada caso.

Ejemplo 1.49 *Resolver el sistema:*

$$\left\{\begin{array}{rcrcrcr} x_1 &+& x_2 &-& x_3 &=& 1 \\ 2x_1 &+& x_2 &+& 3x_3 &=& 2 \\ 3x_1 &+& 2x_2 &+& 2x_3 &=& -1 \end{array}\right. \quad \rightsquigarrow \quad \left\{\begin{array}{rcrcrcr} x_1 &+& x_2 &-& x_3 &=& 1 \\ &-& x_2 &+& 5x_3 &=& 0 \\ &&&&0 &=& -4\ . \end{array}\right.$$

La última ecuación queda $0=-4$, por tanto este sistema es imposible de resolver:
El sistema **no tiene solución**.

Ejemplo 1.50 *Resolver el sistema:*

$$\left\{\begin{array}{rcrcrcr} x_1 &+& x_2 &-& x_3 &=& 1 \\ 2x_1 &+& x_2 &-& 3x_3 &=& -2 \\ x_1 &+& 2x_2 &+& 2x_3 &=& 9 \end{array}\right. \quad \rightsquigarrow \quad \left\{\begin{array}{rcrcrcr} x_1 &+& x_2 &-& x_3 &=& 1 \\ &-& x_2 &-& x_3 &=& -4 \\ &&&&2x_3 &=& 4\ . \end{array}\right.$$

De la última ecuación se obtiene $x_3 = 2$. Sustituyendo en la segunda ecuación, se tiene $x_2 = 2$. Por úlitmo, sustituyendo estos dos valores en la primera ecuación, queda $x_1 = 1$. Por tanto, el sistema **tiene solución única**.

Ejemplo 1.51 *Resolver el sistema:*

$$\left\{\begin{array}{rcrcrcl} x_1 &+& x_2 &-& x_3 &=& 1 \\ 2x_1 &+& x_2 &+& 3x_3 &=& 2 \\ 3x_1 &+& 3x_2 &+& 2x_3 &=& 3 \end{array}\right. \quad \rightsquigarrow \quad \left\{\begin{array}{rcrcrcl} x_1 &+& x_2 &-& x_3 &=& 1 \\ &-& x_2 &+& 5x_3 &=& 0 \\ &&&&0 &=& 0\ . \end{array}\right.$$

La última ecuación, $0 = 0$, se verifica siempre. La segunda ecuación nos dice que $x_2 = 5x_3$. Sustituyendo esto en la primera ecuación, obtenemos $x_1 = 1 - 4x_3$. Ya no quedan más condiciones que imponer, por tanto, tenemos libertad para elegir el valor de x_3. Si le damos, por ejemplo, el valor $x_3 = 1$, obtendremos la solución $x_1 = -3$, $x_2 = 5$, $x_3 = 1$. Si le damos el valor $x_3 = 0$, obtendremos la solución $x_1 = 1$, $x_2 = 0$, $x_3 = 0$. Y así podríamos seguir indefinidamente. Es decir, tendremos una solución distinta para cada valor que le demos a x_3. Por tanto, el sistema **tiene infinitas soluciones**.

Estudiemos ya estos tres casos de forma general. Tomamos el sistema (2) y le aplicamos el método de eliminación de Gauss. Observemos lo siguiente: el primer término no nulo (si existe) de la ecuación i será de la forma $c_{ij}x_j$, para un cierto valor j. Como este índice j depende de i, lo llamaremos j_i. Supongamos que, después de la eliminación de Gauss, nos quedan r ecuaciones no nulas. En este caso tendremos, por construcción: $j_1 < j_2 < \cdots < j_r$. Es decir, el primer elemento no nulo de cada fila estará *más a la derecha* que el de la fila anterior. El sistema obtenido tendrá, por tanto, la siguiente forma:

$$\left\{\begin{array}{lllllllll} c_{11}x_1 + \cdots + & c_{1j_2}x_{j_2} + \cdots + & c_{1j_r}x_{j_r} & + \cdots + & c_{1n}x_n & = & d_1 \\ & c_{2j_2}x_{j_2} + \cdots + & c_{2j_r}x_{j_r} & + \cdots + & c_{2n}x_n & = & d_2 \\ & & \vdots & & \vdots & & \vdots \\ & & c_{rj_r}x_{j_r} & + \cdots + & c_{rn}x_n & = & d_r \\ & & & & 0 & = & d_{r+1} \\ & & & & 0 & = & 0 \\ & & & & \vdots & & \vdots \\ & & & & 0 & = & 0. \end{array}\right.$$

Se nos pueden presentar ahora tres casos:

Caso 1: $\boxed{\text{El término independiente } d_{r+1} \neq 0.}$

En este caso, la ecuación $r + 1$ no puede cumplirse nunca. Por tanto, **no existe solución** para el sistema inicial.

Caso 2: $\boxed{d_{r+1} = 0 \text{ y } r = n.}$

En este caso, hay tantas ecuaciones no nulas como incógnitas. Pero como sabemos que $j_1 < j_2 < \cdots < j_n$, el sistema habrá quedado de la forma:

$$\left\{\begin{array}{rcrcrcrcl} c_{11}x_1 & + & c_{12}x_2 & + \cdots + & c_{1\,n-1}x_{n-1} & + & c_{1n}x_n & = & d_1 \\ & & c_{22}x_2 & + \cdots + & c_{2\,n-1}x_{n-1} & + & c_{2n}x_n & = & d_2 \\ & & & \ddots & \vdots & & \vdots & & \vdots \\ & & & & c_{n-1\,n-1}x_{n-1} & + & c_{nn}x_n & = & d_{n-1} \\ & & & & & & c_{nn}x_n & = & d_n \\ & & & & & & 0 & = & 0 \\ & & & & & & \vdots & & \vdots \\ & & & & & & 0 & = & 0. \end{array}\right.$$

De la n-ésima ecuación, deducimos que el único valor posible para x_n es $x_n = \dfrac{d_n}{c_{nn}}$. Sustituyendo el valor de x_n en la ecuación $n-1$, vemos que también hay un único valor posible para x_{n-1}. Podemos seguir así, sustituyendo y despejando, ya que en la ecuación i, tendremos:

$$x_i = \frac{d_i - c_{i,\,i+1}x_{i+1} - \cdots - c_{in}x_n}{c_{ii}}.$$

Si sabemos que las variables $x_{i+1}, \ldots, x_n$ deben tomar un único valor, pasará lo mismo con x_i. Cuando lleguemos a la primera ecuación, habremos obtenido un único valor para cada variable $x_n, x_{n-1}, \cdots, x_2, x_1$. Es decir, en este caso el sistema tiene **solución única**.

Caso 3: $\boxed{d_{r+1} = 0 \text{ y } r < n.}$

En este caso, tendremos unas variables especiales, $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$, que son las correspondientes al primer término no nulo de cada fila. Vamos a llamarlas *variables pivote*. Procedemos ahora de forma parecida al caso anterior. En la ecuación r, la única variable pivote que aparece es x_{j_r}. Podemos despejarla, por tanto, en función de las variables no-pivote:

$$x_{j_r} = \frac{d_r - c_{rj_r+1}x_{j_r+1} - \cdots - c_{rn}x_n}{c_{rj_r}}.$$

Del mismo modo, en la ecuación $r-1$, podemos despejar $x_{j_{r-1}}$ en función de las variables x_k, con $k > j_{r+1}$. La única variable pivote que aparece es x_{j_r}. Pero ésta ya sabemos escribirla en función de las variables no-pivote. Por tanto, sustituimos su valor, y sabremos escribir $x_{j_{r-1}}$ en función de las variables no-pivote.

Continuamos de este modo, de forma ascendente, y al finalizar sabremos escribir todas las variables pivote $x_{j_1}, x_{j_2}, \ldots, x_{j_r}$, en función de las no-pivote. Es importante darse cuenta de que hemos usado todas las ecuaciones del sistema. Es decir, el sistema no nos impone ninguna condición más. Por tanto, si le damos **cualquier valor** a las variables no-pivote, habremos determinado también el valor de las variables pivote, y por tanto habremos obtenido una solución del sistema. Pero, al tener libertad absoluta para elegir los valores de las variables no-pivote, deducimos que el sistema tiene **infinitas soluciones**. Acabamos de demostrar lo siguiente:

Teorema 1.52 *Si un sistema lineal es compatible indeterminado, entonces tiene infinitas soluciones.*

Nota: Si el sistema es homogéneo, es imposible que se dé el caso 1. Por tanto, **un sistema homogéneo es siempre compatible**. Es más, si el sistema homogéneo es compatible determinado, entonces su única solución es $x_1 = x_2 = \cdots = x_n = 0$, llamada **solución trivial**.

1.13. Método de Gauss-Jordan. Teorema de Rouché-Frobenius.

Al igual que el método de Gauss, el método de Gauss-Jordan para calcular una reducida por filas también puede aplicarse a los sistemas de ecuaciones lineales, para transformarlo en otro sistema equivalente que se resuelva fácilmente.

Método de eliminación de Gauss-Jordan para resolver un sistema lineal:

Paso 1: Aplicar al sistema el método de Gauss.

Paso 2: Multiplicar cada ecuación no nula por un escalar conveniente, de manera que el coeficiente de la variable pivote sea 1.

Paso 3: Comenzando por el pivote más a la derecha, x_{j_r}, eliminar esta variable de cada ecuación (salvo la ecuación r), sumándole un múltiplo conveniente de la ecuación r. Realizar la misma operación con todos los pivotes, de derecha a izquierda.

Nota: Podríamos haber aplicado el método de Gauss-Jordan, de forma clásica, haciendo ceros en las columnas pivote de izquierda a derecha. Hemos preferido hacerlo de derecha a izquierda, ya que se realizan muchas menos operaciones básicas (sumas y multiplicaciones). Por tanto, al implementarlo en un ordenador, resulta mucho más rápido para ejemplos grandes.

Veamos cómo funciona este método con un ejemplo:

Ejemplo 1.53 *Resolver el sistema:*

$$\left\{\begin{array}{rcrcrcrcr} x_1 &+& x_2 &+& x_3 &+& 4x_4 &=& 4 \\ 2x_1 &+& 5x_2 &+& 5x_3 &-& x_4 &=& 11 \\ x_1 &+& 2x_2 &+& x_3 &+& x_4 &=& 3 \end{array}\right. \rightsquigarrow \left\{\begin{array}{rrrrcr} x_1 & & & +\ 7x_4 &=& 3 \\ & x_2 & & -\ 3x_4 &=& -1 \\ & & x_3 & &=& 2\ . \end{array}\right.$$

Para dar la solución del sistema sólo hay que despejar cada variable pivote, con lo que se obtiene:

$$\begin{aligned} x_1 &= 3 - 7x_4 \\ x_2 &= -1 + 3x_4 \\ x_3 &= 2. \end{aligned}$$

Este método, en realidad, realiza las mismas operaciones que el método anterior, cuando íbamos despejando y sustituyendo cada variable pivote. En vez de eso, se aplican más operaciones elementales, de forma que cada variable pivote aparezca sólo en una ecuación, con coeficiente 1. Por tanto, se puede escribir directamente en función de las variables no pivote.

Notemos que el número de variables pivote (r) no cambia. Además, se tienen las mismas tres posiblidades que antes. Es decir, si llamamos d_i al término independiente de la ecuación i, después de aplicar el método de Gauss-Jordan, se tiene:

- Si $d_{r+1} \neq 0$, el sistema no tiene solución.
- Si $d_{r+1} = 0$ y $r = n$, el sistema tiene solución única.
- Si $d_{r+1} = 0$ y $r < n$, el sistema tiene infinitas soluciones.

Este resultado se puede *traducir* de la siguiente manera: Sabemos que el sistema tiene r variables pivote, luego el rango de la matriz de coeficientes es igual a r. La matriz ampliada, por su parte, puede tener rango bien $r+1$, bien r, dependiendo de si admite un pivote más, es decir, de si $d_{r+1} \neq 0$ o bien $d_{r+1} = 0$. Por tanto, las tres posiblidades arriba descritas se pueden expresar usando los rangos de la matriz de coeficientes y la matriz ampliada, con lo que obtenemos el siguiente resultado fundamental del álgebra lineal:

Teorema 1.54 (Teorema de Rouché-Frobenius) *Dado un sistema lineal de m ecuaciones con n incógnitas, sea A su matriz de coeficientes y A' su matriz ampliada. Se tiene:*

- *El sistema es* **incompatible** *si y sólo si* $\quad rg(A) < rg(A')$.

- *El sistema es* **compatible determinado** *si y sólo si* $rg(A) = rg(A') = n$.
- *El sistema es* **compatible indeterminado** *si y sólo si* $rg(A) = rg(A') < n$.

1.14. Regla de Cramer.

Terminamos este tema viendo cómo las propiedades de los determinantes también pueden ayudar a resolver los sistemas lineales. En primer lugar, un sistema lineal puede verse como un producto de matrices. De hecho, si llamamos A a la matriz de coeficientes, $\mathbf{x}$ al vector columna cuyas entradas son las incógnitas, y $\mathbf{b}$ al vector columna cuyas entradas son los términos independientes, entonces el sistema puede escribirse:

$$A\mathbf{x} = \mathbf{b}.$$

Es decir,

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

En el caso particular en el que A sea una matriz cuadrada (tantas ecuaciones como incógnitas), y no singular, el sistema se puede resolver usando inversas o determinantes.

En efecto, si A es no singular, entonces es invertible, y por tanto podemos despejar:

$$A\mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = A^{-1}\mathbf{b}.$$

En otras palabras, cada coordenada del vector $A^{-1}\mathbf{b}$ nos da el valor de cada incógnita. Esto coincide con lo que sabemos: como A es no singular, entonces $\mathrm{rg}(A) = n$, luego el sistema es compatible determinado.

Pero veremos otra forma de resolverlo: la **regla de Cramer**, que nos va a permitir calcular explícitamente el valor de cada incógnita, por medio de los determinantes.

Regla de Cramer: Consideremos el sistema lineal $A\mathbf{x} = \mathbf{b}$, donde $A \in \mathcal{M}_{n\times n}$ es no singular. Para $i = 1, \dots, n$, sea B_i la matriz que se obtiene de A al sustituir su columna i por el vector $\mathbf{b}$. Entonces, la solución del sistema viene dada por:

$$x_i = \frac{\det(B_i)}{\det(A)}, \qquad \forall i = 1, \dots, n.$$

DEMOSTRACIÓN: **(de la regla de Cramer)** Sabemos que $\mathbf{x} = A^{-1}\mathbf{b}$, luego la coordenada x_i será el producto de la fila i de A^{-1} por el vector columna $\mathbf{b}$. Como sabemos que $A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)^t$, la fila i de esta matriz será:

$$\left(\frac{A_{1i}}{\det(A)}, \frac{A_{2i}}{\det(A)}, \ldots, \frac{A_{ni}}{\det(A)}\right).$$

Por tanto, tendremos

$$x_i = \frac{1}{\det(A)} \, (b_1 A_{1i} + b_2 A_{2i} + \cdots + b_n A_{ni}).$$

Pero el factor entre paréntesis es el desarrollo por la columna i del determinante de la matriz B_i, por tanto $x_i = \frac{\det(B_i)}{\det(A)}$, como queríamos demostrar. $\square$

Nota: La regla de Cramer es un resultado clásico, que puede tener aplicaciones teóricas. Pero en la práctica, si se quiere resolver un sistema lineal, es mucho más eficaz calcular la escalonada por filas de la matriz ampliada, usando el método de Gauss, e ir despejando las variables pivote. Además, este último método sirve para cualquier sistema, mientras que la regla de Cramer sólo es válida para matrices cuadradas no singulares.

Ahora que ya sabemos manejar los vectores y las matrices, y conocemos muchas de sus propiedades, vamos a hacer un esfuerzo de **abstracción**. Nos quedaremos sólo con sus propiedades básicas, y veremos que puede haber muchos objetos matemáticos con las mismas propiedades, que podremos usar de la misma manera. A partir de ahora, por tanto, aunque sigamos pensando en matrices y en vectores, estudiaremos un tipo de objetos mucho más general: los elementos de un *espacio vectorial*.

Tema 2. Espacios vectoriales

2.1. Estructuras algebraicas.

En temas anteriores hemos definido matrices y vectores, estudiando algunas de sus propiedades. También hemos trabajado con cuerpos de escalares, suponiendo que se trataba de $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$, pero sin dar más detalles. Ahora vamos a estudiar con rigor estos conceptos. Definiremos algunas de las principales estructuras que se utilizan en álgebra, como son: grupos, anillos, cuerpos y espacios vectoriales. A continuación nos centraremos en la estructura que se estudia en esta asignatura: los espacios vectoriales.

Las **estructuras algebraicas** son conjuntos donde hay definidas ciertas operaciones, que satisfacen unas determinadas propiedades. Las operaciones pueden ser de varios tipos. Por ejemplo, una **operación interna**, definida en un conjunto X, es una función que a dos elementos de X (dados en orden), le hace corresponder otro elemento de X. Es decir, una función

$$p: X \times X \to X.$$

Por ejemplo, p podría ser la suma, la diferencia o la multiplicación de números reales. Observemos que, en ocasiones (la diferencia de números reales, por ejemplo) el orden en que se den los dos elementos implicados influye en el resultado.

Cuando se trabaja con una operación interna, se suele utilizar un símbolo, por ejemplo $*$, de manera que el resultado de aplicar la operación a dos elementos, a y b, se escribe $a * b$. Un ejemplo típico es el símbolo $+$ para la suma de números. En ocasiones, ni siquiera se utiliza símbolo alguno, como en el caso del producto de números, donde ab representa el producto de a y b.

La primera estructura algebraica que estudiaremos, una de las más básicas y utilizadas, es la de **grupo**:

Grupo: Sea G un conjunto no vació, y sea $*$ una operación interna definida en G. Se dice que $(G, *)$ es un **grupo**, si se cumplen las siguientes propiedades:

1. **Asociativa:** $(a * b) * c = a * (b * c), \quad \forall a, b, c \in G.$
2. **Elemento neutro:** $\exists e \in G$ tal que $a * e = e * a = a, \quad \forall a \in G.$
3. **Elemento opuesto:** $\forall a \in G, \quad \exists a' \in G$ tal que $a * a' = a' * a = e.$

Normalmente, la operación interna $*$ será la *suma* o el *producto* de elementos. En la notación *aditiva*, el elemento neutro se denota 0, y el elemento opuesto a a se denota $-a$. En la notación *multiplicativa*, el elemento neutro se denota 1, y el elemento opuesto a a, que en este caso se llama el *inverso* de a, se suele denotar a^{-1}, o bien $\frac{1}{a}$.

Sea $(G, *)$ un grupo. Se dice que G es **conmutativo** o **abeliano** si, además de las propiedades de grupo, verifica la siguiente:

4. **Propiedad conmutativa:** $\quad a * b = b * a, \quad \forall a, b \in G.$

Ejemplo 2.1 *Algunos ejemplos de grupos son los siguientes:*

- $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ *y* $(\mathbb{C}, +)$ *son grupos abelianos aditivos.*
- $(\mathbb{Q}\backslash\{0\}, \cdot)$, $(\mathbb{R}\backslash\{0\}, \cdot)$ *y* $(\mathbb{C}\backslash\{0\}, \cdot)$, *donde* $\cdot$ *se refiere al producto, son grupos abelianos multiplicativos.*
- *El conjunto de matrices* $\mathcal{M}_{m\times n}(K)$, *donde* K *es un cuerpo (ahora veremos la definición de cuerpo), junto con la suma de matrices, es un grupo abeliano aditivo.*
- *El conjunto de matrices cuadradas* **no singulares** *de* $\mathcal{M}_{n\times n}(K)$, *donde* K *es un cuerpo, junto con la multiplicación de matrices, forma un grupo que se llama* **Grupo lineal** *de orden* n *sobre* K, *y se denota* $Gl(n, K)$. *Este grupo* **no es abeliano**.
- *El conjunto de matrices cuadradas de* $\mathcal{M}_{n\times n}(K)$ **con determinante igual a 1**, *junto con la multiplicación de matrices, forma un grupo que se llama* **Grupo especial lineal** *de orden* n *sobre* K, *y se denota* $Sl(n, K)$. *Tampoco es abeliano.*
- *Los vectores de* n *coordenadas, con la suma de vectores, forman un grupo abeliano.*

En ocasiones, se define más de una operación interna sobre un conjunto. Existen estructuras que dependen de dos o más operaciones. Por ejemplo, la más sencilla es la estructura de *anillo*. Usaremos las notaciones tradicionales, $+$ y $\cdot$, para las dos operaciones internas, pero debemos recordar que pueden ser operaciones cualesquiera verificando las condiciones de la definición:

Anillo: Sea A un conjunto no vacío, y sean $+, \cdot$ dos operaciones internas, que llamaremos *suma* y *producto*, definidas en A. Se dice que $(A, +, \cdot)$ es un **anillo**, si se cumplen las siguientes propiedades:

1. $(A, +)$ es un grupo abeliano.
2. **Propiedad asociativa del producto:** $(a \cdot b) \cdot c = a \cdot (b \cdot c), \quad \forall a, b, c \in A.$
3. **Propiedad distributiva del producto respecto a la suma:**

$$\begin{cases} a \cdot (b + c) = a \cdot b + a \cdot c, & \forall a, b, c \in A, \\ \\ (a + b) \cdot c = a \cdot c + b \cdot c, & \forall a, b, c \in A. \end{cases}$$

Si se verifica alguna propiedad más, tenemos tipos especiales de anillos:

Dado un anillo $(A, +, \cdot)$, se dice que es **unitario**, o que tiene **elemento unidad**, si cumple la siguiente propiedad:

- **Elemento neutro:** $\exists u \in A$ tal que $a \cdot u = u \cdot a = a \quad \forall a \in A.$

Dado un anillo $(A, +, \cdot)$, se dice que es **conmutativo** si cumple la siguiente propiedad:

- **Propiedad conmutativa:** $a \cdot b = b \cdot a, \quad \forall a, b \in A.$

Ejemplo 2.2 *Algunos ejemplos de anillo son los siguientes:*

- $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ *y* $(\mathbb{C}, +, \cdot)$ *son anillos conmutativos.*
- *Si* $\mathbb{Z}[x]$ *es el conjunto de los polinomios en la variable* x*, con coeficientes en* $\mathbb{Z}$*, y definimos naturalmente la suma* $(+)$ *y el producto* $(\cdot)$ *de dos polinomios, entonces* $(\mathbb{Z}[x], +, \cdot)$ *es un anillo conmutativo.*
- *De igual modo,* $(\mathbb{Q}[x], +, \cdot)$, $(\mathbb{R}[x], +, \cdot)$, *y* $(\mathbb{C}[x], +, \cdot)$ *son anillos conmutativos.*

- *El conjunto $\mathcal{M}_{n\times n}(K)$, con la suma y el producto de matrices, es un anillo* **no conmutativo**.

En resumen, si $(A,+,\cdot)$ es un anillo, entonces $(A,+)$ es un grupo, y $(A,\cdot)$ es *casi* un grupo: sólo le falta el elemento inverso, y puede que el elemento unidad.

Hay elementos, como el 0 en el caso de los números, que no pueden tener inverso multiplicativo. Pero si cualquier otro elemento puede invertirse, es decir, si $(A\backslash\{0\},\cdot)$ fuera un grupo, y aún más, un grupo abeliano, entonces estaríamos ante un *cuerpo*.

Cuerpo: Sea K un conjunto no vacío, y sean $+,\cdot$ dos operaciones internas, que llamaremos *suma* y *producto*, definidas en K. Se dice que $(K,+,\cdot)$ es un **cuerpo**, si se cumplen las siguientes propiedades:

1. $(K,+)$ es un grupo abeliano.
2. $(K\backslash\{0\},\cdot)$ es un grupo abeliano, donde 0 es el elemento neutro de la suma.
3. **Propiedad distributiva del producto respecto a la suma:**

$$a\cdot(b+c)=a\cdot b+a\cdot c, \quad \forall a,b,c\in K,$$

Observemos que la propiedad distributiva sólo tiene una condición. Esto es porque el producto es conmutativo, luego la otra condición es consecuencia de la primera.

Ejemplo 2.3 *Algunos ejemplos de cuerpo son los siguientes:*

- $(\mathbb{Q},+,\cdot)$, $\quad(\mathbb{R},+,\cdot)$ *y* $(\mathbb{C},+,\cdot)$ *son cuerpos.*
- *Los grupos de matrices invertibles, $Gl(n,k)$, o de determinante 1, $Sl(n,k)$,* **no son cuerpos**, *ya que el producto de matrices no es conmutativo.*

Los cuerpos tienen multitud de propiedades, que no se estudiarán en esta asignatura. Nosotros los usaremos para definir estructuras más complejas, que generalicen las propiedades de los vectores, que hemos visto en los temas anteriores.

Para ello debemos definir las *operaciones externas*. Consideremos un conjunto X, y otro conjunto K que llamaremos *conjunto de escalares*. Llamaremos **operación externa** sobre

X, a una función que tome un elemento de K y un elemento de X, y dé como resultado un elemento de X. Es decir, una función:

$$p : K \times X \to X.$$

Normalmente, a una operación externa de este tipo la denotaremos $\cdot$ y la llamaremos *multiplicación por escalar*; y al resultado de aplicarla a un escalar $\alpha \in K$ y a un elemento $x \in X$, lo denotaremos $\alpha \cdot x$, o simplemente αx, y lo llamaremos producto de α por x.

Por tanto, si tenemos un conjunto X y otro conjunto de escalares K, podemos tener operaciones internas en cada uno de esos conjuntos, y operaciones externas entre ellos. Usando estas dos posiblidades, se definen los *espacios vectoriales*.

Espacio vectorial: Sean V y K conjuntos no vacíos. Sea $+$ una operación interna sobre V, y sea $\cdot$ una operación externa sobre V con conjunto de escalares K, que llamaremos *producto por escalar*. Diremos que V, con estas operaciones, es un **espacio vectorial** si se cumplen las siguientes propiedades:

1. $(V, +)$ es un grupo abeliano.
2. K es un cuerpo.
3. El producto por escalar verifica las siguientes propiedades:
   a) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}, \quad \forall\alpha, \beta \in K, \ \forall\mathbf{v} \in V.$
   b) $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}, \quad \forall\alpha \in K, \ \forall\mathbf{v}, \mathbf{w} \in V.$
   c) $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}, \quad \forall\alpha, \beta \in K, \ \forall\mathbf{v} \in V.$
   d) $1\mathbf{v} = \mathbf{v}, \quad \forall\mathbf{v} \in V$, donde 1 es el elemento neutro de la multiplicación de K.

A los elementos de un espacio vectorial los llamaremos **vectores**, y los denotaremos con una flecha encima. En un espacio vectorial hay, por tanto, cuatro operaciones: la suma de vectores, la suma y producto de escalares, y el producto de vectores por escalares.

Ejemplo 2.4 *Algunos ejemplos de espacios vectoriales son los siguientes:*

- *Los vectores que vimos en los temas anteriores, forman un espacio vectorial. El espacio vectorial de los vectores de n coordenadas obre un cuerpo K, se denota K^n. La suma se realiza coordenada a coordenada, y el producto por escalar también. Ejemplos de este tipo son $\mathbb{R}^2$ o $\mathbb{R}^3$.*

- *Las matrices* $\mathcal{M}_{m\times n}(K)$*, con la suma de matrices y el producto por escalar, forman un espacio vectorial. Observemos que el producto de matrices no se utiliza aquí: En general, no tiene por qué existir una multiplicación de vectores en un espacio vectorial.*

- *El* **espacio vectorial trivial** *es el conjunto* $V = \{0\}$*, con respecto a cualquier cuerpo* K*. Cualquier operación donde intervenga algún vector da como resultado el único elemento:* 0.

- *Los conjuntos de polinomios* $\mathbb{Q}[x]$*,* $\mathbb{R}[x]$ *y* $\mathbb{C}[x]$ *son espacios vectoriales con cuerpo de escalares, respectivamente,* $\mathbb{Q}$*,* $\mathbb{R}$ *y* $\mathbb{C}$*.*

- *Los conjuntos* $\mathbb{Q}[x]_{\leq n}$*,* $\mathbb{R}[x]_{\leq n}$ *y* $\mathbb{C}[x]_{\leq n}$*, formados por polinomios de grado menor o igual a* n*, son espacios vectoriales con cuerpo de escalares, respectivamente,* $\mathbb{Q}$*,* $\mathbb{R}$ *y* $\mathbb{C}$*.*

Terminamos esta sección con algunas consecuencias sencillas de la definición de espacio vectorial:

Proposición 2.5 *Si* V *es un espacio vectorial sobre un cuerpo* K*, se tienen las siguientes propiedades, para todo* $\alpha, \beta \in K$ *y todo* $\mathbf{v}, \mathbf{w} \in V$*:*

1. $\alpha\mathbf{0} = \mathbf{0}$*, donde* $\mathbf{0}$ *es el elemento neutro de la suma en* V*.*
2. $0\mathbf{v} = \mathbf{0}$*, donde* 0 *es el elemento neutro de la suma en* K*.*
3. *Si* $\alpha\mathbf{v} = \mathbf{0}$ *entonces, o bien* $\alpha = 0$ *o bien* $\mathbf{v} = \mathbf{0}$*.*
4. *Si* $\alpha\mathbf{v} = \beta\mathbf{v}$ *y* $\mathbf{v} \neq \mathbf{0}$*, entonces* $\alpha = \beta$*.*
5. *Si* $\alpha\mathbf{v} = \alpha\mathbf{w}$ *y* $\alpha \neq 0$*, entonces* $\mathbf{v} = \mathbf{w}$*.*
6. $(-\alpha)\mathbf{v} = \alpha(-\mathbf{v}) = -\alpha\mathbf{v}$*.*

2.2. Dependencia lineal.

La noción de dependencia o independencia lineal ya la hemos estudiado, en temas anteriores, para vectores de K^n. La definición es exactamente la misma para elementos de un espacio vectorial cualquiera. Repetimos aquí las definiciones y resultados principales:

Sea V un espacio vectorial sobre K. Dados r vectores $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V$, llamamos **combinación lineal** de estos vectores a cualquier expresión de la forma:

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r,$$

donde $\alpha_1, \ldots, \alpha_r \in K$.

Sea V un espacio vectorial. Diremos que un vector $\mathbf{v}$ **depende linealmente** de un conjunto de vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ si $\mathbf{v}$ se puede escribir como combinación lineal de $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

Sea V un espacio vectorial sobre K. Diremos que un sistema (o conjunto) de vectores $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subset V$ es **linealmente dependiente**, si existen r escalares $\alpha_1, \ldots, \alpha_r \in K$, *no todos nulos*, tales que

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r = \mathbf{0}.$$

En caso contrario, es decir, si la única forma de escribir el vector $\mathbf{0}$ como combinación lineal de estos vectores es tomando $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$, diremos que el sistema S es **linealmente independiente** o **libre**.

Lema 2.6 *Sea V un espacio vectorial. Un sistema de vectores $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \subset V$ es linealmente dependiente si y sólo si uno de ellos es combinación lineal de los demás.*

DEMOSTRACIÓN: Supongamos que $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es linealmente dependiente. Entonces existen escalares $\alpha_1, \ldots, \alpha_r$, no todos nulos, tales que $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r = \mathbf{0}$. Sabemos que existe al menos un $\alpha_i \neq 0$. Tendremos entonces:

$$\alpha_i \mathbf{v}_i = -\alpha_1 \mathbf{v}_1 - \cdots - \alpha_{i-1} \mathbf{v}_{i-1} - \alpha_{i+1} \mathbf{v}_{i+1} \cdots - \alpha_r \mathbf{v}_r,$$

y al ser $\alpha_i \neq 0$, podremos despejar

$$\mathbf{v}_i = -\frac{\alpha_1}{\alpha_i} \mathbf{v}_1 - \cdots - \frac{\alpha_{i-1}}{\alpha_i} \mathbf{v}_{i-1} - \frac{\alpha_{i+1}}{\alpha_i} \mathbf{v}_{i+1} \cdots - \frac{\alpha_r}{\alpha_i} \mathbf{v}_r,$$

que es una expresión de $\mathbf{v}_i$ como combinación lineal de los demás, por tanto $\mathbf{v}_i$ depende linealmente de los demás.

Supongamos ahora que un vector $\mathbf{v}_i$ depende linealmente de los demás. Esto quiere decir que existe una combinación lineal

$$\mathbf{v}_i = \beta_1\mathbf{v}_1 + \cdots + \beta_{i-1}\mathbf{v}_{i-1} + \beta_{i+1}\mathbf{v}_{i+1} \cdots + \beta_r\mathbf{v}_r.$$

De esta igualdad se obtiene

$$\beta_1\mathbf{v}_1 + \cdots + \beta_{i-1}\mathbf{v}_{i-1} - \mathbf{v}_i + \beta_{i+1}\mathbf{v}_{i+1} \cdots + \beta_r\mathbf{v}_r = \mathbf{0},$$

que es una expresión del vector $\mathbf{0}$ como combinación lineal de los vectores $\mathbf{v}_1, \ldots, \mathbf{v}_r$ donde no todos los coeficientes son nulos (el coeficiente de $\mathbf{v}_i$ es -1). Por tanto, el sistema $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es linealmente dependiente. $\square$

Lema 2.7 *Si un vector* $\mathbf{u}$ *depende linealmente de los vectores* $\mathbf{v}_1, \ldots, \mathbf{v}_p$*, y cada uno de estos depende linealmente de los vectores* $\mathbf{w}_1, \ldots, \mathbf{w}_q$*, entonces* $\mathbf{u}$ *depende linealmente de* $\mathbf{w}_1, \ldots, \mathbf{w}_q$*.*

Demostración: Por hipótesis, podemos escribir $\mathbf{u} = \alpha_1\mathbf{v}_1 + \cdots \alpha_p\mathbf{v}_p$ y además $\mathbf{v}_i = \beta_{i,1}\mathbf{w}_1 + \cdots + \beta_{i,q}\mathbf{w}_q$ para $i = 1, \ldots, p$. Sustituyendo cada $\mathbf{v}_i$ por la combinación lineal anterior, en la expresión de $\mathbf{u}$, se obtiene:

$$\mathbf{u} = \alpha_1(\beta_{1,1}\mathbf{w}_1 + \cdots + \beta_{1,q}\mathbf{w}_q) + \cdots + \alpha_p(\beta_{p,1}\mathbf{w}_1 + \cdots + \beta_{p,q}\mathbf{w}_q).$$

reorganizando los términos queda:

$$\mathbf{u} = (\alpha_1\beta_{1,1} + \cdots + \alpha_p\beta_{p,1})\mathbf{w}_1 + \cdots + (\alpha_1\beta_{1,q} + \cdots + \alpha_p\beta_{p,q})\mathbf{w}_q.$$

Si llamamos $\gamma_i = \alpha_1\beta_{1,i} + \cdots + \alpha_p\beta_{p,i}$ para $i = 1, \ldots, q$, la expresión anterior se lee

$$\mathbf{u} = \gamma_1\mathbf{w}_1 + \cdots + \gamma_q\mathbf{w}_q,$$

lo que implica que $\mathbf{u}$ depende linealmente de $\{\mathbf{w}_1, \ldots, \mathbf{w}_q\}$. $\square$

Lema 2.8 *Sea* $S \subset V$ *un sistema linealmente independiente. Si* $\mathbf{v}$ *es un vector que no depende linealmente de los vectores de* S*, entonces* $S \cup \{\mathbf{v}\}$ *es un sistema linealmente independiente.*

Demostración: Sea $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$. Por reducción al absurdo, supongamos que $S \cup \{\mathbf{v}\}$ es linealmente dependiente. Esto quiere decir que se puede escribir

$$\alpha_1\mathbf{u}_1 + \cdots \alpha_r\mathbf{u}_r + \beta\mathbf{v} = \mathbf{0},$$

donde no todos los coeficientes son nulos. Si tuviéramos $\beta = 0$, la expresión anterior sería una expresión de $\mathbf{0}$ como una combinación lineal de los elementos de S donde no todos los coeficientes serían nulos, lo cual no es posible porque S es un sistema linealmente independiente. Por tanto, $\beta \neq 0$. Podemos entonces despejar $\mathbf{v}$ en la expresión anterior, obteniendo:

$$\mathbf{v} = -\frac{\alpha_1}{\beta}\mathbf{u}_1 - \cdots - \frac{\alpha_r}{\beta}\mathbf{u}_r.$$

Por tanto $\mathbf{v}$ depende linealmente de S. Contradicción. $\square$

2.3. Sistemas de generadores y bases.

En esta sección veremos cómo el concepto de dependencia lineal sirve para expresar los elementos de un espacio vectorial utilizando sólo un conjunto (posiblemente finito) de vectores.

Sistema de generadores: Sea V un espacio vectorial. Diremos que un sistema de vectores $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ es un **sistema de generadores** de V si todo vector de V puede escribirse como combinación lineal de los vectores de S.

En este caso diremos que V está generado por S, o por los vectores de S.

Un espacio vectorial puede tener muchos sistemas de generadores diferentes. Incluso puede haber sistemas de generadores donde "sobre" algún vector. Por ejemplo, si tenemos un sistema con cuatro vectores en $\mathbb{R}^3$, nos basta con tres de ellos para generar todo el espacio. Esto nos va a llevar al concepto de base. Pero antes debemos hacer una restricción, puesto que existen espacios vectoriales demasiado "grandes".

Un espacio vectorial V se dice que es de **tipo finito** si está generado por un número finito de vectores. Es decir, si existe un sistema de generadores $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

Para estos espacios vectoriales de tipo finito, podemos definir sin problemas la noción de *base*:

Base: Sea V un espacio vectorial de tipo finito. Diremos que un sistema de vectores $B \subset V$ es una **base** de V si cumple:

1. B es un sistema de generadores de V.

2. B es linealmente independiente.

En otras palabras, una base es un sistema de generadores de un espacio vectorial en el que no sobra ningún vector, ya que, al ser linealmente independiente, ninguno de ellos puede escribirse como combinación lineal de los demás.

Teorema 2.9 *Sea V un espacio vectorial de tipo finito, y sea B sistema de vectores de V. Entonces B es una base si y sólo si todo vector de V se puede expresar* **de una única manera** *como combinación lineal de los vectores de B.*

Demostración: Supongamos que $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base de V. Dado un vector $\mathbf{v} \in V$, como B es sistema de generadores podremos escribir $\mathbf{v} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. Si existiera otra forma de expresar $\mathbf{v}$, digamos $\mathbf{v} = \beta_1\mathbf{u}_1 + \cdots + \beta_n\mathbf{u}_n$, entonces tendríamos

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (\alpha_1 - \beta_1)\mathbf{u}_1 + \cdots (\alpha_n - \beta_n)\mathbf{u}_n.$$

Pero como B es un sistema linealmente independiente, los coeficientes de la expresión anterior deben ser todos nulos. Es decir, $\alpha_i - \beta_i = 0$, o lo que es lo mismo, $\alpha_i = \beta_i$ para todo $i = 1, \ldots, n$. Por tanto, la forma de expresar $\mathbf{v}$ como combinación lineal de los elementos de B es única.

Recíprocamente, sea $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es un sistema tal que todo vector $\mathbf{v} \in V$ se puede expresar de forma única como combinación lineal de los vectores de B. Por un lado, B es sistema de generadores, puesto que todo vector de V se puede expresar como combinación lineal de B. Por otra parte, consideremos el vector $\mathbf{0} \in V$. Sabemos que siempre se tiene la combinación lineal obvia:

$$\mathbf{0} = 0\mathbf{u}_1 + \cdots + 0\mathbf{u}_n.$$

Por la propiedad que le suponemos a B, esta es la única forma de escribir $\mathbf{0}$ como combinación lineal de los vectores de B. Por tanto, B es un sistema linealmente independiente, luego es una base. $\square$

Ahora veamos que un espacio vectorial de tipo finito, que no sea trivial, siempre tiene una base. Además veremos cómo se construye, a partir de un sistema de generadores.

Teorema 2.10 (de existencia de base) *Sea $V \neq \{\mathbf{0}\}$ un espacio vectorial de tipo finito. Dado cualquier sistema finito de generadores $G \subset V$, existe una base B de V formada por vectores de G.*

Demostración: Consideremos el sistema de generadores $G = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. Si es libre, entonces es una base, y hemos acabado. Si no, hay un elemento $\mathbf{v}_i \in G$ que depende linealmente de los demás. Pero entonces $G_1 = G \backslash \{\mathbf{v}_i\}$ sigue siendo sistema de generadores. Si es libre, G_1 es una base. Si no, existirá otro vector $\mathbf{v}_j$ que depende linealmente de los demás vectores de G_1, y también lo podremos eliminar.

Continuamos este proceso mientras el sistema de generadores sea linealmente dependiente. Pero como mucho podremos eliminar $p-1$ vectores ya que, como $V \neq \{\mathbf{0}\}$, al menos debe haber un vector en cualquier sistema de generadores. Por tanto, en algún momento debemos tener algún G_i que sea libre, luego será una base contenida en G. $\square$

2.4. Teorema de la base. Dimensión.

en esta sección definiremos un concepto esencial del álgebra lineal: la *dimensión* de un espacio vectorial. Necesitamos primero el siguiente resultado:

Teorema 2.11 *Sea V un espacio vectorial. Si $G = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ es un sistema de generadores de V, y $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es un sistema linealmente independiente, entonces $n \leq m$.*

Demostración: Supongamos que $n > m$. Como G es un sistema de generadores, podemos escribir cada $\mathbf{v}_i$ como combinación lineal de los elementos de G:

$$\mathbf{v}_i = a_{1i}\mathbf{u}_1 + \cdots + a_{mi}\mathbf{u}_m.$$

Por otra parte, como S es linealmente independiente, la ecuación

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = 0$$

sólo puede admitir la solución trivial, $x_1 = \cdots = x_n = 0$. Ahora bien, sustituyendo cada $\mathbf{v}_i$, obtenemos la ecuación equivalente:

$$x_1(a_{11}\mathbf{u}_1 + \cdots + a_{m1}\mathbf{u}_m) + \cdots + x_n(a_{1n}\mathbf{u}_1 + \cdots + a_{mn}\mathbf{u}_m) = 0,$$

donde, sacando factor común los $\mathbf{u}_i$, se tiene:

$$(a_{11}x_1 + \cdots + a_{1n}x_n)\mathbf{u}_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)\mathbf{u}_m = 0.$$

Una posible solución para esta ecuación se obtendría si cada coeficiente fuera cero, es decir, si

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

Este sistema homogéneo tiene, como máximo, rango m, ya que tiene m filas. Ahora bien, si $n > m$, el Teorema de Rouché-Frobenius nos dice que es un sistema compatible indeterminado, es decir, existe una solución para $x_1, \ldots, x_n$ donde no todos son cero. Esto contradice que S sea un sistema libre. $\square$

Veamos entonces qué es la *dimensión* de un espacio vectorial:

Teorema 2.12 (Teorema de la base) *Sea V un espacio vectorial de tipo finito. Todas las bases de V tienen el mismo número de elementos. A este número se le llama* **dimensión** *de V.*

Demostración: Sean B_1 y B_2 dos bases de V, de m y n vectores respectivamente. Como B_1 es sistema de generadores, y B_2 es libre, entonces $n \leq m$ por el Teorema 2.11. Pero como B_2 es sistema de generadores, B_1 es libre, se tiene $m \leq n$. Por tanto, $m = n$. $\square$

Dimensión: La dimensión de un espacio vectorial V, que denotamos $\dim(V)$, se define como sigue:

- Si $V = \{\mathbf{0}\}$, entonces $\dim(V) = 0$.
- Si V es de tipo finito, su dimensión es el número de elementos de cualquier base de V.
- Si V no es de tipo finito, diremos que tiene dimensión infinita, y escribiremos $\dim V = \infty$.

Ejemplo 2.13 *El espacio vectorial $\mathbb{R}^n$ tiene dimensión n. Una base, llamada la* **base canónica**, *está formada por los vectores $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, donde*

$$\mathbf{e}_i = (0, \ldots, 0, \overset{(i)}{1}, 0, \ldots, 0).$$

Ejemplo 2.14 *El conjunto de polinomios, $\mathbb{R}[x]$, es un espacio vectorial de dimensión infinita. En efecto, supongamos que existe un sistema de generadores G de $\mathbb{R}[x]$, formado por un número finito de polinomios. Sea entonces m el mayor grado de todos los polinomios de G. Entonces, cualquier combinación lineal de los polinomios de G tiene como máximo grado m, luego no podríamos obtener los polinomios de grado mayor que m, y G no sería sistema de generadores. Por tanto,* $\dim(\mathbb{R}[x]) = \infty$.

2.5. Dimensión y sistemas de vectores. Coordenadas.

La dimensión de un espacio vectorial nos impone restricciones sobre el tamaño que pueden tener los sistemas libres, o los sistemas de generadores:

Proposición 2.15 *Sea $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ un sistema de vectores de un espacio vectorial V de dimensión finita. Se tiene:*

1. *Si S es un sistema de generadores, entonces $m \geq \dim V$.*
2. *Si S es linealmente independiente, entonces $m \leq \dim V$.*
3. *Si S es sistema de generadores, y $m = \dim V$, entonces S es base de V.*
4. *Si S es linealmente independiente, y $m = \dim V$, entonces S es base de V.*

Demostración: Es consecuencia directa del Teorema 2.11, y del teorema de existencia de base. □

Una propiedad importante de las bases es la siguiente:

Teorema 2.16 (Teorema de la base incompleta) *Sea V un espacio vectorial de tipo finito. Todo sistema linealmente independiente puede completarse hasta obtener una base. Es decir, si $\dim V = n$, y $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ es un sistema libre, con $m < n$, entonces existen $n - m$ vectores $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n \in V$ tales que el sistema $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es base de V. Además, los vectores $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ pueden tomarse de cualquier base de V.*

Demostración: Sea S como en el enunciado, y sea $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ una base de V. Si cada elemento de B depende linealmente de los elementos de S, entonces S es sistema de generadores, luego sería una base. Imposible. Tendremos entonces un vector de B, supongamos que es $\mathbf{u}_1$, que no depende linealmente de S. Tomamos entonces el sistema $S \cup \{\mathbf{u}_1\}$, que será linealmente independiente.

Si $m+1 < n$, entonces $S \cup \{\mathbf{u}_1\}$ no es base. Por tanto, debe existir otro vector en B (que no puede ser $\mathbf{u}_1$), que no dependa linealmente de $S \cup \{\mathbf{u}_1\}$. Digamos que es $\mathbf{u}_2$. Entonces $S \cup \{\mathbf{u}_1, \mathbf{u}_2\}$ es linealmente independiente. Continuamos este proceso hasta obtener $S \cup \{\mathbf{u}_1, \ldots, \mathbf{u}_{n-m}\}$, sistema linealmente independiente de n vectores, es decir, base de V. $\square$

La principal ventaja de la existencia de bases, en los espacios vectoriales de tipo finito, es que vamos a poder estudiarlos, sea cual sea el espacio vectorial, como si fuera K^n. Esto lo vamos a conseguir mediante el uso de *coordenadas*.

Primero necesitamos hacer una precisión. Hasta ahora, cuando hablábamos de un sistema de vectores, o de una base, no nos importaba el orden en que estuvieran los vectores. Pero para definir las coordenadas de un vector, es necesario fijar un orden. Por tanto, a partir de ahora, escribiremos la bases de la siguiente forma: $B = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$. El uso de paréntesis, en lugar de llaves, indica que los vectores están ordenados, luego podremos hablar del i-ésimo vector de una base, de forma rigurosa.

Coordenadas: Sea V un espacio vectorial de dimensión n sobre un cuerpo K. Dada una base $B = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ sabemos que, para todo vector $\mathbf{v} \in V$, existe una única combinación lineal

$$\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n.$$

Los escalares $\alpha_1, \ldots, \alpha_n$ definen, por tanto, al vector $\mathbf{v}$, y los llamaremos **coordenadas** de $\mathbf{v}$ respecto a B. Escribiremos:

$$\mathbf{v}_B = (\alpha_1, \ldots, \alpha_n).$$

Cuando la base B esté clara por el contexto, escribiremos simplemente

$$\mathbf{v} = (\alpha_1, \ldots, \alpha_n).$$

Por tanto, no importa cómo sea V como espacio vectorial; si fijamos una base, vamos a poder representar los elementos de V como elementos del conocido espacio vectorial K^n. Pero la correspondencia entre V y K^n es todavía más fuerte: las operaciones de suma y producto por escalar son iguales en ambos espacios. Veamos esto con más detalle:

Teorema 2.17 *Sea V un espacio vectorial de dimensión n sobre un cuerpo K, y sea B*

una base de V. Sea

$$\mathcal{C}_B : V \to K^n$$

la aplicación que a cada elemento de V le hace corresponder el vector de sus coordenadas. Entonces $\mathcal{C}_B$ es una aplicación biyectiva, y además se tiene:

1. $\mathcal{C}_B(\mathbf{u}+\mathbf{v}) = \mathcal{C}_B(\mathbf{u}) + \mathcal{C}_B(\mathbf{v}) \qquad \forall \mathbf{u}, \mathbf{v} \in V.$
2. $\mathcal{C}_B(\alpha\, \mathbf{u}) = \alpha\, \mathcal{C}_B(\mathbf{u}) \qquad \forall \mathbf{u} \in V,\ \forall \alpha \in K.$

Demostración: La aplicación es biyectiva por el Teorema 2.9. Las propiedades de la suma y del producto por escalar se prueban de forma directa. □

Este resultado nos dice que los espacios vectoriales V y K^n son **isomorfos**. Por tanto, si necesitamos trabajar con un espacio vectorial de dimensión n, podemos trabajar simplemente con K^n.

2.6. Cambio de base.

Observemos que las coordenadas de un vector de V dependen de la base B que hayamos elegido. Si tuviéramos otra base B', las coordenadas del mismo vector serían diferentes. Vamos a ver entonces cómo están relacionados estos dos tipos de coordenadas.

Supongamos que tenemos un espacio vectorial V de dimensión n, y sean $B = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ y $B' = (\mathbf{u}'_1, \ldots, \mathbf{u}'_n)$ dos bases de V. Como B es base, podremos escribir cada vector de B' respecto a B, es decir, tendremos:

$$\begin{aligned}
\mathbf{u}'_1 &= a_{11}\mathbf{u}_1 + a_{21}\mathbf{u}_2 + \cdots + a_{n1}\mathbf{u}_n, \\
\mathbf{u}'_2 &= a_{12}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + \cdots + a_{n2}\mathbf{u}_n, \\
&\vdots \\
\mathbf{u}'_n &= a_{1n}\mathbf{u}_1 + a_{2n}\mathbf{u}_2 + \cdots + a_{nn}\mathbf{u}_n.
\end{aligned}$$

Con esta notación, se tiene lo siguiente:

Teorema 2.18 *Si las coordenadas de $\mathbf{v} \in V$ respecto a B y B' son, respectivamente $\mathbf{v}_B = (x_1, \ldots, x_n)$ y $\mathbf{v}_{B'} = (x'_1, \ldots, x'_n)$, entonces se tiene la relación:*

$$
\begin{aligned}
\mathbf{x}_1 &= a_{11}\mathbf{x}'_1 + a_{12}\mathbf{x}'_2 + \cdots + a_{1n}\mathbf{x}'_n,\\
\mathbf{x}_2 &= a_{21}\mathbf{x}'_1 + a_{22}\mathbf{x}'_2 + \cdots + a_{2n}\mathbf{x}'_n,\\
&\vdots\\
\mathbf{x}_n &= a_{n1}\mathbf{x}'_1 + a_{n2}\mathbf{x}'_2 + \cdots + a_{nn}\mathbf{x}'_n.
\end{aligned}
$$

DEMOSTRACIÓN: Por un lado, tenemos $\mathbf{v} = x_1\mathbf{u}_1 + \cdots x_n\mathbf{u}_n$. Y por otro lado, $\mathbf{v} = x'_1\mathbf{u}'_1 + \cdots x'_n\mathbf{u}'_n$. Si sustituimos cada $\mathbf{u}'_i$ por $a_{1i}\mathbf{u}_1 + a_{2i}\mathbf{u}_2 + \cdots + a_{ni}\mathbf{u}_n$ en la expresión anterior, y agrupamos coeficientes, obtendremos:

$$\mathbf{v} = (a_{11}x'_1 + \cdots + a_{1n}x'_n)\mathbf{u}_1 + \cdots + (a_{n1}x'_1 + \cdots + a_{nn}x'_n)\mathbf{u}_1.$$

Como la forma de expresar $\mathbf{v}$ como combinación lineal de B es única, los coeficientes de esta última combinación lineal han de ser iguales a $x_1, \ldots, x_n$, lo que demuestra el resultado. $\square$

Una de las principales ventajas de trabajar con K^n es que podemos usar matrices. El teorema anterior, por ejemplo, se puede ver mucho mejor de forma matricial. Sea

$$A_{B',B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

la *matriz del cambio de base*. Es decir, las **columna** i de $A_{B',B}$ contiene las coordenadas del vector $\mathbf{v}'_i$ de B' respecto de la base B. Entonces la relación entre las coordenadas $(x_1, \ldots, x_n)$ y $(x'_1, \ldots, x'_n)$, respecto a B y B', de un vector cualquiera es:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}.$$

Escrito de otra manera,

$$X = A_{B',B}\, X',$$

donde X y X' son los vectores columna que representan las coordenadas de un vector respecto a B y a B'. Por tanto, la matriz $A_{B',B}$ transforma las coordenadas respecto a B' en coordenadas respecto a B (mediante multiplicación a izquierda).

Teorema 2.19 *Sea V un espacio vectorial de dimensión n, y sea B una base de V. Dado un sistema B' de n vectores, sea $A_{B',B} \in \mathcal{M}_{n\times n}(K)$ la matriz cuyas columnas contienen las coordenadas de los vectores de B' respecto a B. Entonces B' es una base si y sólo si $A_{B',B}$ es no singular.*

Demostración: B' es base de V si y sólo si sus vectores son linealmente independientes. Esto es, si y sólo si las columnas de $A_{B',B}$ son linealmente independientes, lo que ocurre si y sólo si $\mathrm{rg}(A_{B',B}) = n$, es decir, si $A_{B',B}$ es no singular. $\square$

Otra forma de demostrar que, dadas dos bases B y B', la matriz $A_{B',B}$ es invertible, es la siguiente: consideremos la matriz $A_{B,B'}$. Esta matriz transforma coordenadas respecto de B en coordenadas respecto de B'. Por tanto, tenemos por un lado $X = A_{B',B}\, X'$, y por otro $X' = A_{B,B'}\, X$. Uniendo estas dos igualdades, se tiene:

$$X = A_{B',B}\, X' = (A_{B',B}A_{B,B'})X.$$

Como esta igualdad se tiene para cualquier vector $X \in K^n$, deducimos que $A_{B',B}A_{B,B'} = I$. Análogamente, se obtiene $A_{B,B'}A_{B',B} = I$. Por tanto:

Dadas dos bases B y B' de un espacio vectorial de dimensión n, la matriz de cambio de base $A_{B',B}$ es invertible, y su inversa es $A_{B,B'}$.

Usando este tipo de matrices, podremos ver la similitud existente entre los conceptos definidos para espacios vectoriales y los definidos para matrices. Pero esto lo haremos mejor en el tema siguiente, donde definimos las variedades lineales.

Tema 3. Variedades lineales

3.1. Definición y propiedades básicas.

En los ejemplos que hemos dado en $\mathbb{R}^3$, vimos que un vector define una recta, o que dos vectores (no proporcionales) definen un plano. Son estas estructuras las que en realidad nos interesan, y en las que se centra la mayor parte del álgebra lineal. En esta sección veremos cómo estas estructuras, llamadas *variedades lineales* o *subespacios vectoriales*, también son espacios vectoriales, y estudiaremos sus propiedades. La definición precisa es la siguiente:

Variedad lineal: Sea V un espacio vectorial sobre un cuerpo K, y sea L un subconjunto de V. Diremos que L es un **subespacio vectorial**, o una **variedad lineal** de V si, con las mismas operaciones de suma y producto por escalar, L es un espacio vectorial sobre K.

Observemos que los elementos de L, al ser elementos de V, satisfacen todas las propiedades de un espacio vectorial. Pero hay un detalle importante: tanto la suma de vectores de L, como el producto por escalares, deben dar como resultado vectores de L. Si no, no estaríamos ante operaciones en L, y por tanto L no sería espacio vectorial. Por tanto, lo único que hay que verificar para saber si $L \subset V$ es subespacio vectorial, es lo siguiente:

Proposición 3.1 *Dado un espacio vectorial V sobre un cuerpo K, un subconjunto $L \subset V$ es una variedad lineal de V si se cumplen las siguientes propiedades:*

1. $\forall \mathbf{v}, \mathbf{w} \in L, \qquad \mathbf{v} + \mathbf{w} \in L.$
2. $\forall \alpha \in K, \quad \forall \mathbf{v} \in L, \qquad \alpha\mathbf{v} \in L.$

La siguiente propiedad es consecuencia directa de la definición.

Proposición 3.2 *Sea V un espacio vectorial sobre un cuerpo K, y sea $\mathbf{0}$ el elemento neutro de la suma de vectores. Se tiene:*

1. *El espacio vectorial trivial $\{\mathbf{0}\}$ es una variedad lineal de V.*
2. *Cualquier variedad lineal $L \subset V$ contiene al vector $\mathbf{0}$.*

El ejemplo principal de espacio vectorial que vamos a utilizar es $\mathbb{R}^n$. Recordemos que, si tenemos un sólo vector $\mathbf{v} \in \mathbb{R}^3$, los vectores que se pueden escribir como combinación lineal de $\mathbf{v}$ forman una recta: la que pasa por el origen y tiene la dirección de $\mathbf{v}$. Por otra parte, si tenemos dos vectores $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, los vectores que se pueden escribir como combinación lineal de $\mathbf{v}$ y $\mathbf{w}$ forman un plano: el que pasa por el origen y contiene a la recta definida por $\mathbf{v}$ y a la recta definida por $\mathbf{w}$. Al estudiar sistemas de vectores, lo que de verdad nos interesa es esa recta o ese plano, es decir, el conjunto de todos los vectores que se pueden escribir como combinación lineal de los vectores del sistema:

Teorema 3.3 *Sea V un espacio vectorial, y sea S un sistema de vectores de V. El conjunto de combinaciones lineales de los vectores de S, que llamaremos $\langle S \rangle$, es una variedad lineal de V.*

Demostración: Directa. $\square$

Sean $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ y $T = \{\mathbf{w}_1, \cdots \mathbf{w}_q\}$ dos sistemas de vectores de un espacio vectorial V. Diremos que S y T son **equivalentes** si $\langle S \rangle = \langle T \rangle$.

Otra posible definición de equivalencia de sistemas es la que viene dada por el siguiente resultado:

Proposición 3.4 *Sea V un espacio vectorial. Dos sistemas de vectores $S, T \in V$ son equivalentes si y sólo si todo vector de S puede escribirse como combinación lineal de los vectores de T, y viceversa.*

Demostración: Directa. $\square$

En el caso de $V = \mathbb{R}^3$, dos sistemas de dos vectores son equivalentes si y sólo si definen el mismo plano. De hecho, en $\mathbb{R}^3$, las variedades lineales son la siguientes: el origen (que corresponde al subespacio trivial $\{\mathbf{0}\}$), las rectas que pasan por el origen, los planos que pasan por el origen, y todo $\mathbb{R}^3$. Esto nos da una idea de las dimensiones de estas variedades lineales: en $\mathbb{R}^3$, que tiene dimensión 3, existen variedades de dimensión 0, 1, 2 o 3. Más generalmente, se tiene:

Teorema 3.5 *Sea V un espacio vectorial de dimensión finita, y sea L una variedad lineal de V. Entonces L también tiene dimensión finita, y $\dim L \leq \dim V$. Además, la igualdad sólo se da si $L = V$.*

Demostración: Si $L = \{\mathbf{0}\}$, el resultado es evidente. Supongamos entonces que L contiene vectores no nulos. Entonces L contiene sistemas libres. Pero cualquier sistema libre de L es también un sistema libre de V, luego tiene como máximo n vectores, donde $n = \dim V$. Sea m el número máximo de vectores que puede tener un sistema libre de L (ya sabemos que $m \leq n$), y sea B un sistema libre de m vectores de L. Como no existe otro sistema libre de L con más de m vectores, entonces todo vector de L depende linealmente de B, es decir, B es una base de L. Por tanto $\dim L = m \leq n = \dim V$.

Si tuviéramos $\dim L = \dim V$, entonces una base B de L sería un sistema libre de V con n elementos, luego sería base de V. Por tanto, $L = V$. $\square$

Rango de un sistema de vectores: Sea V un espacio vectorial, y sea S un sistema finito de vectores de V. Llamamos **rango** de S a la dimensión de la variedad lineal generada por S. Es decir:

$$\mathrm{rg}(S) = \dim(\langle S \rangle).$$

Dicho de otra forma, el rango de S es el mayor número de vectores linealmente independientes que se pueden tomar en S.

Tenemos entonces el siguiente resultado, que relaciona el rango de un sistema de vectores y el rango de una matriz:

Proposición 3.6 *En un espacio vectorial V de dimensión finita, sea B una base de V, S un sistema finito de vectores de V, y $A_{S,B}$ la matriz cuyas columnas (o cuyas filas) son las coordenadas de los vectores de S respecto a B. Entonces*

$$rg(S) = rg(A_{S,B}).$$

Demostración: Si $V = K^n$, ya hemos demostrado que el rango de una matriz es el máximo número de columnas (o filas) linealmente independientes que tiene. Si $V \neq K^n$, el resultado es consecuencia del isomorfismo $\mathcal{C}_B$, que a cada vector de V le asocia sus coordenadas. $\square$

Nota: Observemos que el rango de la matriz $A_{S,B}$ **no depende** de la base B, ya que es igual al rango del sistema de vectores S, que está definido sin tener que recurrir a ninguna base. Otra forma de ver esto es la siguiente: si B y B' son dos bases distintas, cuya matriz de cambio de base es $A_{B',B}$, entonces se tiene:

$$A_{S,B} = A_{B',B}\, A_{S,B'}.$$

Como sabemos que $A_{B',B}$ es no singular, entonces $\mathrm{rg}(A_{S,B}) = \mathrm{rg}(A_{S,B'})$.

3.2. Ecuaciones paramétricas e implícitas.

Volviendo a nuestro ejemplo principal, $\mathbb{R}^3$, el rango de un sistema de vectores $S \subset \mathbb{R}^3$, nos dice si $\langle S \rangle$ es un punto, una recta, un plano o todo el espacio, según sea 0, 1, 2 ó 3. Y para ver cuál es ese rango, basta calcular el rango de la matriz cuyas columnas son los vectores de S.

Hasta ahora sólo hemos visto una forma de determinar una variedad lineal: mediante un sistema de generadores. Esta forma es equivalente a dar unas *ecuaciones paramétricas.*

Ecuaciones paramétricas de una variedad lineal: Sea L una variedad lineal de un espacio vectorial V de dimensión n, y sea $G = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ un sistema de generadores de L. Supongamos que las coordenadas de $\mathbf{v}_i$ respecto a una base B de V son: $\mathbf{v}_i = (a_{1i}, \cdots, a_{ni})$. Entonces, como todo vector $\mathbf{v} \in L$, con coordenadas $(x_1, \ldots, x_n)$ se escribe como combinación lineal de G, existirán unos escalares $\lambda_1, \ldots, \lambda_m$ tales que $\mathbf{v} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_m \mathbf{v}_m$. Es decir:

$$\left\{\begin{array}{lclclcl} x_1 & = & a_{11}\lambda_1 & + & \cdots & + & a_{1m}\lambda_m \\ x_2 & = & a_{21}\lambda_1 & + & \cdots & + & a_{2m}\lambda_m \\ \vdots & & \vdots & & & & \vdots \\ x_n & = & a_{n1}\lambda_1 & + & \cdots & + & a_{nm}\lambda_m \end{array}\right.$$

Unas ecuaciones de este tipo, conde los escalares λ_i son parámetros indeterminados, se llaman unas **ecuaciones paramétricas** de L.

En otras palabras, unas ecuaciones paramétricas nos dicen cómo son las coordenadas de un vector cualquiera de L, dependiendo de los coeficientes que tomemos en la combinación lineal de los generadores.

> **Ejemplo 3.7** *Un plano en $\mathbb{R}^3$ que pasa por el origen (es decir, una variedad lineal de $\mathbb{R}^3$ de dimensión 2), puede venir dada por las siguientes ecuaciones paramétricas:*
>
> $$\left\{\begin{array}{l} x_1 = 2\lambda_1 - 3\lambda_2 \\ x_2 = \lambda_1 + 5\lambda_2. \end{array}\right.$$
>
> *En este caso se trata del plano generado por los vectores $(2,1)$ y $(-3,5)$.*

Las ecuaciones paramétricas, en el fondo, equivalen a definir una variedad lineal dando un sistema de generadores. Pero existe otra forma, más interesante, de determinar una variedad lineal: mediante unas *ecuaciones implícitas*. El resultado que necesitamos es el siguiente:

Teorema 3.8 *Sea V un espacio vectorial de dimensión n, y sea B una base de V. Consideremos un sistema lineal homogéneo:*

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

Sea L el conjunto de vectores cuyas coordenadas (respecto de B) son una solución de este sistema lineal. Entonces L es una variedad lineal.

Demostración: Si dos vectores $\mathbf{v} = (x_1, \ldots, x_n)$ y $\mathbf{v}' = (x_1', \ldots, x_n')$ pertenecen a L, entonces satisfacen cada ecuación del sistema, es decir, $a_{i1}x_1 + \cdots + a_{in}x_n = 0$ y además $a_{i1}x_1' + \cdots + a_{in}x_n' = 0$. Pero entonces,

$$(a_{i1}x_1 + \cdots + a_{in}x_n) + (a_{i1}x_1' + \cdots + a_{in}x_n') = a_{i1}(x_1 + x_1') + \cdots a_{in}(x_n + x_n') = 0,$$

por tanto, el vector $\mathbf{v} + \mathbf{v}' = (x_1 + x_1', \ldots, x_n + x_n')$ es solución del sistema, luego pertenece a L.

Por otra parte, dado cualquier $\alpha \in K$, se tiene

$$\alpha(a_{i1}x_1 + \cdots + a_{in}x_n) = a_{i1}(\alpha x_1) + \cdots + a_{in}(\alpha x_n) = 0,$$

luego $\alpha\mathbf{v} = (\alpha x_1, \ldots, \alpha x_n) \in L$. Por tanto, L es una variedad lineal. $\square$

Ecuaciones implícitas de una variedad lineal: Sea V un espacio vectorial de dimensión n, y sea B una base de V. Unas **ecuaciones implícitas** de una variedad lineal L es un sistema de ecuaciones

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

tal que los vectores de L sean exactamente aquellos cuyas coordenadas (respecto a B) son una solución del sistema.

En otras palabras, si unas ecuaciones paramétricas nos dicen *cómo son* las coordenadas de los vectores de L, unas ecuaciones implícitas nos dicen *qué relaciones deben verificar* entre ellas. Podríamos decir que en unas ecuaciones implícitas los vectores de L están más escondidos, ya que a simple vista no podríamos determinar ninguno de ellos: hay que resolver el sistema.

Observemos que el teorema anterior nos ha dado una nueva motivación para estudiar variedades lineales, ya que las soluciones de un sistema lineal homogéneo *son* variedades lineales.

3.3. Ecuaciones y dimensión.

Veamos ahora cómo, a partir de unas ecuaciones paramétricas o implícitas de una variedad lineal, podemos calcular la dimensión de la variedad.

Proposición 3.9 *Sea V un espacio vectorial de dimensión n, sean*

$$\left\{\begin{array}{rcl} x_1 & = & a_{11}\lambda_1 + \cdots + a_{1m}\lambda_m \\ x_2 & = & a_{21}\lambda_1 + \cdots + a_{2m}\lambda_m \\ \vdots & & \vdots \\ x_n & = & a_{n1}\lambda_1 + \cdots + a_{nm}\lambda_m \end{array}\right.$$

unas ecuaciones paramétricas de una variedad lineal L, y sea

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

la matriz de los coeficientes. Entonces

$$\dim L = rg(A).$$

Demostración: Esto es una consecuencia inmediata de los teoremas que conocemos sobre la base de una variedad lineal, sabiendo que las columnas de A son generadores de L. $\square$

Proposición 3.10 *Sea V un espacio vectorial de dimensión n, y sean*

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

unas ecuaciones implícitas de una variedad lineal L, y sea A la matriz de coeficientes del sistema homogéneo. Entonces:

$$\dim L = n - rg(A).$$

Demostración: Recordemos cómo se usa el método de eliminación de Gauss-Jordan para resolver un sistema lineal. Si la matriz A tiene rango r, obtendremos r variables pivote. Por simplificar, diremos que las variables pivote son $x_1, \ldots, x_r$, aunque la demostración funciona igual si son otras. Despejando las variables pivote respecto a las demás, se obtiene que la solución general del sistema es de la forma:

$$\begin{array}{c} x_1 = c_{1r+1}x_{r+1} + c_{1r+2}x_{r+2} + \cdots + c_{1n}x_n, \\ x_2 = c_{2r+1}x_{r+1} + c_{2r+2}x_{r+2} + \cdots + c_{2n}x_n, \\ \vdots \\ x_r = c_{rr+1}x_{r+1} + c_{rr+2}x_{r+2} + \cdots + c_{rn}x_n, \end{array}$$

donde las variables no pivote $x_{r+1}, \ldots, x_n$ pueden tomar cualquier valor. Pero si le damos a las variables $x_{r+1}, \ldots, x_n$ los valores (indeterminados) $\lambda_1, \ldots, \lambda_{n-r}$, se obtiene que la solución general del sistema (es decir, la variedad L) viene dada por:

$$\left\{\begin{array}{cccccccc} x_1 & = c_{1r+1}\lambda_1 & + & c_{1r+2}\lambda_2 & + & \cdots & + & c_{1n}\lambda_{n-r} \\ x_2 & = c_{2r+1}\lambda_1 & + & c_{2r+2}\lambda_2 & + & \cdots & + & c_{2n}\lambda_{n-r} \\ \vdots & \vdots & & & & \vdots & & \\ x_r & = c_{rr+1}\lambda_1 & + & c_{rr+2}\lambda_2 & + & \cdots & + & c_{rn}\lambda_{n-r} \\ x_{r+1} & = \quad \lambda_1 & & & & & & \\ x_{r+2} & = & & \lambda_2 & & & & \\ \vdots & & & & & \ddots & & \\ x_n & = & & & & & & \lambda_{n-r}. \end{array}\right.$$

Pero estas resultan ser unas ecuaciones paramétricas de la variedad L, donde la matriz de coeficientes tiene rango $n-r$, ya que tiene $n-r$ columnas, y sus $n-r$ últimas filas son claramente libres. Luego, por el resultado anterior, se sigue que $\dim L = n - r$. Es decir, $\dim L = n - \mathrm{rg}(A)$. $\square$

En muchas ocasiones es importante saber, dada una variedad lineal L, transformar unas ecuaciones implícitas en unas ecuaciones paramétricas, y viceversa. El primer caso es sencillo:

Observación: Si tenemos unas ecuaciones implícitas de una variedad lineal L, es decir, un sistema homogéneo que determina los elementos de L, la forma de calcular unas ecuaciones paramétricas es simplemente **resolviendo el sistema**, como hemos visto en el resultado anterior.

Observemos además que la solución obtenida despejando las variables pivote nos da una **base** de L, formada por los vectores que son coeficientes de los parámetros λ_1, λ_2, etc.

Si por el contrario tenemos unas ecuaciones paramétricas de L, es decir, un sistema de generadores, la forma de calcular unas ecuaciones implícitas es la siguiente:

Método para transformar un sistema de generadores S de una variedad L en unas ecuaciones implícitas. Suponemos fijada una base B del espacio vectorial V, y procedemos de la siguiente manera.

1. Se considera la matriz $A_{S,B}$, cuyas filas son las coordenadas de los elementos de S respecto de la base B.
2. Mediante el método del orlado, se identifican el máximo número posible de filas independientes, con lo que se obtiene una base B_1 de la variedad L. Digamos que B_1 tiene r elementos, es decir, $\dim(L) = r$.
3. Se considera la matriz $A_{B_1,B} \in \mathcal{M}_{n\times r}$, cuyas filas son una base de L, y la matriz M que resulta al añadir a esta matriz una fila de incógnitas $(x_1, \ldots, x_n)$.
4. Un vector $(x_1, \ldots, x_n)$ estará en L si y sólo si es combinación lineal de las filas de $A_{B_1,B}$. Es decir, si y sólo si la matriz M tiene rango r. Imponemos entonces que la matriz M tenga rango r. Usando el método del orlado (orlando un menor no nulo de tamaño r en las r primeras filas de M), esto significa imponer que $n-r$ determinantes sean nulos. Estos $n-r$ determinantes son $n-r$ ecuaciones implícitas que definen L.

Este proceso nos sirve además para demostrara el siguiente resultado:

Teorema 3.11 *Toda variedad lineal, L, de un espacio vectorial de dimensión finita, puede ser representada por unas ecuaciones paramétricas, y por unas ecuaciones implícitas.*

Demostración: Toda variedad lineal en un espacio de dimensión finita tiene un sistema finito de generadores, luego admite unas ecuaciones paramétricas. El método anterior nos explica cómo conseguir unas ecuaciones implícitas a partir de éstas, luego L admite también unas ecuaciones implícitas. $\square$

3.4. Intersección y suma de variedades.

Ya hemos visto cómo se puede determinar una variedad lineal usando ecuaciones paramétricas o implícitas, y cómo calcular su dimensión. Continuaremos con algunas propiedades sencillas de las variedades lineales:

Proposición 3.12 *Si L_1 y L_2 son dos variedades lineales de un espacio vectorial V, entonces $L_1 \cap L_2$ es una variedad lineal.*

Demostración: Sean $\mathbf{v}_1, \mathbf{v}_2 \in L_1 \cap L_2$. Como pertenecen a L_1, entonces $\mathbf{v}_1 + \mathbf{v}_2 \in L_1$, al ser L_1 variedad lineal. Pero como también pertenecen a L_2, entonces $\mathbf{v}_1 + \mathbf{v}_2 \in L_2$. Por tanto, $\mathbf{v}_1 + \mathbf{v}_2 \in L_1 \cap L_2$.

Análogamente se demuestra que si $\alpha \in K$ y $\mathbf{v} \in L_1 \cap L_2$, entonces $\alpha\mathbf{v} \in L_1 \cap L_2$. Por tanto, $L_1 \cap L_2$ satisface las dos propiedades necesarias y suficientes para ser una variedad lineal. $\square$

Proposición 3.13 *Sean S y T dos sistemas de vectores de un espacio vectorial V. Se tiene:*

1. $S \subset \langle S \rangle$.
2. $S = \langle S \rangle \ \Leftrightarrow\ S$ *es una variedad lineal.*
3. $S \subset T \ \Rightarrow\ \langle S \rangle \subset \langle T \rangle$.
4. $\langle S \cap T \rangle \ \subset\ \langle S \rangle \cap \langle T \rangle$.
5. $\langle S \rangle \cup \langle T \rangle \ \subset\ \langle S \cup T \rangle$.

Demostración:

1. Trivial.

2. Evidente a partir de las definiciones, ya que $\langle S \rangle$ es una variedad lineal.

3. Si $\mathbf{v} \in \langle S \rangle$, entonces es combinación lineal de los vectores de S. Pero como $S \subset T$, $\mathbf{v}$ es combinación lineal de los vectores de T, es decir, $\mathbf{v} \in \langle T \rangle$.

4. Si un vector es combinación lineal de los vectores de $S \cap T$, entonces es combinación lineal de los vectores de S, y también es combinación lineal de los vectores de T, es decir, pertenece a $\langle S \rangle \cap \langle T \rangle$.

5. Como $S \subset S \cup T$, se tiene $\langle S \rangle \subset \langle S \cup T \rangle$. Del mismo modo $\langle T \rangle \subset \langle S \cup T \rangle$. Por tanto, $\langle S \rangle \cup \langle T \rangle \subset \langle S \cup T \rangle$.

□

Observación: Si conocemos L_1 y L_2, y queremos conocer $L_1 \cap L_2$, sólo tenemos que tomar unas ecuaciones implícitas de L_1 y unas ecuaciones implícitas de L_2. El conjunto formado por **todas las ecuaciones** formará unas ecuaciones implícitas de $L_1 \cap L_2$. En efecto, un vector está en $L_1 \cap L_2$, es decir, está en L_1 y en L_2, si y sólo si satisface las ecuaciones que definen L_1 y además las que definen L_2.

Una vez que hemos visto que la intersección de variedades lineales es una variedad lineal, y hemos estudiado algunas de sus propiedades, podríamos intentar hacer lo mismo con la unión de variedades lineales. Pero hay que tener cuidado:

Nota: Aunque la intersección de dos variedades lineales es una variedad lineal, la **unión** de dos variedades lineales **no es una variedad lineal**, en general. Por ejemplo, en $\mathbb{R}^3$, la unión de dos rectas que pasan por el origen no tiene por qué ser una recta, y por supuesto no es un punto, ni un plano, ni todo el espacio.

De todas formas, aunque $L_1 \cup L_2$ no sea una variedad lineal, si lo que necesitamos es una variedad que contenga a L_1 y a L_2, nos basta tomar $\langle L_1 \cup L_2 \rangle$. Tenemos entonces la siguiente definición:

Suma de variedades lineales: Sean L_1 y L_2 dos variedades lineales de un espacio vectorial V. Se llama **suma** de L_1 y L_2 a la variedad lineal:

$$L_1 + L_2 = \langle L_1 \cup L_2 \rangle.$$

Observación: Por definición, si conocemos L_1 y L_2 y queremos hallar L_1+L_2, sólo tenemos que tomar un sistema de generadores S_1 de L_1 y un sistema de generadores S_2 de L_2. La unión de estos dos conjuntos, $S_1 \cup S_2$, será un sistema de generadores de $L_1 + L_2$.

3.5. Propiedades de la suma de variedades. Fórmula de la dimensión.

Dadas dos variedades L_1 y L_2 en un espacio vectorial de dimensión finita V, hemos definido la variedad suma $L_1 + L_2$. La causa de que esta variedad lineal se llame *suma*, se encuentra en el siguiente resultado:

Proposición 3.14 *Sean L_1 y L_2 dos variedades lineales de un espacio vectorial V. se tiene:*

$$L_1 + L_2 = \{\mathbf{v}_1 + \mathbf{v}_2; \quad \mathbf{v}_1 \in L_1, \ \mathbf{v}_2 \in L_2\}.$$

Demostración: Si $\mathbf{v} \in L_1+L_2$, entonces es combinación lineal de los vectores de $L_1 \cup L_2$. Separemos esta combinación lineal en dos sumandos $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, donde en $\mathbf{v}_1$ están todos los términos en que aparece un vector de L_1, y $\mathbf{v}_2$ contiene el resto de los términos, que necesariamente consta de vectores de L_2. Entonces $\mathbf{v}_1 \in \langle L_1 \rangle = L_1$, y $\mathbf{v}_2 \in \langle L_2 \rangle = L_2$.

La otra inclusión es trivial. $\square$

Veamos ahora que la suma de dos variedades, $L_1 + L_2$, es en realidad la variedad más pequeña que hubiéramos podido escoger, conteniendo a $L_1 \cup L_2$.

Proposición 3.15 *Dado un sistema de vectores S de un espacio vectorial V, la variedad lineal $\langle S \rangle$ es la menor variedad lineal que contiene a S. Es decir, si L es una variedad lineal que contiene a S, entonces $\langle S \rangle \subseteq L$.*

Demostración: Si una variedad L contiene a S, es decir, si $S \subset L$, entonces $\langle S \rangle \subset \langle L \rangle = L$. $\square$

Corolario 3.16 *$L_1 + L_2$ es la menor variedad lineal que contiene a L_1 y a L_2.*

Pero no tenemos por qué restringirnos a sumar sólo dos variedades. Podemos sumar tantas como queramos, siempre que sea un número finito.

Sea V un espacio vectorial, y sean $L_1, \ldots, L_m$ variedades lineales de V. Se define la suma de todas estas variedades como la variedad lineal

$$\sum_{i=1}^{m} L_i = L_1 + L_2 + \cdots + L_m = \langle L_1 \cup L_2 \cup \cdots \cup L_m \rangle.$$

De forma análoga a la proposición anterior, se demuestra lo siguiente:

Proposición 3.17 *Si $L_1, \ldots, L_m$ son variedades lineales de un espacio vectorial V, entonces*

$$L_1 + \cdots + L_m = \{\mathbf{v}_1 + \cdots + \mathbf{v}_m; \quad \mathbf{v}_i \in L_i, \ \forall i = 1, \ldots, m\}.$$

Finalizamos esta sección con uno de los teoremas más importantes del álgebra lineal, que relaciona las dimensiones de dos variedades cualesquiera, su suma y su intersección. Este teorema es muy útil para calcular dimensiones de variedades lineales.

Teorema 3.18 (Fórmula de la dimensión) *Sean L_1 y L_2 dos variedades lineales de un espacio vectorial V de dimensión finita. Se tiene:*

$$\dim L_1 + \dim L_2 = \dim(L_1 + L_2) + \dim(L_1 \cap L_2).$$

Demostración: Sea $B_0 = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ una base de $L_1 \cap L_2$. Por el teorema de la base incompleta, podemos ampliar B_0 hasta una base de L_1, y también la podemos ampliar hasta una base de L_2. Es decir, existen dos sistemas de vectores, $S_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$ y $S_2 = \{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$ tales que $B_1 = B_0 \cup S_1$ es una base de L_1, y $B_2 = B_0 \cup S_2$ es una base de L_2.

Sea $B = B_0 \cup S_1 \cup S_2$. Vamos a demostrar que B es base de $L_1 + L_2$, y con eso habremos probado el teorema, ya que $\dim L_1 = r + s$, $\dim L_2 = r + t$, $\dim(L_1 \cap L_2) = r$, y en este caso $\dim(L_1 + L_2) = r + s + t$.

B es sistema de generadores de $L_1 + L_2$, ya que $B = B_1 \cup B_2$. Por tanto, sólo tenemos que ver que es linealmente independiente. Consideremos una combinación lineal:

$$\sum_{i=1}^{r} \alpha_i \mathbf{u}_1 + \sum_{j=1}^{s} \beta_j \mathbf{v}_j + \sum_{k=1}^{t} \gamma_k \mathbf{w}_k = \mathbf{0}.$$

Hay que demostrar que todos los coeficientes deben ser nulos. Sea

$$\mathbf{v} = \sum_{i=1}^{r} \alpha_i \mathbf{u}_1 + \sum_{j=1}^{s} \beta_j \mathbf{v}_j = -\sum_{k=1}^{t} \gamma_k \mathbf{w}_k.$$

De la primera forma de escribir $\mathbf{v}$ se obtiene que $\mathbf{v} \in L_1$, y de la segunda, que $\mathbf{v} \in L_2$. Por tanto, $\mathbf{v} \in L_1 \cap L_2$, y así $\mathbf{v}$ se escribe de forma única como combinación lineal de los vectores de B_0. Como también se escribe de forma única como combinación lineal de los vectores de B_1 (la fórmula anterior), y $B_0 \subset B_1$, estas dos formas de escribirlo deben ser la misma. Por tanto, $\beta_1 = \cdots = \beta_s = 0$.

Después de esto, nos queda

$$\sum_{i=1}^{r} \alpha_i \mathbf{u}_1 + \sum_{k=1}^{t} \gamma_k \mathbf{w}_k = \mathbf{0},$$

pero esta es una combinación lineal de los vectores de B_2, que es una base, luego todos los coeficientes son nulos. $\square$

3.6. Descomposición de variedades. Espacio producto y cociente.

como vimos en la sección precedente, las variedades lineales se pueden intersecar o sumar. En esta sección veremos que, si $L = L_1 + L_2$, hay ocasiones en que las propiedades de la variedad L se pueden estudiar fácilmente a partir de las propiedades de L_1 y L_2. Para ver cómo esto es posible, definiremos la *suma directa* de variedades:

Suma directa: Diremos que dos variedades lineales L_1, L_2 son **independientes**, o que su suma $L_1 + L_2$ es **suma directa**, que escribiremos $L_1 \oplus L_2$, si

$$L_1 \cap L_2 = \{\mathbf{0}\}.$$

La suma directa recuerda mucho al concepto de base. En particular, por el siguiente resultado:

Proposición 3.19 *Sean L_1 y L_2 dos variedades lineales de un espacio vectorial V. La suma $L_1 + L_2$ es directa si y sólo si cualquier vector $\mathbf{v} \in L_1 + L_2$ se puede escribir, de una única forma, como $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, donde $\mathbf{v}_1 \in L_1$ y $\mathbf{v}_2 \in L_2$.*

Demostración: Supongamos que $L_1 \oplus L_2$. Si un vector $\mathbf{v}$ se pudiera escribir de dos maneras distintas, $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_1' + \mathbf{v}_2'$, donde $\mathbf{v}_1, \mathbf{v}_1' \in L_1$ y $\mathbf{v}_2, \mathbf{v}_2' \in L_2$, entonces $\mathbf{v}_1 \neq \mathbf{v}_1'$ (si no, tendríamos también $\mathbf{v}_2 = \mathbf{v}_2'$, y la descomposición sería la misma). Consideremos $\mathbf{u} = \mathbf{v}_1 - \mathbf{v}_1' \neq \mathbf{0}$. Entonces $\mathbf{u} \in L_1$, pero además

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_1' + (\mathbf{u} + \mathbf{v}_2) = \mathbf{v}_1' + \mathbf{v}_2' \quad \Rightarrow \quad \mathbf{u} = \mathbf{v}_2' - \mathbf{v}_2 \in L_2.$$

Por tanto, $\mathbf{u} \in L_1 \cap L_2$, lo que contradice que la suma de L_1 y L_2 sea directa.

Supongamos ahora que cualquier vector se puede escribir de forma única como suma de vectores de L_1 y L_2. Si existiera un vector $\mathbf{v} \in L_1 \cap L_2$, entonces podríamos escribir $\mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v}$, que serían dos descomposiciones distintas. Esto es imposible, por tanto $L_1 \cap L_2 = \{\mathbf{0}\}$, y la suma de estas dos variedades es directa. □

Corolario 3.20 *Sean L_1 y L_2 dos variedades lineales de un espacio vectorial V. La suma $L_1 + L_2$ es directa si y sólo si, si se tiene $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$, con $\mathbf{v}_1 \in L_1$ y $\mathbf{v}_2 \in L_2$, entonces $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$.*

Demostración: Si $L_1 \oplus L_2$, entonces $\mathbf{0}$ se puede escribir de una única forma como suma de vectores de L_1 y L_2. Por tanto, si $\mathbf{0} = \mathbf{v}_1 + \mathbf{v}_2$, sólo hay una posiblidad: $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$.

Por otra parte, supongamos que el vector $\mathbf{0}$ sólo se puede escribir $\mathbf{0} + \mathbf{0}$ como suma de vectores de L_1 y L_2. Si la suma de L_1 y L_2 no fuera directa, existiría un vector $\mathbf{v} \in L_1 + L_2$ que se podría escribir de dos formas distintas como suma de vectores de L_1 y L_2, digamos $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_1' + \mathbf{v}_2'$. Pero entonces

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (\mathbf{v}_1 - \mathbf{v}_1') - (\mathbf{v}_2 - \mathbf{v}_2'),$$

donde $\mathbf{v}_1 - \mathbf{v}_1' \in L_1$ y $\mathbf{v}_2 - \mathbf{v}_2' \in L_2$, por tanto $\mathbf{v}_1 - \mathbf{v}_1' = \mathbf{0}$ y $\mathbf{v}_2 - \mathbf{v}_2' = \mathbf{0}$, es decir, la descomposición es la misma. Por tanto, se tiene $L_1 \oplus L_2$. □

Estas dos caracterizaciones nos permiten extender la definición de suma directa a más de dos variedades lineales.

Suma directa: Dadas m variedades lineales $L_1, \ldots, L_m$ de un espacio vectorial V, se dice que son **independientes**, o que su suma $L_1 + \cdots + L_m$ es **suma directa**, que escribiremos $L_1 \oplus L_2 \oplus \cdots \oplus L_m$, si cualquier vector $\mathbf{v}$ de dicha suma se puede escribir, de una única forma, como

$$\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_m,$$

donde $\mathbf{v}_i \in L_i$ para todo $i = 1, \ldots, m$.

También se tiene la caracterización análoga al caso de dos variedades lineales, con la misma demostración:

Proposición 3.21 *Sean* $L_1, \ldots, L_m$ *variedades lineales de un espacio vectorial* V*. Su suma es directa si y sólo si, si se tiene* $\mathbf{0} = \mathbf{v}_1 + \cdots + \mathbf{v}_m$*, con* $\mathbf{v}_i \in L_i$ *para todo* i*, entonces* $\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_m = \mathbf{0}$.

Estos conceptos de suma y de suma directa de variedades lineales se ven más claramente cuando todas las variedades son de dimensión 1. En ese caso, se tiene:

Proposición 3.22 *Sea* V *un espacio vectorial y sea* $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ *un sistema finito de vectores de* V*. Se tiene:*

1. $\langle S \rangle = \langle \mathbf{v}_1 \rangle + \cdots + \langle \mathbf{v}_m \rangle$.
2. S *es linealmente independiente si y sólo si* $\langle S \rangle = \langle \mathbf{v}_1 \rangle \oplus \cdots \oplus \langle \mathbf{v}_m \rangle$.

DEMOSTRACIÓN: Las dos propiedades se obtienen directamente a partir de las definiciones.
□

Un caso especial, e importante, de suma directa de dos subespacios es el siguiente:

Dado un espacio vectorial V, dos variedades lineales L_1 y L_2 de V se dicen **suplementarias** si $L_1 \oplus L_2 = V$.

De la misma forma que hemos probado los resultados anteriores, se tiene:

Proposición 3.23 *Sea* V *un espacio vectorial, y sean* L_1 *y* L_2 *dos variedades lineales de* V*. Las condiciones siguientes son equivalentes:*

1. L_1 *y* L_2 *son suplementarios.*
2. $L_1 + L_2 = V$, *y* $L_1 \cap L_2 = \{\mathbf{0}\}$.
3. *Todo vector de* $\mathbf{v} \in V$ *se descompone de forma única como una suma* $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$*, donde* $\mathbf{v}_1 \in L_1$ *y* $\mathbf{v}_2 \in L_2$.

3.7. Propiedades de la suma directa. Espacio producto.

La importancia de los espacios suplementarios procede de la facilidad para manejar sus bases y dimensiones:

Proposición 3.24 *Sea V un espacio vectorial de dimensión finita, y sean L_1 y L_2 dos espacios suplementarios, con bases respectivas B_1 y B_2. Se tiene:*

1. $B_1 \cup B_2$ *es base de* V.
2. $\dim L_1 + \dim L_2 = \dim V$.

Demostración: Como $L_1 \oplus L_2 = V$, entonces todo vector de V puede escribirse de una única forma como suma de un vector de L_1 y otro de L_2. Pero como B_1 es base de L_1 y B_2 es base de L_2, estos dos vectores se escriben de forma única como combinación lineal de los vectores de B_1 y B_2. Es decir, cualquier vector de V se escribe de forma única como combinación lineal de los vectores de $B_1 \cup B_2$, luego este conjunto es una base de V.

La segunda propiedad es una consecuencia directa de la primera. $\square$

El recíproco del resultado anterior también s cierto:

Proposición 3.25 *Sea $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_s, \mathbf{u}_{s+1}, \ldots, \mathbf{u}_t\}$ una base de un espacio vectorial V. Sean $B_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$ y $B_2 = \{\mathbf{u}_{s+1}, \ldots, \mathbf{u}_t\}$. Entonces $\langle B_1 \rangle$ y $\langle B_2 \rangle$ son dos variedades suplementarias de V.*

Demostración: Directa. $\square$

Y por último, este resultado es una reescritura de un resultado anterior:

Proposición 3.26 *Sea V un espacio vectorial de tipo finito. Toda variedad lineal de V tiene alguna variedad suplementaria.*

Demostración: Esto es consecuencia del resultado anterior, y del teorema de la base incompleta. $\square$

Hemos visto, por tanto, cómo una variedad lineal L (es decir, un espacio vectorial) se puede descomponer en dos o más subespacios, $L_1 \oplus \cdots \oplus L_m$ de forma óptima: La dimensión de L es la suma de las dimensiones de cada L_i, y si conocemos una base de cada L_i, su unión es una base de L. Ahora veamos la operación contraria: dados dos o más espacios vectoriales sobre K, de tipo finito, $V_1, \ldots, V_m$, aunque no tengan nada que ver, podremos construir un espacio vectorial más grande, V, tal que $V = V_1 \oplus \cdots \oplus V_m$.

Producto de espacios vectoriales Dados dos espacios vectoriales de dimensión finita, V_1 y V_2 sobre un mismo cuerpo K, se define el **espacio producto** de V_1 y V_2 como el conjunto

$$V_1 \times V_2 = \{(\mathbf{v}_1, \mathbf{v}_2)\ ; \quad \mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2\},$$

donde se definen las siguientes operaciones internas:

- **Suma:** $(\mathbf{u}_1, \mathbf{u}_2) + (\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{v}_1, \mathbf{u}_2 + \mathbf{v}_2)$.
- **Producto por escalar:** $\alpha(\mathbf{v}_1, \mathbf{v}_2) = (\alpha\mathbf{v}_1, \alpha\mathbf{v}_2)$.

Proposición 3.27 *Dados dos espacios vectoriales de tipo finito, V_1 y V_2, sobre un mismo cuerpo K, el espacio producto $V_1 \times V_2$ es un espacio vectorial. Además,* $\dim(V_1 \times V_2) = \dim(V_1) + \dim(V_2)$.

Demostración: Se prueba que $V_1 \times V_2$ es un espacio vectorial directamente a partir de la definición. Para probar que su dimensión es la suma de las de V_1 y V_2, tomemos una base $B_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ de V_1, y una base $B_2 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ de V_2. Se prueba de forma directa que el sistema de vectores

$$B = ((\mathbf{u}_1, 0), \ldots, (\mathbf{u}_m, 0), (0, \mathbf{v}_1), \ldots, (0, \mathbf{v}_n))$$

es base de $V_1 \times V_2$. Por tanto, $\dim(V_1 \times V_2) = m + n = \dim(V_1) + \dim(V_2)$. $\square$

3.8. Espacio cociente.

Terminaremos esta sección, y este tema, estudiando una noción que es básica en muchas ramas de las matemáticas, en particular en el álgebra lineal: el *espacio cociente*. Fijaremos a partir de ahora un espacio vectorial V, y una variedad lineal $L \subset V$. Básicamente, se puede pensar en el espacio cociente de V sobre L como si fuera el espacio V, pero donde los vectores de L no tienen ningún valor: es decir, cualquier vector de L representa el vector $\mathbf{0}$ del espacio cociente; y si sumamos a cualquier vector del cociente un vector de L, éste se queda igual. Vamos a definirlo de forma rigurosa:

Dos vectores $\mathbf{u}$ y $\mathbf{v}$ de V se dicen **L-equivalentes** si su diferencia pertenece a L. Escribiremos:

$$\mathbf{u} \sim_L \mathbf{v} \quad \Leftrightarrow \quad \mathbf{u} - \mathbf{v} \in L.$$

Proposición 3.28 *La L-equivalencia es una relación de equivalencia.*

Demostración: Hay que demostrar las propiedades simétrica, reflexiva y transitiva. Las tres son directas a partir de la definición de variedad lineal. $\square$

Cuando se define, en cualquier conjunto, una relación de equivalencia, se pueden considerar los subconjuntos de elementos que están relacionados entre sí. Estos subconjuntos se llaman *clases de equivalencia*. En este caso, las clases de equivalencia se llaman *variedades lineales afines*.

Variedad lineal afín: Sea L una variedad lineal de un espacio vectorial V, y sea $\mathbf{v}$ un vector de V. Llamaremos **variedad lineal afín** que pasa por $\mathbf{v}$ con dirección L, y la notaremos $\mathbf{v} + L$, a la *clase de L-equivalencia* de $\mathbf{v}$, es decir, al conjunto formado por todos los vectores de V que son L-equivalentes a $\mathbf{v}$:

$$\mathbf{v} + L = \{\mathbf{u} \in V \; ; \quad \mathbf{u} \sim_L \mathbf{v}\} = \{\mathbf{v} + \mathbf{w} \; ; \quad \mathbf{w} \in L\}.$$

Ejemplo 3.29 *Si $V = \mathbb{R}^3$ y L es un plano que pasa por el origen de coordenadas, dos vectores $\mathbf{u}$ y $\mathbf{v}$ son L-equivalentes si su vector diferencia pertenece a L, es decir, si el segmento que une los puntos finales de $\mathbf{u}$ y $\mathbf{v}$ es paralelo al plano L. Por tanto, la variedad lineal afín que pasa por un vector $\mathbf{v}$ con dirección L, está formada por todos los vectores cuyos puntos finales forman un plano: el que contiene al punto final de $\mathbf{v}$ y es paralelo a L. Así, las variedades lineales con dirección L son, en cierto modo, todos los planos paralelos a L.*

Ejemplo 3.30 *Al igual que en el ejemplo anterior, si $V = \mathbb{R}^3$ y L es una recta que pasa por el origen, entonces las variedades lineales afines con dirección L vienen determinadas por las rectas paralelas a L, es decir las que tienen la misma dirección que la recta L.*

Una propiedad evidente de las variedades lineales afines es la siguiente:

Proposición 3.31 *Dados $\mathbf{u}, \mathbf{v} \in V$, se tiene:*

$$\mathbf{u} + L = \mathbf{v} + L \quad \Leftrightarrow \quad \mathbf{u} \sim_L \mathbf{v} \quad \Leftrightarrow \quad \mathbf{u} - \mathbf{v} \in L.$$

Nota: Aunque L sea una variedad lineal, las variedades lineales afines correspondientes **no son variedades lineales**, en general. Esto se puede ver en los dos ejemplos anteriores (los planos o rectas que no pasan por el origen no determinan variedades lineales), o bien por el siguiente razonamiento: Si $\mathbf{u} \in \mathbf{v} + L$, entonces $2\mathbf{u} \in \mathbf{v} + L$ si y sólo si $\mathbf{u} \in L$. Pero en ese caso, $\mathbf{v} \sim_L \mathbf{u} \sim_L \mathbf{0}$, luego $\mathbf{v} + L = \mathbf{0} + L$. Por tanto, la única variedad lineal afín con dirección L que es una variedad lineal es $\mathbf{0} + L$, es decir, la misma L.

De todas formas, aunque las variedades lineales afines no sean variedades lineales, sí van a ser los elementos de un nuevo espacio vectorial, llamado *espacio cociente*.

Espacio cociente: Sea L una variedad lineal de un espacio vectorial V. Llamaremos **espacio cociente** de V sobre L, y lo denotaremos V/L, al conjunto formado por las variedades lineales afines con dirección L, donde definimos las siguientes operaciones:

- **Suma:** $(\mathbf{u} + L) + (\mathbf{v} + L) = (\mathbf{u} + \mathbf{v}) + L$.
- **Producto por escalar:** $\alpha(\mathbf{u} + L) = (\alpha\mathbf{u}) + L$.

Proposición 3.32 *La suma y el producto que acabamos de dar, están bien definidos.*

DEMOSTRACIÓN: Necesitamos este resultado ya que, si queremos sumar variedades lineales afines, la suma no puede depender del representante (el vector) que tomemos. Es decir, debemos demostrar que, si $\mathbf{u} + L = \mathbf{u}' + L$ y además $\mathbf{v} + L = \mathbf{v}' + L$, entonces las clases de equivalencia $(\mathbf{u} + \mathbf{v}) + L$ y $(\mathbf{u}' + \mathbf{v}') + L$ son iguales. Pero sabemos que $\mathbf{u} \sim_L \mathbf{u}'$, luego $\mathbf{u} - \mathbf{u}' \in L$. Análogamente $\mathbf{v} - \mathbf{v}' \in L$. Por tanto, $(\mathbf{u} - \mathbf{u}') + (\mathbf{v} - \mathbf{v}') = (\mathbf{u} + \mathbf{v}) - (\mathbf{u}' + \mathbf{v}') \in L$. Es decir, $(\mathbf{u} + \mathbf{v}) \sim_L (\mathbf{u}' + \mathbf{v}')$, luego $(\mathbf{u} + \mathbf{v}) + L = (\mathbf{u}' + \mathbf{v}') + L$ como queríamos demostrar.

Por otro lado, si $\mathbf{u}+L=\mathbf{u}'+L$ y $\alpha\in K$, entonces $(\mathbf{u}-\mathbf{u}')\in L$, luego $\alpha(\mathbf{u}-\mathbf{u}')=\alpha\mathbf{u}-\alpha\mathbf{u}'\in L$. Por tanto $(\alpha\mathbf{u})+L=(\alpha\mathbf{u}')+L$, y se obtiene el resultado. $\square$

Teorema 3.33 *Sea L una variedad lineal de un espacio vectorial V sobre K. El espacio cociente V/L, con las dos operaciones que acabamos de definir, es un espacio vectorial sobre K. Además, si V es de dimensión finita, se tiene:*

$$\dim(V/L)=\dim(V)-\dim(L).$$

Demostración: La demostración de que V/L es un espacio vectorial, es directa. Observemos que el elemento neutro de la suma de clases es la clase $\mathbf{0}+L$. Para probar la fórmula que relaciona sus dimensiones, tomemos una base $B_1=(\mathbf{u}_1,\ldots,\mathbf{u}_r)$ de L. Esta base se podrá ampliar a una base $B=(\mathbf{u}_1,\ldots,\mathbf{u}_r,\mathbf{u}_{r+1},\ldots,\mathbf{u}_n)$ de V. Vamos a probar que $B_2=(\mathbf{u}_{r+1}+L,\ldots,\mathbf{u}_n+L)$ es una base de V/L, y esto demostrará el resultado.

Probemos primero que B_2 es sistema de generadores. Sea $\mathbf{v}+L$ una clase de equivalencia cualquiera. Como $\mathbf{v}\in V$, podremos escribirlo como combinación lineal de los elementos de B. Es decir, $\mathbf{v}=\alpha_1\mathbf{u}_1+\cdots+\alpha_n\mathbf{u}_n$. Sea $\mathbf{u}=\alpha_1\mathbf{u}_1+\cdots+\alpha_r\mathbf{u}_r$. Claramente $\mathbf{u}\in L$, luego $\mathbf{u}'=\mathbf{v}-\mathbf{u}\sim_L\mathbf{v}$, donde $\mathbf{u}'=\alpha_{r+1}\mathbf{u}_{r+1}+\cdots+\alpha_n\mathbf{u}_n$. Pero en ese caso $\mathbf{v}+L=\mathbf{u}'+L=\alpha_{r+1}(\mathbf{u}_{r+1}+L)+\cdots+\alpha_n(\mathbf{u}_n+L)$. Es decir, cualquier clase de equivalencia, $\mathbf{v}+L$, puede escribirse como combinación lineal de los elementos de B_2.

La demostración estará completa si probamos que B_2 es un sistema libre. Supongamos que tenemos una combinación lineal

$$\alpha_{r+1}(\mathbf{u}_{r+1}+L)+\cdots+\alpha_n(\mathbf{u}_n+L)=\mathbf{0}+L.$$

Esto implica que

$$(\alpha_{r+1}\mathbf{u}_{r+1}+\cdots+\alpha_n\mathbf{u}_n)+L=\mathbf{0}+L,$$

es decir, $(\alpha_{r+1}\mathbf{u}_{r+1}+\cdots+\alpha_n\mathbf{u}_n)\in L$. Pero la variedad lineal generada por los vectores $(\mathbf{u}_{r+1},\ldots,\mathbf{u}_n)$ es suplementaria a L (ya que B es una base), luego la única posiblidad es que $(\alpha_{r+1}\mathbf{u}_{r+1}+\cdots+\alpha_n\mathbf{u}_n)=\mathbf{0}$, por lo que $\alpha_{r+1}=\cdots=\alpha_n=0$. Esto nos dice que los elementos de B_2 son linealmente independientes. $\square$

Ejemplo 3.34 *Si L es un plano de $V=\mathbb{R}^3$, que pasa por el origen, los elementos del espacio cociente son los planos paralelos a L. La suma de dos planos Π_1 y Π_2, da como resultado otro plano Π_3: si se toma un vector $\mathbf{u}_1$ cuyo punto final esté en Π_1, y un vector $\mathbf{u}_2$, cuyo punto final esté en Π_2, el punto final del vector $\mathbf{u}_1+\mathbf{u}_2$ estará en Π_3. Del mismo modo, el producto de α por Π_1 es el plano que contiene al punto final del vector $\alpha\mathbf{u}_1$.*

Esta noción de espacio cociente será utilizada en el tema siguiente. Pero lo más importante de las variedades lineales afines es su relación con los sistemas de ecuaciones lineales. En el tema siguiente veremos que las soluciones de un sistema lineal cualquiera, forman siempre una variedad lineal afín.

Tema 4. Aplicaciones lineales

4.1. Definición y propiedades.

Cuando en matemáticas se estudia un tipo de conjuntos, se deben estudiar también las aplicaciones (o funciones) entre ellos. Si además, estos conjuntos tienen definidas operaciones internas o externas, nos interesarán las aplicaciones que preserven estas operaciones. Como estamos estudiando espacios vectoriales, veamos qué tipo de aplicaciones preservan sus dos operaciones: suma de vectores, y producto de vectores por escalares.

Aplicación lineal: Sean V y V' dos espacios vectoriales sobre un mismo cuerpo K. Sea f una aplicación de V en V', que escribiremos:

$$f: V \longrightarrow V'.$$

Esto quiere decir que a cada elemento $\mathbf{v} \in V$ le hacemos corresponder un elemento $f(\mathbf{v}) \in V'$. Diremos que f es una **aplicación lineal**, o un **homomorfismo**, si se cumplen las condiciones siguientes:

- $f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}) \qquad \forall \mathbf{u}, \mathbf{v} \in V.$
- $f(\alpha\mathbf{v}) = \alpha f(\mathbf{v}) \qquad \forall \alpha \in K, \ \forall \mathbf{v} \in V.$

Ejemplo 4.1 *Hay dos ejemplos triviales de aplicaciones lineales. En primer lugar, si $V = V'$, tenemos la llamada* **aplicación identidad**, *$id: V \longrightarrow V$, definida por $id(\mathbf{v}) = \mathbf{v}$ para todo $\mathbf{v} \in V$.*

Por otra parte, para cualesquiera V y V', siempre existe la **aplicación nula**, *$\mathcal{O}: V \longrightarrow V'$, definida por $\mathcal{O}(\mathbf{v}) = \mathbf{0}$, para todo $\mathbf{v} \in V$. Tanto la aplicación identidad como la aplicación nula son claramente aplicaciones lineales.*

Algunas propiedades básicas de las aplicaciones lineales son las siguientes:

Proposición 4.2 *Sea $f: V \to V'$ una aplicación lineal entre espacios vectoriales. Se tiene:*

1. $f(\mathbf{0}) = \mathbf{0}$, *para cualquier aplicación lineal f.*

2. $f(-\mathbf{v}) = -f(\mathbf{v}),\quad$ *para todo* $\mathbf{v} \in V$.

DEMOSTRACIÓN:

1. Para todo $\mathbf{v} \in V$, $f(\mathbf{v}) = f(\mathbf{v}+\mathbf{0}) = f(\mathbf{v}) + f(\mathbf{0})$, luego $f(\mathbf{0}) = f(\mathbf{v}) - f(\mathbf{v}) = \mathbf{0}$.
2. A partir de la definición, se tiene $f(-\mathbf{v}) = f((-1)\mathbf{v}) = (-1)f(\mathbf{v}) = -f(\mathbf{v})$.

□

Proposición 4.3 *Sea* $f: V \to V'$ *una aplicación lineal entre espacios vectoriales. Se tiene:*

1. $f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_r\mathbf{v}_r) = \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_r f(\mathbf{v}_r),\ \forall \alpha_1, \dots, \alpha_r \in K,\ \forall \mathbf{v}_1, \dots, \mathbf{v}_r \in V$.

2. Si $S = \{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ *es un sistema de vectores de* V *linealmente dependiente, entonces* $f(S) = \{f(\mathbf{u}_1), \dots, f(\mathbf{u}_r)\}$ *es un sistema de vectores de* V' *linealmente dependiente.*

3. Si $g: V' \to V''$ *es otra aplicación lineal, entonces la composición* $g \circ f: V \to V''$ *es una aplicación lineal.*

DEMOSTRACIÓN:

1. El resultado es cierto por definición para $r = 2$, ya que $f(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) = f(\alpha_1\mathbf{v}_1) + f(\alpha_2\mathbf{v}_2) = \alpha_1 f(\mathbf{v}_1) + \alpha_2 f(\mathbf{v}_2)$. Si suponemos el resultado cierto para $r-1$, con $r > 2$, se prueba para r de la siguiente manera:
$$f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_r\mathbf{v}_r) = f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_{r-1}\mathbf{v}_{r-1} + \alpha_r\mathbf{v}_r) =$$
$$f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_{r-1}\mathbf{v}_{r-1}) + f(\alpha_r\mathbf{v}_r) = \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_{r-1} f(\mathbf{v}_{r-1}) + \alpha_r f(\mathbf{v}_r).$$
2. Si S es un sistema linealmente dependiente, existen unos escalares $\alpha_1, \dots, \alpha_r$, no todos nulos, tales que $\alpha_1\mathbf{v}_1 + \cdots + \alpha_r\mathbf{v}_r = \mathbf{0}$. Aplicando f a ambos términos de esta igualdad, se obtiene
$$f(\alpha_1\mathbf{v}_1 + \cdots + \alpha_r\mathbf{v}_r) = f(0),$$
es decir, por el apartado anterior y por el resultado anterior,
$$\alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_r f(\mathbf{v}_r) = \mathbf{0}.$$
Por tanto, hemos obtenido el vector $\mathbf{0} \in V'$ como combinación lineal de los vectores de $f(S)$, donde no todos los coeficientes son nulos, luego $f(S)$ es linealmente dependiente.

3. Dados $\mathbf{u}, \mathbf{v} \in V$ y $\alpha \in K$, se tiene $(g \circ f)(\mathbf{u}+\mathbf{v}) = g(f(\mathbf{u}+\mathbf{v})) = g(f(\mathbf{u}) + f(\mathbf{v})) = g(f(\mathbf{u})) + g(f(\mathbf{v})) = (g\circ f)(\mathbf{u}) + (g \circ f)(\mathbf{v})$, y por otra parte, $(g \circ f)(\alpha \mathbf{v}) = g(f(\alpha \mathbf{v})) = g(\alpha f(\mathbf{v})) = \alpha g(f(\mathbf{v})) = \alpha\,(g \circ f)(\mathbf{v})$. Luego $g \circ f$ es una aplicación lineal.

□

4.2. Imagen y núcleo.

Dados dos espacios vectoriales, V y V', al conjunto de los homomorfismos (aplicaciones lineales) de V en V' lo denotaremos $\mathrm{Hom}(V, V')$. Por tanto, a partir de ahora en lugar de decir: "sea $f : V \to V'$ una aplicación lineal", diremos: "sea $f \in \mathrm{Hom}(V, V')$".

Imagen y núcleo: Sea $f \in \mathrm{Hom}(V, V')$. Se llama **imagen** de f, denotada por $\mathrm{Im}(f)$ o por $f(V)$, al siguiente subconjunto de V':

$$\mathrm{Im}(f) = \{f(\mathbf{v}); \quad \mathbf{v} \in V\}.$$

Se llama **núcleo** de f, denotado por $\ker(f)$ o por $f^{-1}(\mathbf{0})$, al siguiente subconjunto de V:

$$\ker(f) = \{\mathbf{v} \in V \mid f(\mathbf{v}) = 0\}.$$

Proposición 4.4 *Dada $f \in \mathrm{Hom}(V, V')$, los conjuntos Im$(f)$ y $\ker(f)$ son variedades lineales de V' y V, respectivamente.*

Demostración: Dados $\mathbf{u}, \mathbf{v} \in \ker(f)$ y $\alpha \in K$, se tiene $f(\mathbf{u}) = \mathbf{0}$ y $f(\mathbf{v}) = \mathbf{0}$. entonces $f(\mathbf{u}+\mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$., luego $\mathbf{u}+\mathbf{v} \in \ker(f)$, y además $f(\alpha\mathbf{v}) = \alpha\, f(\mathbf{v}) = \alpha\mathbf{0} = \mathbf{0}$, por lo que $\alpha\mathbf{v} \in \ker(f)$. Por tanto $\ker(f)$ es una variedad lineal de V.

Por otra parte, dados $\mathbf{u}', \mathbf{v}' \in \mathrm{Im}(f)$ y $\alpha \in K$, existen $\mathbf{u}, \mathbf{v} \in V$ tales que $f(\mathbf{u}) = \mathbf{u}'$ y $f(\mathbf{v}) = \mathbf{v}'$. Entonces se tiene $\mathbf{u}' + \mathbf{v}' = f(\mathbf{u}) + f(\mathbf{v}) = f(\mathbf{u}+\mathbf{v})$, luego $\mathbf{u}'+\mathbf{v}' \in \mathrm{Im}(f)$, y además $\alpha\mathbf{v}' = \alpha f(\mathbf{v}) = f(\alpha\mathbf{v})$, luego $\alpha\mathbf{v}' \in \mathrm{Im}(f)$. Por tanto $\mathrm{Im}(f)$ es una variedad lineal de V'.

□

Proposición 4.5 *Si* $G = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *es un sistema de generadores de* V*, entonces* $f(G) = \{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)\}$ *es un sistema de generadores de Im*(f)*.*

DEMOSTRACIÓN: Dado un vector $\mathbf{v}' \in \mathrm{Im}(f)$, existe un vector $\mathbf{v} \in V$ tal que $f(\mathbf{v}) = \mathbf{v}'$. Como G es un sistema de generadores de V, existen unos escalares $\alpha_1, \ldots, \alpha_n \in K$ tales que $\mathbf{v} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. Aplicando f as ambos lados de esta ecuación, se tiene $f(\mathbf{v}) = f(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n)$, es decir, $\mathbf{v}' = \alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_n f(\mathbf{u}_n)$. Por tanto $f(G)$ es sistema de generadores de $\mathrm{Im}(f)$. $\square$

Rango: Dada una aplicación lineal $f \in \mathrm{Hom}(V, V')$, llamamos **rango** de f a la dimensión de $\mathrm{Im}(f)$.

Teorema 4.6 *Si* V *es un espacio vectorial de dimensión finita, y* $f \in \mathrm{Hom}(V, V')$*, entonces*

$$\dim\ker(f) + \dim\mathrm{Im}(f) = \dim V.$$

DEMOSTRACIÓN: Como V es de dimensión finita, entonces $\ker(f) \subset V$ también tiene dimensión finita. Sea $B_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ una base de $\ker(f)$. Ampliemos esta base hasta una base $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ de V. Por la proposición anterior, el sistema $f(B)$ será un sistema de generadores de $\mathrm{Im}(f)$. Pero como $f(\mathbf{u}_i) = \mathbf{0}$ para todo $i \leq r$, se tiene: $f(B) = \{\mathbf{0}, \ldots, \mathbf{0}, f(\mathbf{u}_{r+1}), \ldots, f(\mathbf{u}_n)\}$.

Por tanto, el sistema $S = \{f(\mathbf{u}_{r+1}), \ldots, f(\mathbf{u}_n)\}$ es un sistema de generadores de $\mathrm{Im}(f)$. Para terminar la demostración, necesitamos probar que S es libre. Si tuviéramos una combinación lineal

$$\alpha_{r+1} f(\mathbf{u}_{r+1}) + \cdots + \alpha_n f(\mathbf{u}_n) = \mathbf{0},$$

tendríamos, al ser f una aplicación lineal:

$$f(\alpha_{r+1}\mathbf{u}_{r+1} + \cdots + \alpha_n\mathbf{u}_n) = \mathbf{0} \quad \Rightarrow \quad \alpha_{r+1}\mathbf{u}_{r+1} + \cdots + \alpha_n\mathbf{u}_n \in \ker(f).$$

Pero el espacio generado por $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ es suplementario a $\ker(f)$ (al ser B una base de V), por tanto, $\alpha_{r+1}\mathbf{u}_{r+1} + \cdots + \alpha_n\mathbf{u}_n = 0$. Pero entonces, al ser B una base, se tiene $\alpha_{r+1} = \cdots = \alpha_n = 0$. Es decir, S es libre, como queríamos demostrar. $\square$

4.3. Imagen e imagen inversa de variedades lineales. Aplicaciones inyectivas.

Si tenemos una aplicación $f \in \mathrm{Hom}(V, V')$, podemos preguntarnos en qué se transforman, mediante f, las variedades lineales de V y de V'.

Proposición 4.7 *Sea $f \in \mathrm{Hom}(V, V')$ y sea L una variedad lineal de V. Entonces $f(L) = \{f(\mathbf{v});\ \mathbf{v} \in L\}$ es una variedad lineal de V'.*

Demostración: Se puede demostrar de forma directa, pero hay otra demostración más interesante. Consideremos la aplicación $f_{|L} : L \to V'$, que a cada vector $\mathbf{v} \in L$ le asocia $f_{|L}(\mathbf{v}) = f(\mathbf{v})$. Esta aplicación se llama *restricción de f a L*. Como $f_{|L}$ coincide con f en todos los vectores de L, y L es una variedad lineal (i.e. un espacio vectorial), $f_{|L}$ satisface todas las propiedades de aplicación lineal, es decir $f_{|L} \in \mathrm{Hom}(L, V')$. Pero es evidente que $\mathrm{Im}(f_{|L}) = f_{|L}(L) = f(L)$. Como sabemos que la imagen de cualquier aplicación lineal es una variedad lineal, se sigue que $f(L)$ es variedad lineal, como queríamos demostrar. $\square$

Sea $f \in \mathrm{Hom}(V, V')$. Dado $\mathbf{v}' \in V'$, llamamos **imagen inversa** de $\mathbf{v}'$ por f, al conjunto:

$$f^{-1}(\mathbf{v}') = \{\mathbf{v} \in V \mid f(\mathbf{v}) = \mathbf{v}'\}.$$

Dada una variedad lineal $L' \subset V'$, llamamos **imagen inversa** de L' por f, al conjunto:

$$f^{-1}(L') = \{\mathbf{v} \in V \mid f(\mathbf{v}) \in L'\} = \bigcup_{\mathbf{v}' \in L'} f^{-1}(\mathbf{v}').$$

Proposición 4.8 *Sea $f \in \mathrm{Hom}(V, V')$. La imagen inversa por f de cualquier variedad lineal de V' es una variedad lineal de V.*

Demostración: Directa. $\square$

Nota: Observemos que este resultado nos demuestra, de otra manera, que $\ker(f)$ es una variedad lineal, ya que $\ker(f) = f^{-1}(\{\mathbf{0}\})$.

Pero podemos decir todavía más:

Proposición 4.9 *Sea $f \in \mathrm{Hom}(V, V')$, y sea $\mathbf{v}' \in V'$. Si $\mathbf{v}' \notin \mathrm{Im}(f)$, entonces $f^{-1}(\mathbf{v}') = \emptyset$. Pero si $\mathbf{v}' \in \mathrm{Im}(f)$, entonces la imagen inversa $f^{-1}(\mathbf{v}')$ es una variedad lineal afín, con dirección $\ker(f)$.*

Demostración: La primera afirmación es evidente. Supongamos entonces que $\mathbf{v}' \in \mathrm{Im}(f)$. Esto quiere decir que existe un vector $\mathbf{v} \in V$ tal que $f(\mathbf{v}) = \mathbf{v}'$. Como, en este caso, $\mathbf{v}$ es un vector de $f^{-1}(\mathbf{v}')$, tenemos que demostrar que

$$f^{-1}(\mathbf{v}') = \mathbf{v} + \ker(V).$$

Pero un vector $\mathbf{v}_0$ pertenece a $f^{-1}(\mathbf{v}')$ si y sólo si $f(\mathbf{v}_0) = \mathbf{v}'$. Esto ocurre si y sólo si $f(\mathbf{v}) - f(\mathbf{v}_0) = f(\mathbf{v} - \mathbf{v}_0) = \mathbf{0}$, es decir, si y sólo si $\mathbf{v} - \mathbf{v}_0 \in \ker(f)$. Hemos probado entonces que un vector está en $f^{-1}(\mathbf{v}')$ si y sólo si está en la variedad lineal afín $\mathbf{v} + \ker(f)$. Esto es lo que queríamos demostrar. $\square$

Recordemos que una aplicación $f : V \to V'$ se dice **inyectiva** si no hay dos elementos de V cuya imagen por f sea la misma. Y f se dice **sobreyectiva** si $f(V) = V'$, o dicho de otra forma, si todo elemento de V' tiene una *preimagen* por f. Una aplicación inyectiva y sobreyectiva se dice **biyectiva**. En ese caso cada elemento de V' está relacionado (mediante f) con uno, y sólo uno, de los elementos de V.

Los homomorfismos inyectivos pueden determinarse fácilmente mediante su núcleo:

Proposición 4.10 *Una aplicación lineal $f \in \mathrm{Hom}(V, V')$ es inyectiva si y sólo si $\ker(f) = \{\mathbf{0}\}$.*

Demostración: En primer lugar, supongamos que $\ker(f) \neq \{\mathbf{0}\}$. Entonces existirá un vector no nulo $\mathbf{v} \in \ker(f)$, luego se tiene $\mathbf{v} \neq \mathbf{0}$ y $f(\mathbf{v}) = \mathbf{0} = f(\mathbf{0})$. Por tanto, f no es inyectiva.

Recíprocamente, supongamos que f no es inyectiva. Entonces existirán dos vectores $\mathbf{u} \neq \mathbf{v}$ en V tales que $f(\mathbf{u}) = f(\mathbf{v})$. Pero en ese caso tenemos $f(\mathbf{u} - \mathbf{v}) = f(\mathbf{u}) - f(\mathbf{v}) = \mathbf{0}$. Por tanto $\mathbf{u} - \mathbf{v} \in \ker(f)$, luego en $\ker(f)$ hay un vector no nulo, es decir, $\ker(f) \neq \{\mathbf{0}\}$. $\square$

Aparte de esta caracterización, las aplicaciones lineales inyectivas tienen otras propiedades interesantes:

Proposición 4.11 *Sea $f \in \mathrm{Hom}(V, V')$, donde V tiene dimensión finita. Las siguientes condiciones son equivalentes:*

1. *f es inyectiva.*

2. $\dim V = \dim f(V)$.

3. *Si $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base de V, entonces $f(B) = \{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)\}$ es una base de $f(V)$.*

4. *Si $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base de V, entonces $f(B)$ es un sistema libre.*

5. *Para todo sistema libre $S \subset V$, el sistema $f(S)$ también es libre.*

Demostración: Las dos primeras condiciones son equivalentes, ya que f es inyectiva si y sólo si $\ker(f) = \{\mathbf{0}\}$, es decir, si y sólo si $\dim \ker(f) = 0$. Por la fórmula que relaciona las dimensiones de V, $\ker(f)$ e $\mathrm{Im}(f)$, esto es equivalente a $\dim V = \dim \mathrm{Im}(f) = \dim f(V)$.

Por otra parte, sabemos que si $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ es una base de V, entonces $f(B)$ es un sistema de generadores de $f(V)$. Por tanto, $f(B)$ será base de $f(V)$ si y sólo si $f(B)$ es un sistema libre, lo que sucede si y sólo si $\dim V = \dim f(V)$. Por tanto, las condiciones 2, 3 y 4 son equivalentes.

Como la condición 4 es un caso particular de la condición 5, ésta última implica la anterior. Sólo nos queda por demostrar, entonces, que cualquiera de las 4 primeras condiciones implica la condición 5. Supongamos entonces que f es inyectiva, y sea $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ un sistema libre de vectores de V. Para demostrar que $f(S)$ es libre, tomemos una combinación lineal:

$$\alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_r f(\mathbf{u}_r) = \mathbf{0}.$$

Hay que demostrar que todos los coeficientes son nulos. Pero como f es una aplicación lineal, tenemos:

$$\alpha_1 f(\mathbf{u}_1) + \cdots + \alpha_r f(\mathbf{u}_r) = f(\alpha_1 \mathbf{u}_1 + \cdots + \alpha_r \mathbf{u}_r) = \mathbf{0}.$$

Como f es inyectiva, $\alpha_1 \mathbf{u}_1 + \cdots + \alpha_r \mathbf{u}_r = \mathbf{0}$, y como S es un sistema libre, todos los coeficientes son cero. Esto termina la demostración. $\square$

4.4. Isomorfismos.

Ya hemos estudiado un poco las aplicaciones inyectivas. Otros tipos de aplicaciones lineales son los siguientes:

Endomorfismo: Es un homomorfismo $f \in Hom(V, V)$ (de V en sí mismo). Al conjunto de los endomorfismos de V se le suele denotar: $\mathrm{End}(V) = \mathrm{Hom}(V, V)$.
Isomorfismo: Es un homomorfismo $f \in Hom(V, V')$ *biyectivo* (es decir, inyectivo y sobreyectivo).
Automorfismo: Es un endomorfismo biyectivo. Es decir, un isomorfismo de V en sí mismo.

Por ejemplo, la aplicación identidad $\mathrm{id} \in \mathrm{Hom}(V, V)$ es un automorfismo. Pero la aplicación nula $\mathcal{O} \in Hom(V, V')$ sólo sería un automorfismo si $V = V' = \{\mathbf{0}\}$.

Espacios isomorfos: Dos espacios vectoriales V y V' se dicen **isomorfos** si existe un isomorfismo $f \in \mathrm{Hom}(V, V')$.

El concepto de espacios isomorfos es muy importante. Si dos espacios son isomorfos, todas las propiedades que demostremos para uno de ellos (usando las propiedades de los espacios vectoriales), son válidas para el otro. Por tanto, si nos sentimos más cómodos trabajando con uno de ellos, podemos hacerlo sin ningún problema. Esto es lo que hicimos en el tema anterior cuando definimos las *coordenadas* de un vector: definimos un isomorfismo $\mathcal{C}_B : V \to K^n$. Es decir, demostramos que todo espacio vectorial V de dimensión n sobre un cuerpo K **es isomorfo** a K^n.

Algunas propiedades de los isomorfismos son las siguientes:

Proposición 4.12 *Se tienen las siguientes propiedades:*

1. *La composición de dos isomorfismos es un isomorfismo.*
2. $f \in \mathrm{Hom}(V, V')$ *es un isomorfismo si y sólo si* $\ker(f) = \{\mathbf{0}\}$ *e* $\mathrm{Im}(f) = V'$.
3. *Si* V *es de dimensión finita,* $f \in \mathrm{Hom}(V, V')$ *es un isomorfismo si y sólo si* $\dim V = \dim f(V) = \dim V'$.
4. *Si* V *es de dimensión finita,* $f \in End(V)$ *es un automorfismo si y sólo y es inyectiva. Y esto ocurre si y sólo si* f *es sobreyectiva.*
5. *Si* $f \in \mathrm{Hom}(V, V')$ *es un isomorfismo, entonces la aplicación inversa* $f^{-1} : V' \to V$ *es también un isomorfismo.*
6. *Dos espacios vectoriales de dimensión finita, sobre un mismo cuerpo, son isomorfos si y sólo si tienen la misma dimensión.*

Demostración:

1. Esto es consecuencia de que la composición de dos funciones biyectivas es una función biyectiva.
2. f es inyectiva si y sólo si $\ker(f) = \{\mathbf{0}\}$, y es sobreyectiva si y sólo si $\operatorname{Im}(f) = V'$, por tanto, será biyectiva si y sólo si las dos condiciones son ciertas.
3. Análogo a lo anterior: f es inyectiva si y sólo si $\dim V = \dim f(V)$, y es sobreyectiva si y sólo si $f(V) = V'$, es decir, si y sólo si $\dim f(V) = \dim V'$.
4. Esto es consecuencia de la propiedad anterior, ya que en este caso $V' = V$, y tanto la inyectividad como la sobreyectividad de f son equivalentes a $\dim V = \dim f(V)$.
5. Si f es un isomorfismo, es decir, una aplicación biyectiva, podemos definir su inversa, que también será biyectiva. Sólo tenemos que demostrar, entonces, que f^{-1} es una aplicación lineal. Sean $\mathbf{u}', \mathbf{v}' \in V'$, y sea $\alpha \in K$. Como f es biyectiva, existen $\mathbf{u}, \mathbf{v} \in V$ tales que $f(\mathbf{u}) = \mathbf{u}'$ y $f(\mathbf{v}) = \mathbf{v}'$. Entonces se tiene:

   $$f^{-1}(\mathbf{u}' + \mathbf{v}') = f^{-1}(f(\mathbf{u}) + f(\mathbf{v})) = f^{-1}f(\mathbf{u} + \mathbf{v}) = \mathbf{u} + \mathbf{v} = f^{-1}(\mathbf{u}') + f^{-1}(\mathbf{v}').$$

   Por otra parte,

   $$f^{-1}(\alpha\mathbf{u}') = f^{-1}(\alpha f(\mathbf{u})) = f^{-1}(f(\alpha\mathbf{u})) = \alpha\mathbf{u} = \alpha f^{-1}(\mathbf{u}').$$

   Por tanto, f^{-1} es una aplicación lineal, y como es biyectiva, es un isomorfismo.
6. Si V tiene dimensión n, tomemos cualquier base B de V. Sabemos que existe una aplicación lineal biyectiva $\mathcal{C}_B : V \to K^n$, es decir, V es isomorfo a K^n. Si V' es otro espacio vectorial de dimensión n, tomamos una base B' y consideramos el isomorfismo $\mathcal{C}_{B'} : V' \to K^n$. Entonces la aplicación lineal $\mathcal{C}_{B'}^{-1}\mathcal{C}_B$ es un isomorfismo de V en V'.

   Visto de otra forma, como ser isomorfo es una relación de equivalencia, y todos los espacios vectoriales de dimensión n son isomorfos a K^n, todos ellos son isomorfos entre sí.

□

4.5. Aplicaciones lineales y matrices I.

Hasta ahora hemos visto las definiciones y algunas propiedades de las aplicaciones lineales. Pero, ¿cómo podemos definirlas? Es decir, como V tiene un número infinito de vectores,

¿Hay que saber la imagen de cada uno de ellos para saber cómo es una aplicación lineal? Veamos que, afortunadamente, esto no es necesario: nos basta conocer la imagen de los elementos de una base de V.

Proposición 4.13 *Sean V y V' dos espacios vectoriales sobre un mismo cuerpo K. Sea $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ es una base de V, y sea $S = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_1\}$ un sistema cualquiera de vectores de V'. Entonces existe una única aplicación lineal $f \in \mathrm{Hom}(V, V')$ tal que $f(\mathbf{v}_i) = \mathbf{v}'_i$ para $i = 1, \ldots, n$.*

Demostración: La aplicación f se define de la siguiente manera: Dado cualquier vector $\mathbf{v} \in V$, se puede escribir de manera única como $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$. Definimos entonces: $f(\mathbf{v}) = \alpha_1 f(\mathbf{v}_1) + \cdots + \alpha_n f(\mathbf{v}_n) = \alpha_1\mathbf{v}'_1 + \cdots + \alpha_n\mathbf{v}'_n$. Esta aplicación está bien definida, ya que los coeficientes $\alpha_1, \ldots, \alpha_n$ están unívocamente determinados por $\mathbf{v}$.

Se demuestra que f es aplicación lineal de forma directa.

Por último, si existiera otra aplicación lineal g, tal que $g(\mathbf{v}_i) = \mathbf{v}'_i$ para $i = 1, \ldots, n$, entonces la imagen de un vector $\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$ sería $g(\mathbf{v}) = g(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 g(\mathbf{v}_1) + \cdots + \alpha_n g(\mathbf{v}_n) = \alpha_1\mathbf{v}'_1 + \cdots + \alpha_n v'_n = f(\mathbf{v})$. Por tanto, como $\mathbf{v}$ es un vector cualquiera, la aplicación g es igual a f. $\square$

Acabamos de demostrar, por tanto, que para conocer cómo es una aplicación lineal, basta conocer las imágenes de los elementos de una base. Observemos que, por la unicidad demostrada en el resultado anterior, toda aplicación lineal es de la forma descrita: si sabemos las imágenes por f de los elementos de una base, entonces f tiene que ser obligatoriamente la función definida arriba.

Ahora veamos que, si el espacio de llegada V' también es de dimensión finita, entonces las aplicaciones lineales se describen todavía más fácilmente.

Proposición 4.14 *Sean V y V' dos espacios vectoriales, sobre un mismo cuerpo K, de dimensiones n y m respectivamente. Sea B una base de V, sea B' una base de V', y consideremos una aplicación lineal $f \in \mathrm{Hom}(V, V')$.*
Sea $A \in \mathcal{M}_{m\times n}$ la matriz cuya columnas representan los vectores de $f(B)$ (es decir, sus coordenadas respecto de la base B'). Entonces, dado un vector cualquiera $\mathbf{v} \in V$, con coordenadas $\mathbf{v}_B = (x_1, \ldots, x_n)$, las coordenadas de $f(\mathbf{v})$, que denotaremos $f(\mathbf{v})_{B'} = (y_1, \ldots, y_m)$,

están determinadas por:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix},$$

es decir: $f(\mathbf{v})_{B'} = A\,\mathbf{v}_B$.

Demostración: Si $B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ y $B' = (\mathbf{v}'_1, \ldots, \mathbf{v}'_m)$, sabemos, por construcción, que $f(\mathbf{v}_i) = a_{1i}\mathbf{v}'_1 + \cdots + a_{mi}\mathbf{v}'_m$. Entonces, dado un vector $\mathbf{v} \in V$, con coordenadas $\mathbf{v}_B = (x_1, \ldots, x_n)$, se tiene $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$. Por tanto,

$$f(\mathbf{v}) = x_1 f(\mathbf{v}_1) + \cdots + x_n f(\mathbf{v}_n) = x_1(a_{11}\mathbf{v}'_1 + \cdots + a_{m1}\mathbf{v}'_m) + \cdots + x_n(a_{1n}\mathbf{v}'_1 + \cdots + a_{mn}\mathbf{v}'_m).$$

Agrupando los coeficientes de cada $\mathbf{v}'_i$, tenemos:

$$f(\mathbf{v}) = (a_{11}x_1 + \cdots + a_{1n}x_n)\mathbf{v}'_1 + \cdots + (a_{m1}x_1 + \cdots + a_{mn}x_n)\mathbf{v}'_m.$$

Por tanto, si $f(\mathbf{v})_{B'} = (y_1, \ldots, y_m)$, tendremos

$$y_i = a_{i1}x_1 + \cdots + a_{in}x_n,$$

que es lo que queríamos demostrar. □

Acabamos de demostrar que una aplicación lineal, entre un espacio de dimensión n y un espacio de dimensión m (sobre un mismo cuerpo), está completamente determinada por una matriz $m \times n$. Y a la inversa: toda matriz $m \times n$ determina una aplicación lineal. Por tanto, hemos demostrado lo siguiente:

Corolario 4.15 *Dados dos espacios vectoriales V y V', de dimensiones m y n, sobre el mismo cuerpo K, existe una biyección* $M : \mathrm{Hom}(V, V') \to \mathcal{M}_{m\times n}$.

Demostración: Basta fijar una base B de V y una base B' de V', y asociar a cada aplicación lineal $f \in \mathrm{Hom}(V, V')$ la matriz $M(f)$ definida en la proposición anterior. □

Otra consecuencia importante de la proposición anterior es la relación entre la composición de aplicaciones lineales y el producto de matrices:

Corolario 4.16 *Sean V, V' y V'' tres espacios vectoriales de dimensión finita, donde fijamos tres bases, B, B' y B'', respectivamente. Sean $f \in \mathrm{Hom}(V, V')$, $g \in \mathrm{Hom}(V', V'')$, y sean $M(f)$ y $M(g)$ sus matrices correspondientes. Entonces la matriz correspondiente a $g \circ f \in \mathrm{Hom}(V, V'')$ es:*

$$M(g \circ f) = M(g)M(f).$$

Demostración: Si $\mathbf{v} \in V$, sabemos que $f(\mathbf{v}) = M(f)\mathbf{v}$, y dado $\mathbf{v}' \in V'$ sabemos que $g(\mathbf{v}') = M(g)\mathbf{v}'$. Por tanto,

$$g \circ f(\mathbf{v}) = g(f(\mathbf{v})) = g(M(f)\mathbf{v}) = M(g)M(f)\mathbf{v},$$

por lo que la matriz de $g \circ f$ es la matriz producto $M(g)M(f)$, como queríamos demostrar.
$\square$

4.6. Aplicaciones lineales y matrices II.

La relación entre las aplicaciones lineales y las matrices va más allá de la mera fórmula para describir coordenadas. La mayoría de las propiedades que conocemos sobre las matrices, tienen sentido al hablar de las aplicaciones lineales, y nos ayudarán a estudiar estas últimas. Recordemos, por ejemplo, que el rango de una aplicación lineal f es la dimensión de $\mathrm{Im}(f)$. Se tiene:

Proposición 4.17 *Sea $f \in \mathrm{Hom}(V, V')$, donde V y V' tienen dimensión n y m respectivamente. Sea $M(f) \in \mathcal{M}_{m\times n}$ la matriz asociada a f respecto de dos bases cualesquiera de V y V'. Entonces se tiene:*

1. *El rango de f es igual al rango de $M(f)$.*
2. *f es inyectiva si y sólo si $\mathrm{rg}(M(f)) = n$.*
3. *f es sobreyectiva si y sólo si $\mathrm{rg}(M(f)) = m$.*
4. *f es un isomorfismo si y sólo si $M(f)$ es cuadrada y no singular.*

Demostración: La primera propiedad se demuestra como sigue: si tomamos una base cualquiera $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V, el rango de f es igual a la dimensión de la variedad generada por $\{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)\}$. Pero las columnas de la matriz $M(f)$ representan a estos n vectores, luego esta dimensión es igual al rango de $M(f)$, como queríamos demostrar.

Por otra parte, el rango de f es la dimensión de $f(V)$, y sabemos que f es inyectiva si y sólo si esta dimensión es igual a la de V. Es decir, si y sólo si $\mathrm{rg}(M(f)) = n$. Esto demuestra la segunda propiedad.

Demostramos la tercera como sigue: f será sobreyectiva si y sólo si los vectores $\{f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)\}$ generan un espacio de dimensión m (la dimensión de V'). Pero esto pasa si y sólo si en las columnas de V hay m vectores linealmente independientes. Es decir, si $\mathrm{rg}(M(f)) = m$.

Para demostrar la cuarta condición, recordemos que V y V' sólo pueden ser isomorfos (y lo son) si tienen la misma dimensión. Por tanto, supondremos que $\dim V = \dim V' = n$, luego $M(f)$ será una matriz cuadrada. Hay que demostrar que f es un isomorfismo si y sólo si $M(f)$ es no singular. Pero por las dos propiedades anteriores, f es isomorfismo, es decir, f es biyectiva, si y sólo si $\mathrm{rg}(M(f)) = n$, esto es, si y sólo si $M(f)$ es no singular. $\square$

Una consecuencia evidente de este resultado es la siguiente:

- Si $n < m$, entonces f podría ser inyectiva, pero nunca podría ser sobreyectiva.
- Si $n > m$, entonces f podría ser sobreyectiva, pero nunca podría ser inyectiva.
- Si $n = m$, entonces f es inyectiva si y sólo si es sobreyectiva.

En este último caso, sabemos que f admite una función inversa f^{-1}, que es también una aplicación lineal. Por supuesto, la matriz de f^{-1} es la inversa de la matriz de f:

Proposición 4.18 *Sean V y V' dos espacios vectoriales isomorfos, de dimensión finita. Sea $f \in \mathrm{Hom}(V, V')$ un isomorfismo, y sea $M(f)$ su matriz asociada (respecto de dos bases fijadas). Entonces la matriz de f^{-1} es: $M(f^{-1}) = M(f)^{-1}$.*

Demostración: Sabemos que si f es un isomorfismo, entonces $M(f)$ es no singular, y por tanto existe su matriz inversa $M(f)^{-1}$. Pero si $f(\mathbf{v}) = M(f)\mathbf{v}$, entonces $\mathbf{v} = M(f)^{-1}f(\mathbf{v})$, para cualquier $\mathbf{v} \in V$. Es decir, $f^{-1}(\mathbf{v}') = M(f)^{-1}\mathbf{v}'$, para cualquier $\mathbf{v}' \in V'$. Esto quiere decir que la matriz de f^{-1} es $M(f)^{-1}$. $\square$

Si estudiamos aplicaciones lineales usando matrices, también podemos calcular los elementos principales de una aplicación, como su núcleo o su imagen.

Proposición 4.19 *Sea* $f \in \operatorname{Hom}(V, V')$*. Fijemos dos bases de* V *y* V'*, y sea* $M(f)$ *la matriz de* f *respecto de estas dos bases. Sea* $\mathbf{x}$ *un vector que representa las coordenadas de un elemento cualquiera de* V*. Entonces:*

1. *Las columnas de* $M(f)$ *son un sistema de generadores de* $\operatorname{Im}(f)$*.*
2. $M(f)\mathbf{x} = \mathbf{0}$ *son unas ecuaciones implícitas de* $\ker(f)$*.*

Demostración: Las dos propiedades son una aplicación directa de las definiciones. $\square$

Algo análogo podemos hacer para las variedades lineales de V y V':

Proposición 4.20 *Sea* $f \in \operatorname{Hom}(V, V')$ *y* $M(f)$ *la matriz de* f*, como antes. Sean* L *y* L' *variedades lineales cualesquiera de* V *y* V'*, respectivamente. Supongamos que* A *es una matriz cuyas* **columnas** *forman un sistema de generadores de* L*, y que* $B\mathbf{x}' = \mathbf{0}$ *son unas ecuaciones implícitas de* L'*. Entonces:*

1. *Las columnas de* $M(f)A$ *forman un sistema de generadores de* $f(L)$*.*
2. $BM(f)\,\mathbf{x} = \mathbf{0}$ *son unas ecuaciones implícitas de* $f^{-1}(L)$*.*

Demostración: Las dos propiedades son consecuencia de la fórmula $\mathbf{x}' = M(f)\mathbf{x}$, que relaciona las coordenadas de un vector $\mathbf{v} \in V$ con las del vector $f(\mathbf{v}) \in V$. $\square$

4.7. Primer teorema de isomorfía.

Una de las aplicaciones más importantes de la relación entre las aplicaciones lineales y las matrices, es el estudio de los sistemas de ecuaciones lineales. Recordemos que un sistema lineal puede escribirse de forma matricial:

$$A\mathbf{x} = \mathbf{b}.$$

Pero ahora sabemos que toda matriz A puede verse como la matriz de una aplicación lineal, es decir, podemos considerar que $A = M(f)$, para una cierta aplicación lineal f. Pero entonces el sistema anterior se lee:

$$f(\mathbf{x}) = \mathbf{b}.$$

Como $\mathbf{x}$ es el vector incógnita, resolver el sistema consiste en encontrar los vectores cuya imagen por f sea $\mathbf{b}$. Es decir, la solución del sistema es exactamente la variedad lineal afín $f^{-1}(\mathbf{b})$. Con este razonamiento tan sencillo, hemos demostrado un resultado importante:

Teorema 4.21 *El conjunto de soluciones de un sistema lineal es una variedad lineal afín.*

Pero podemos decir todavía más. Sabemos que $f^{-1}(\mathbf{b})$ es igual a $\mathbf{v} + \ker(f)$, donde $\mathbf{v}$ es cualquier vector tal que $f(\mathbf{v}) = \mathbf{b}$. Es decir, si conocemos una sola solución, $\mathbf{v}$, del sistema, entonces obtenemos *todas* las soluciones sumándole los elementos de $\ker(f)$. Sabemos que un vector $\mathbf{v}_0$ pertenece a $\ker(f)$ si y sólo si $M(f)\mathbf{v}_0 = \mathbf{0}$. En nuestro caso, $A\mathbf{v}_0 = \mathbf{0}$. ¡Pero este es el sistema homogéneo asociado al sistema de partida! (es decir, el que resulta al hacer cero todos los términos independientes). En efecto, las soluciones del sistema homogéneo $A\mathbf{x} = \mathbf{0}$ forman la variedad lineal $\ker(f)$. Por tanto, hemos demostrado el siguiente resultado.

Teorema 4.22 *Consideremos un sistema de ecuaciones lineales completo, $A\mathbf{x} = \mathbf{b}$, y su sistema homogéneo asociado $A\mathbf{x} = \mathbf{0}$. Supongamos que conocemos una solución particular $\mathbf{v}$ del sistema completo, y la variedad lineal L de soluciones del sistema homogéneo. Entonces la solución general del sistema completo es la variedad lineal afín $\mathbf{v} + L$.*

Por tanto, cuando tengamos un sistema compatible indeterminado con matriz de coeficientes $A \in \mathcal{M}_{m\times n}$, no sólo sabemos que tiene infinitas soluciones, sino que éstas forman una variedad lineal afín, cuya variedad de dirección ($\ker(f)$) tiene dimensión $n - \mathrm{rg}(A)$.

Terminaremos esta sección con un teorema importante, llamado *primer teorema de isomorfía*, que puede dar una idea más precisa de cómo son las aplicaciones lineales, sus núcleos y sus imágenes. Recordemos que si L es una variedad lineal de una espacio vectorial V, entonces podemos considerar el espacio cociente V/L. Si tenemos una aplicación $f \in \mathrm{Hom}(V, V')$, su núcleo, $\ker(f)$ es una variedad lineal de V, por tanto, podremos considerar el espacio cociente $V/\ker(f)$. En teorema es el siguiente:

Teorema 4.23 (Primer teorema de isomorfía) *Dada $f \in \mathrm{Hom}(V, V')$, el espacio cociente $V/\ker(f)$ es isomorfo a $\mathrm{Im}(f)$. Un isomorfismo entre estos dos espacios es el siguiente:*

$$\varphi : V/\ker(f) \longrightarrow \mathrm{Im}(f),$$

definido por $\varphi(\mathbf{v} + \ker(f)) = f(\mathbf{v})$.

DEMOSTRACIÓN: Primero hay que demostrar que φ está bien definida, es decir, que si $\mathbf{v}_1 + \ker(f) = \mathbf{v}_2 + \ker(f)$, entonces $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Pero $\mathbf{v}_1 + \ker(f) = \mathbf{v}_2 + \ker(f)$ si y sólo si $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(f)$, es decir, $f(\mathbf{v}_1 - \mathbf{v}_2) = f(\mathbf{v}_1) - f(\mathbf{v}_2) = 0$, como queríamos probar.

Después, se demuestra de forma directa que φ es una aplicación lineal. Por último, para demostrar que es isomorfismo sólo hay que probar que es inyectiva y sobreyectiva, lo cual también se hace de forma directa. $\square$

4.8. Cambio de base. Matrices equivalentes.

Hasta ahora hemos relacionado las aplicaciones lineales y las matrices, fijando una base en el espacio de partida, y otra en el espacio de llegada. Pero esta elección no es, evidentemente, única: una misma aplicación lineal puede estar representada por distintas matrices, dependiendo de las bases respecto de las que estén definidas.

Recordemos que si V es un espacio vectorial de dimensión n, y consideramos dos bases B_1 y B_2 de V, se define la *matriz del cambio de base* como la matriz $A_{B_1,B_2} \in \mathcal{M}_{n\times n}$, donde las columnas de A_{B_1,B_2} representan los elementos de B_1 respecto de la base B_2. Esta matriz transforma coordenadas respecto de B_1 en coordenadas respecto de B_2, por multiplicación a izquierda:

$$(A_{B_1,B_2})\mathbf{v}_{B_1} = \mathbf{v}_{B_2}.$$

Ahora consideremos una aplicación lineal $f \in \mathrm{Hom}(V, V')$, donde V tiene dimensión n y V' tiene dimensión m. Si fijamos una base B_1 de V y una base B_1' de V', obtendremos una matriz $M(f) = M(f)_{B_1,B_1'}$. Pero si hubiéramos fijado otra base B_2 de V, y también otra base B_2' de V', habríamos obtenido otra matriz para f, que llamaremos $M(f)_{B_2,B_2'}$. Nos interesa saber cuál es la relación entre estas dos matrices. Como cabe esperar, podremos pasar de una a otra multiplicando por las matrices de cambio de base:

Proposición 4.24 *Con las notaciones anteriores, se tiene:*

$$M(f)_{B_2,B_2'} = (A_{B_1',B_2'})\,(M(f)_{B_1,B_1'})\,(A_{B_2,B_1}).$$

DEMOSTRACIÓN: Respecto a las bases B_1 y B_1' tenemos, para cualquier vector $\mathbf{v} \in V$,

$$(M(f)_{B_1,B_1'})\,\mathbf{v}_{B_1} = f(\mathbf{v})_{B_1'},$$

Veamos que la matriz del enunciado transforma cualquier vector $\mathbf{v}_{B_2}$ en su imagen $f(\mathbf{v})_{B'_2}$, y así habremos demostrado que es igual a la matriz $M(f)_{B_2,B'_2}$. Se tiene:

$$(A_{B'_1,B'_2})\,(M(f)_{B_1,B'_1})\,\underbrace{(A_{B_2,B_1})\,\mathbf{v}_{B_2}} = (A_{B'_1,B'_2})\,\overbrace{(M(f)_{B_1,B'_1})\,\mathbf{v}_{B_1}} = (A_{B'_1,B'_2})\,f(\mathbf{v})_{B'_1} = f(\mathbf{v})_{B'_2},$$

luego el resultado es cierto. $\square$

Esto nos lleva a la siguiente definición:

Matrices equivalentes: Dos matrices $A, B \in \mathcal{M}_{m\times n}$ se dicen **equivalentes** si existen dos matrices invertibles $P \in \mathcal{M}_{m\times m}$ y $Q \in \mathcal{M}_{n\times n}$ tales que

$$PAQ = B.$$

Proposición 4.25 *Sean V y V' dos espacios vectoriales de dimensiones respectivas m y n sobre un cuerpo K. Dos matrices $A, B \in \mathcal{M}_{m\times n}(K)$ son equivalentes si y sólo si son las matrices de una misma aplicación lineal $f \in \mathrm{Hom}(V, V')$, respecto de distintas bases.*

Demostración: Supongamos que las matrices son equivalentes. Entonces existen matrices invertibles P y Q tales que $PAQ = B$. Fijemos una base B_1 de V y una base B'_1 de V'. entonces la matriz A representa a una aplicación lineal $f \in \mathrm{Hom}(V, V')$. Sea B_2 la base de V cuyos elementos son las columnas de la matriz Q, y sea B'_2 la base de V' suyos elementos son las columnas de la matriz P^{-1}. Sabemos que B_2 y B'_2 son bases porque P y Q son invertibles. Entonces $P = A_{B'_1,B'_2}$ y $Q = A_{B_2,B_1}$. Por tanto, la proposición anterior nos dice que PAQ, es decir B, es la matriz $M(f)_{B_2,B'_2}$. Por tanto A y B son dos matrices que representan a la misma aplicación lineal f.

Recíprocamente, supongamos que $A = M(f)_{B_1,B'_1}$ y $B = M(f)_{B_2,B'_2}$ para una cierta aplicación lineal f, y unas bases B_1, B_2 de V, y B'_1, B'_2 de V'. Para probar que A y B son equivalentes basta tomar las matrices de cambio de base: $P = A_{B'_1,B'_2}$ y $Q = A_{B_2,B_1}$. $\square$

Veamos que la palabra *equivalente* no ha sido escogida al azar:

Proposición 4.26 *La equivalencia de matrices es una relación de equivalencia.*

DEMOSTRACIÓN: Las propiedades simétrica, reflexiva y transitiva se demuestran de forma directa. $\square$

Nos podemos preguntar ahora si habrá muchas clases de equivalencias de matrices, es decir, si podremos encontrar muchas matrices $m \times n$ que no sean equivalentes dos a dos. La respuesta es que no, ya que la clase de equivalencia de una matriz sólo depende de su rango:

Proposición 4.27 *Toda matriz $A \in \mathcal{M}_{m\times n}$ de rango r es equivalente a la matriz C_r, que tiene la forma:*

$$C_r = \left(\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right),$$

donde I_r es la matriz identidad de orden r, y O denota a matrices nulas de las dimensiones requeridas.

DEMOSTRACIÓN: Comencemos con una matriz A de rango r. Haciendo transformaciones elementales de filas, obtenemos su reducida por filas, A', y una matriz invertible P tal que $A' = PA$. Como A tiene rango r, A' tendrá r filas distintas de cero. Ahora aplicamos a A' transformaciones elementales de columnas, y la transformamos en A'', su escalonada por columnas, obteniendo una matriz invertible Q tal que $A'' = A'Q = PAQ$. Pero la escalonada por columnas de A' debe tener sólo r columnas distintas de cero, y como sólo tiene r filas distintas de cero, en esas primeras r filas deben estar todos los pivotes. Es decir, $A'' = C_r$, luego $C_r = PAQ$, y así A y C_r son equivalentes, como queríamos demostrar. $\square$

Corolario 4.28 *Dos matrices de las mismas dimensiones son equivalentes si y sólo si tienen el mismo rango.*

4.9. Endomorfismos. Matrices semejantes.

Para la definición de equivalencia de matrices, hemos considerado dos espacios vectoriales V y V', y las aplicaciones lineales entre ellos. Un caso particular importante se da cuando $V = V'$, es decir, cuando estudiamos endomorfismos de V. Si estudiamos este caso igual que el caso general, estamos permitiendo que una aplicación $f \in \mathrm{End}(V)$ tome vectores de V respecto de una base, y los envíe a vectores de V respecto de *otra base*.

Pero normalmente, cuando trabajamos en un espacio V fijo, se supone que fijamos una base B, y que tanto el vector $\mathbf{v}$ como su imagen $f(\mathbf{v})$ deben estar representados respecto de la misma base. En este caso, el cambio de base de V cambiará la matriz $M(f)$ de la siguiente manera:

Proposición 4.29 *Sea V un espacio vectorial de dimensión n. Sea $f \in \mathrm{End}(V)$. Si B_1 y B_2 son dos bases de V, y las matrices de f con respecto a estas bases son, respectivamente, $M(f)_{B_1}$ y $M(f)_{B_2}$, entonces se tiene:*

$$M(f)_{B_2} = A_{B_1,B_2}\, M(f)_{B_1}\, A_{B_2,B_1}.$$

DEMOSTRACIÓN: Esta es la fórmula ya conocida para el cambio de bases en las aplicaciones lineales entre dos espacios, si nos damos cuenta de que $M(f)_{B_1} = M(f)_{B_1,B_1}$ y $M(f)_{B_2} = M(f)_{B_2,B_2}$. $\square$

Recordemos que $A_{B_1,B_2} = A^{-1}_{B_2,B_1}$. Esto nos da lugar a la siguiente definición:

Matrices semejantes: Dos matrices cuadradas $A, B \in \mathcal{M}_{n\times n}$ se dicen **semejantes** si existe una matriz invertible P tal que

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

Las matrices semejantes son a los endomorfismos lo que las matrices equivalentes eran a los homomorfismos:

Proposición 4.30 *Sea V un espacio vectorial de dimensión n sobre un cuerpo K. Dos matrices $A, B \in \mathcal{M}_{n\times n}(K)$ son semejantes si y sólo si son las matrices de una misma aplicación lineal $f \in \mathrm{End}(V)$, respecto de distintas bases.*

DEMOSTRACIÓN: La demostración es análoga al resultado correspondiente para matrices equivalentes. $\square$

Proposición 4.31 *La semejanza de matrices es una relación de equivalencia.*

DEMOSTRACIÓN: Directa. $\square$

En el caso de matrices equivalentes, vimos que el rango de una matriz determinaba su clase de equivalencia. Para matrices semejantes no es tan fácil. Sin embargo, tenemos un invariante para matrices semejantes que ya conocemos:

Proposición 4.32 *Si dos matrices $A, B \in \mathcal{M}_{n\times n}$ son semejantes, entonces se tiene* $\det(A) = \det(B)$.

Demostración: Si A y B son semejantes, existe una matriz invertible P tal que $B = P^{-1}AP$. Entonces:

$$\det(B) = \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \frac{1}{\det(P)}\det(A)\det(P) = \det(A).$$

□

Hemos demostrado entonces que la siguiente definición tiene sentido:

Determinante de un endomorfismo: Sea V un espacio vectorial de dimensión finita, y sea $f \in \mathrm{End}(V)$. Se define el **determinante** de f, denotado $\det f$, como el determinante de la matriz $M(f)$ respecto de cualquier base de V.

4.10. El espacio vectorial $\mathrm{Hom}(V, V')$.

Llevamos todo este tema estudiando las propiedades de las aplicaciones lineales entre espacios vectoriales. Hemos visto, sobre todo, que si los espacios V y V' son de dimensiones finitas, n y m, entonces las aplicaciones de $\mathrm{Hom}(V, V')$ se pueden identificar con las matrices de $\mathcal{M}_{m\times n}$.

Pero en $\mathcal{M}_{m\times n}$ hay más estructura que la de un simple conjunto. Sabemos que dos matrices de $\mathcal{M}_{m\times n}$ se pueden sumar, y una matriz se puede multiplicar por un escalar. Además, vimos que estas dos operaciones dotan a $\mathcal{M}_{m\times n}$ de estructura de espacio vectorial. Pues bien, la correspondencia entre $\mathrm{Hom}(V, V')$ y $\mathcal{M}_{m\times n}$ llega hasta ese punto: $\mathrm{Hom}(V, V')$ también tiene estructura de espacio vectorial, y este espacio es isomorfo a $\mathcal{M}_{m\times n}$.

Operaciones con aplicaciones lineales: Sean V y V' dos espacios vectoriales sobre un cuerpo K, y sea $\mathrm{Hom}(V, V')$ el conjunto de las aplicaciones lineales de V en V'. Dadas $f, g \in Hom(V, V')$ y $\alpha \in K$, se definen las siguientes operaciones:

- **Suma de aplicaciones:** La aplicación $f + g$ está definida por

$$(f + g)(\mathbf{v}) = f(\mathbf{v}) + g(\mathbf{v}).$$

- **Producto de aplicación por escalar:** La aplicación αf está definida por

$$(\alpha f)(\mathbf{v}) = \alpha f(\mathbf{v}).$$

Proposición 4.33 *Si $f, g \in \mathrm{Hom}(V, V')$ y $\alpha \in K$, las aplicaciones $f + g$ y αf definidas arriba son aplicaciones lineales.*

Demostración: Directa. □

Teorema 4.34 *Con las operaciones anteriores, $\mathrm{Hom}(V, V')$ tiene estructura de espacio vectorial. Y si V y V' tienen dimensiones respectivas n y m, entonces $\mathrm{Hom}(V, V')$ es isomorfo a $\mathcal{M}_{m\times n}$.*

Demostración: Se comprueba de forma directa que $\mathrm{Hom}(V, V')$ es espacio vectorial.

Para ver que es isomorfo a $\mathcal{M}_{m\times n}$, se prueba directamente que la aplicación $M : \mathrm{Hom}(V, V') \to \mathcal{M}_{m\times n}$, que a cada aplicación f le asocia su matriz correspondiente $M(f)$ respecto de dos bases fijadas de V y V', es un isomorfismo. □

Veamos además que estos dos espacios son de tipo finito.

Proposición 4.35 *El espacio vectorial $\mathcal{M}_{m\times n}(K)$ tiene dimensión mn.*

Demostración: Para demostrar este resultado daremos una base de $\mathcal{M}_{m\times n}$. Dado $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$, definimos la matriz E_{ij} como aquella cuyas entradas son todas nulas salvo un 1 en la posición (i, j). Veamos que $B = \{E_{ij}; \quad i = 1, \ldots, m, \ j = 1, \ldots, n\}$ es una base de $\mathcal{M}_{m\times n}$.

En primer lugar, dada una matriz cualquiera $A \in \mathcal{M}_{m\times n}$, cuyas entradas denotaremos por a_{ij}, se puede escribir claramente como combinación lineal de las matrices de B:

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}.$$

Por tanto B es sistema de generadores de $\mathcal{M}_{m\times n}$. Para ver que B es un sistema libre, recordemos que la matriz nula, $\mathcal{O}$, es el elemento neutro de la suma de matrices. Si tenemos una combinación lineal:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \alpha_{ij}E_{ij} = \mathcal{O},$$

en el primer término de la igualdad tenemos la matriz cuya entrada (i, j) es α_{ij}, y en el segundo término la matriz nula. Esto implica que cada α_{ij} es nulo, luego B es un sistema de generadores libre, es decir, una base de $\mathcal{M}_{m\times n}$. Como en B hay exactamente mn elementos, se obtiene el resultado. $\square$

Corolario 4.36 *Si V y V' son dos espacios vectoriales sobre K, con dimensiones respectivas m y n, el espacio vectorial* $\mathrm{Hom}(V, V')$ *tiene dimensión mn.*

DEMOSTRACIÓN: Esto es consecuencia del resultado anterior, ya que $\mathrm{Hom}(V, V')$ es isomorfo a $\mathcal{M}_{m\times n}$. $\square$

Nota: En realidad hay más estructura en común entre $\mathrm{Hom}(V, V')$ y $\mathcal{M}_{m\times n}$, ya que hemos visto que la composición de aplicaciones equivale al producto de matrices. Pero dos matrices sólo se pueden multiplicar si tienen las dimensiones adecuadas. Si queremos multiplicar sin problemas, podemos restringirnos a los endomorfismos de un espacio V de dimensión n, ya que sus matrices asociadas son cuadradas. Esto nos dice que $\mathrm{End}(V)$ y $\mathcal{M}_{n\times n}$ son *anillos* isomorfos. Pero la definición de anillo se dará en otra asignatura.

Tema 5. Endomorfismos

5.1. Autovalores y autovectores.

En este tema seguiremos con el estudio de las aplicaciones lineales, pero en el caso en el que el espacio de partida es el mismo que el de llegada. Es decir, estudiaremos más a fondo los *endomorfismos* de un espacio vectorial V. Para simplificar el estudio, supondremos de ahora en adelante que V es de dimensión finita, n, y por tanto los endomorfismos de n se representan por matrices $n \times n$.

Recordemos que dos matrices $A, B \in \mathcal{M}_{n\times n}$ representan al mismo endomorfismo (respecto de distintas bases de V), si y sólo si son *semejantes*, esto es, si existe una matriz invertible P tal que $P^{-1}AP = B$. En ese caso, la matriz P es la matriz del cambio de base. Recordemos también que podemos saber si dos matrices son equivalentes simplemente mirando sus rangos, pero no conocemos (por ahora) ningún criterio para comprobar si dos matrices son semejantes. En este tema veremos el siguiente criterio: dos matrices son semejantes si y sólo si tienen la misma *forma canónica*.

La idea es la siguiente: nos interesa saber si, dado un endomorfismo $f \in \mathrm{End}(V)$, hay alguna base de V respecto de la cual la matriz $M(f)$ resulte ser lo más simple posible, y que además nos dé información sobre el comportamiento de f.

Vamos ya a adelantar algo: si la matriz de un endomorfismo f, respecto de una cierta base, es *diagonal*, entonces el endomorfismo actúa de una forma muy simple: Si los elementos de la diagonal son $d_1, d_2, \ldots, d_n$, entonces f transforma el vector de coordenadas $(x_1, \ldots, x_n)$ en el vector $(d_1x_1, d_2x_2, \ldots, d_nx_n)$. En otras palabras, si $B = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ es una base de V tal que $M(f)_B$ es diagonal, entonces f transforma el espacio V "expandiendo o contrayendo" cada vector, en la dirección de cada $\mathbf{e}_i$, por un factor d_i. Por tanto, si la matriz $M(f)$ es diagonal, sabemos perfectamente cómo actúa f, y es muy sencillo y efectivo hacer cálculos con f respecto de la base B.

Nos interesará, por tanto, si tenemos un endomorfismo cualquiera dado por una matriz $n \times n$, saber si existe un cambio de base que la transforme en diagonal. Para eso definimos los autovalores y autovectores:

Autovalores y autovectores de un endomorfismo: Sea $f \in \mathrm{End}(V)$. Se dice que un vector *no nulo* $\mathbf{v} \in V$ es un **autovector** de f si $f(\mathbf{v})$ es un múltiplo de $\mathbf{v}$. Es decir, $\mathbf{v}$ es autovector si $f(\mathbf{v}) = \lambda\mathbf{v}$ para un cierto escalar λ. En este caso, λ se llama **autovalor** de f, y se dice que $\mathbf{v}$ es un autovector asociado al autovalor λ.

Análogamente se definen los mismos conceptos para matrices:

Autovalores y autovectores de una matriz: Sea $A \in \mathcal{M}_{n\times n}$. Se dice que un vector *no nulo* $\mathbf{v} \in V$ es un **autovector** de A si $A\mathbf{v}$ es un múltiplo de $\mathbf{v}$. Es decir, $\mathbf{v}$ es autovector si $A\mathbf{v} = \lambda\mathbf{v}$ para un cierto escalar λ. En este caso, λ se llama **autovalor** de A, y se dice que $\mathbf{v}$ es un autovector asociado al autovalor λ.

Ejemplo 5.1 *Sea* $f \in \operatorname{End}(\mathbb{R}^3)$ *dado por la matriz diagonal:*

$$M(f) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -3 \end{pmatrix}.$$

Entonces el vector $(1,0,0)$ *es un autovector asociado al autovalor* 2*, ya que* $f(1,0,0) = (2,0,0)$*. Análogamente,* $(0,1,0)$ *es un autovector asociado al autovalor* 4*, y* $(0,0,1)$ *es un autovector asociado al autovalor* -3*.*

En este ejemplo vemos algo interesante: cuando la matriz es diagonal, los autovectores son precisamente los elementos de la base de V respecto de la cual la matriz está escrita, y los autovalores correspondientes son los elementos de la diagonal. Esto nos servirá más adelante para *diagonalizar* la matriz. Pero antes veamos cómo se pueden calcular los autovalores y autovectores de un endomorfismo o, análogamente, de una matriz, en el caso general.

A partir de ahora, I denotará a la matriz identidad de orden n. Se tiene:

Proposición 5.2 *Dada una matriz* $A \in \mathcal{M}_{n\times n}$*, los autovalores de* A *(o del endomorfismo que representa) son las soluciones de la ecuación dada por*

$$\det(A - \lambda I) = 0.$$

Demostración: Un escalar λ es un autovalor si y sólo si existe un autovector $\mathbf{v}$ tal que $A\mathbf{v} = \lambda\mathbf{v}$. Pero $I\mathbf{v} = \mathbf{v}$, luego esta expresión se puede transformar como sigue:

$$A\mathbf{v} = \lambda\mathbf{v} \quad\Leftrightarrow\quad A\mathbf{v} = \lambda I\mathbf{v} \quad\Leftrightarrow\quad A\mathbf{v} - \lambda I\mathbf{v} = \mathbf{0} \quad\Leftrightarrow\quad (A - \lambda I)\mathbf{v} = \mathbf{0}.$$

Pero esto último es un sistema lineal homogéneo, cuya matriz de coeficientes es $A - \lambda I \in \mathcal{M}_{n\times n}$. Entonces λ será un autovalor si y sólo si este sistema tiene solución no trivial. Como es un sistema homogéneo, tendrá solución no trivial si y sólo si su matriz de coeficientes tiene rango menor que n, es decir, si y sólo si $\det(A - \lambda I) = 0$. $\square$

Por tanto, si tenemos

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

entonces para hallar los autovalores de A hay que resolver:

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0.$$

Observemos que, al desarrollar este determinante, obtenemos un polinomio de grado n, en la variable λ. Las raíces de la ecuación que resulta al igualar este polinomio a cero, son los autovalores de A.

Polinomio y ecuación característica: Dada una matriz $A \in \mathcal{M}_{n\times n}$, llamamos **polinomio característico** de A al polinomio $|A - \lambda I|$. Y llamamos **ecuación característica** de A a la ecuación $|A - \lambda I| = 0$.

Por supuesto, el método para calcular los autovectores asociados a un autovalor fijado, λ_0, consiste en resolver el sistema lineal

$$(A - \lambda_0 I)\mathbf{x} = \mathbf{0}.$$

Cualquier solución de este sistema será un autovector asociado a λ_0. Por tanto, los autovectores asociados a un autovalor λ_0 forman una variedad lineal de V, concretamente $\ker(A - \lambda_0 I)$, cuya dimensión es exactamente $n - \mathrm{rg}(A - \lambda_0 I)$.

Subespacio propio: Dada una matriz A, y un autovalor λ de A, llamamos **subespacio propio** asociado a λ, al subespacio $V_1(\lambda) \subset V$ formado por los autovectores de A asociados a λ. Es decir, si llamamos f a la aplicación lineal determinada por A, y g a la aplicación lineal $f - \lambda\,\mathrm{id}$, determinada por $A - \lambda I$, entonces:

$$V_1(\lambda) = \ker(g).$$

Si el autovalor λ al que nos referimos está claro por el contexto, a veces escribiremos V_1 en vez de $V_1(\lambda)$.

Podríamos preguntarnos qué relación hay entre los subespacios $V_1(\lambda)$, para diferentes valores de λ. La respuesta la da el siguiente resultado:

Proposición 5.3 *Dada una matriz $A \in \mathcal{M}_{n\times n}$, y m autovalores distintos, $\lambda_1, \ldots, \lambda_m$ de A, los espacios $V_1(\lambda_1), \ldots, V_1(\lambda_m)$ son independientes. Es decir,*

$$V_1(\lambda_1) + \cdots + V_1(\lambda_m) = V_1(\lambda_1) \oplus \cdots \oplus V_1(\lambda_m).$$

DEMOSTRACIÓN: Supongamos que el resultado no es cierto. Existirán entonces unos vectores $\mathbf{v}_1 \in V_1(\lambda_1), \ldots, \mathbf{v}_m \in V_1(\lambda_m)$, no todos nulos, tales que $\mathbf{v}_1 + \cdots + \mathbf{v}_m = \mathbf{0}$.

Veamos que esto es imposible por inducción en m. Si $m = 1$, tendríamos $\mathbf{v}_1 = \mathbf{0}$, pero esto es imposible porque $\mathbf{v}_1$ es un autovector. Supongamos que $m > 1$, y que el resultado es cierto para $m - 1$. Multiplicando por A la suma de estos vectores, se tiene:

$$A(\mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m) = \mathbf{0} \quad \Rightarrow \quad \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}.$$

Pero por otro lado, alguno de los autovalores debe ser no nulo (supongamos que es λ_m). Multiplicamos entonces la suma inicial por λ_m, y tenemos:

$$\lambda_m\mathbf{v}_1 + \lambda_m\mathbf{v}_2 + \cdots + \lambda_m\mathbf{v}_m = \mathbf{0}.$$

Restando estas dos expresiones, concluimos:

$$(\lambda_1 - \lambda_m)\mathbf{v}_1 + (\lambda_2 - \lambda_m)\mathbf{v}_2 + \cdots + (\lambda_{m-1} - \lambda_m)\mathbf{v}_{m-1} + \mathbf{0} = \mathbf{0}.$$

Pero los $m-1$ vectores de esta expresión son no nulos, luego esto es imposible por hipótesis de inducción. $\square$

Corolario 5.4 *Una matriz $A \in \mathcal{M}_{n\times n}$ no puede tener más de n autovalores distintos.*

DEMOSTRACIÓN: Como la suma de sus espacios propios es directa, la dimensión de la suma de todos los espacios propios es la suma de las dimensiones de cada espacio. Como esta suma no puede ser mayor que n (la dimensión de V), se concluye que no puede haber más de n espacios propios, luego no puede haber más de n autovalores. $\square$

Hemos definido entonces los autovalores y autovectores de una matriz, o de un endomorfismo de V. Una propiedad importante es que los autovalores de una matriz no cambian si cambiamos de base. Además, las dimensiones de los subespacios propios también se mantienen. Es decir:

Proposición 5.5 *Si dos matrices $A, B \in \mathcal{M}_{n\times n}$ son semejantes, sus polinomios característicos coinciden, y los subespacios propios correspondientes a cada autovalor tienen la misma dimensión.*

Demostración: Sabemos que $P^{-1}AP = B$ para una cierta matriz P. Entonces se tiene:

$$P^{-1}(A - \lambda I)P = (P^{-1}A - \lambda P^{-1})P = P^{-1}AP - \lambda P^{-1}P = B - \lambda I.$$

Por tanto,

$$|B - \lambda I| = |P^{-1}(A - \lambda I)P| = |P^{-1}||A - \lambda I||P| = |A - \lambda I|,$$

es decir, los polinomios característicos de A y B (y por tanto sus autovalores) coinciden.

Por otra parte, si fijamos un autovalor λ de A y B, la dimensión del subespacio propio correspondiente viene determinada por el rango de la matriz $A - \lambda I$, o $B - \lambda I$, en cada caso. Pero hemos visto que

$$B - \lambda I = P^{-1}(A - \lambda I)P,$$

donde P es una matriz no singular. Por tanto los rangos de $A - \lambda I$ y de $B - \lambda I$ coinciden.
□

5.2. Multiplicidad algebraica y geométrica. Diagonalización.

Volvamos al problema de inicio. Estamos intentando saber, dada una matriz A, si existe una matriz semejante que sea diagonal. Vimos que si D era una matriz diagonal, entonces existen n autovectores linealmente independientes (los de la base canónica). Esto quiere decir que, si D es diagonal, la suma de las dimensiones de todos los subespacios propios debe ser n. Como estas dimensiones son invariantes por semejanza, esta misma propiedad la deben satisfacer *todas las matrices diagonalizables*. Es una condición necesaria para que A sea diagonalizable. Veremos que también es una condición suficiente. Para definirla con más propiedad, comenzaremos definiendo las multiplicidades de los autovalores:

Multiplicidad algebraica y geométrica de un autovalor: Sea $A \in \mathcal{M}_{n\times n}$ (y f el endomorfismo que representa). Sea λ_0 un autovalor de A.

Se define la **multiplicidad algebraica** de λ_0 como el número de veces que aparece λ_0 como raíz de la ecuación característica de A. Es decir, la multiplicidad algebraica de λ_0 es m si el polinomio característico de A se puede escribir:

$$|A-\lambda I| = (\lambda-\lambda_0)^m p(\lambda),$$

donde $p(\lambda)$ es un polinomio que no tiene a λ_0 como raíz.

Se define la **multiplicidad geométrica** de λ_0 como la dimensión del subespacio propio $V_1(\lambda_0)$. Es decir, $\dim(\ker(g))$, donde g es la aplicación lineal determinada por la matriz $A-\lambda_0 I$. Dicho de otra forma, la multiplicidad geométrica de λ_0 es $n-\operatorname{rg}(A-\lambda_0 I)$.

Proposición 5.6 *Sea λ_0 un autovalor de un matriz $A \in \mathcal{M}_{n\times n}$. Sea m su multiplicidad algebraica y sea g su multiplicidad geométrica. Entonces $1 \leq g \leq m$.*

Demostración: La desigualdad $1 \leq g$ es muy sencilla de demostrar: si λ_0 es un autovalor, esto significa que tiene algún autovector asociado, es decir, que la dimensión $g = \dim(V_1(\lambda_0))$ debe ser al menos 1.

Por otro lado, sea $B_0 = \{\mathbf{e}_1, \dots, \mathbf{e}_g\}$ una base de $V_1(\lambda_0)$. Por el teorema de la base incompleta, podemos completar B_0 hasta una base B de V. Podemos entonces cambiar de base, y escribir la matriz A respecto de la base B. Esto quiere decir que tenemos una matriz A', semejante a A, que representa al mismo endomorfismo (f) que A, pero respecto de la base B. Ahora bien, sabemos que $f(\mathbf{e}_i) = \lambda_0 \mathbf{e}_i$, para $i = 1, \dots, g$. También sabemos que las columnas de A' representan $f(\mathbf{e}_i)$ para $i = 1, \dots, n$. Por tanto, se tiene:

$$A' = \left(\begin{array}{c|c} \lambda_0 I_g & M \\ \hline \mathcal{O} & N \end{array}\right),$$

para unas ciertas submatrices M y N. Pero entonces el polinomio característico de A', que coincide (al ser semejantes) con el polinomio característico de A, es de la forma:

$$|A'-\lambda I| = \left(\begin{array}{c|c} (\lambda_0-\lambda) I_g & M \\ \hline \mathcal{O} & N-\lambda I \end{array}\right) = (\lambda_0-\lambda)^g |N-\lambda I| = (\lambda_0-\lambda)^g p(\lambda),$$

para un cierto polinomio $p(\lambda)$ que podrá, o no, contener a λ_0 como raíz. Por tanto, la multiplicidad algebraica de λ_0 es al menos g, como queríamos demostrar. □

En esta demostración hemos visto cómo se puede diagonalizar un trozo de matriz: simplemente tomando autovectores como elementos de la base. Esto es exactamente lo que hay que hacer en el caso general. Por tanto, el resultado que buscábamos es el siguiente:

Teorema 5.7 *Una matriz $A \in \mathcal{M}_{n\times n}$ es diagonalizable, es decir, existe P invertible tal que $P^{-1}AP$ es diagonal, si y sólo si A admite n autovectores linealmente independientes. Es decir, si la multiplicidad algebraica de cada autovalor coincide con su multiplicidad geométrica, y la suma de todas las multiplicidades es igual a n.*

Demostración: Si D es una matriz diagonal, entonces los vectores $e_i = (0, \ldots, 0, \overset{(i)}{1}, 0, \ldots, 0)$ son autovectores de D. Por tanto D admite n autovectores linealmente independientes. Esto es equivalente a $g_1 + \cdots + g_d = \dim(V_1(\lambda_1)) + \cdots + \dim(V_1(\lambda_d)) = n$, donde $\lambda_1, \ldots, \lambda_d$ son los autovalores de D. Pero $g_i \leq m_i$ (donde m_i es la multiplicidad algebraica de λ_i, para $i = 1, \ldots, d$), y la suma de todas las multiplicidades algebraicas nunca puede ser mayor que n (que es el grado del polinomio característico). Por tanto, se tiene:

$$n = g_1 + \cdots + g_n \leq m_1 + \cdots + m_n \leq n,$$

es decir,

$$n = g_1 + \cdots + g_n = m_1 + \cdots + m_n = n,$$

y por tanto $g_i = m_i$ para todo $i = 1, \ldots, d$.

Ahora bien, si A es diagonalizable, entonces $P^{-1}AP = D$. Hemos demostrado que si dos matrices son semejantes, entonces sus polinomios característicos, y las dimensiones de sus subespacios propios, coinciden. Por tanto, las multiplicidades algebraicas y geométricas de los autovalores de A y D coinciden. Si estas multiplicidades son m_i y g_i, para $i = 1, \ldots, d$, se tiene $g_i = m_i$ y $g_1 + \cdots + g_d = m_1 + \cdots + m_d = n$. Como podemos tomar g_i vectores linealmente independientes de cada $V_1(\lambda_i)$, y todos estos subespacios son independientes, concluimos que A admite n autovectores linealmente independientes.

Recíprocamente, si A admite n autovectores linealmente independientes, basta formar una nueva base B con estos n autovectores, y tomar P como la matriz del cambio de base. La matriz resultante: $P^{-1}AP$ es diagonal, y los elementos de la diagonal principal son los autovalores de A. $\square$

Algunas observaciones sencillas, que nos pueden ayudar a determinar si una matriz es diagonalizable, son las siguientes:

Proposición 5.8 *Sea $A \in \mathcal{M}_{n\times n}$. Sean $\lambda_1, \ldots, \lambda_d$ sus autovalores, y sean g_i y m_i las multiplicidades geométrica y algebraica, respectivamente, de λ_i. Se tiene:*

1. *Si $d = n$, es decir, si A tiene n autovalores distintos, entonces A es diagonalizable.*

2. *Si $g_i < m_i$ para un valor de i, entonces A no es diagonalizable.*

DEMOSTRACIÓN: La primera propiedad se tiene ya que $1 \leq g_i$ para todo i. Por tanto, si $n = d$, tenemos $n \leq g_1 + \cdots + g_n \leq n$, por tanto $g_1 + \cdots + g_n = n$, y A es diagonalizable. La segunda propiedad es consecuencia directa del resultado anterior. $\square$

5.3. Forma canónica de Jordan. Subespacios propios generalizados.

Continuamos en esta sección estudiando los endomorfismos de un espacio vectorial V de dimensión n, o análogamente, las matrices de $\mathcal{M}_{n\times n}(K)$. Vimos en la sección anterior que si una matriz $A \in \mathcal{M}_{m\times n}$ admite n autovectores linealmente independientes, entonces podemos formar una base con esos autovectores, y al representar A respecto de esta nueva base, obtenemos una matriz diagonal, D, semejante a A, donde los elementos de la diagonal principal son los autovalores de A.

Pero no todas las matrices son diagonalizables. En el caso en que sólo existan $m < n$ autovectores linealmente independientes, vamos a buscar otros $n - m$ vectores, que completen los autovectores hasta una base de V, tales que al cambiar de base la matriz se convierta en otra lo más simple posible. Veremos en esta sección el caso en que A tenga exactamente n autovalores (contando multiplicidades). Es decir, si A tiene p autovalores, de multiplicidades algebraicas $m_1, \ldots, m_p$, y se tiene $m_1 + \cdots + m_p = n$, entonces existirá una matriz J, semejante a A, que se llama *forma canónica de Jordan*, y que es suficientemente simple, aunque no sea diagonal.

Nota: Normalmente, en los ejemplos que usamos, el cuerpo K es igual a $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$. De estos tres cuerpos, $\mathbb{C}$ es el más aconsejable, ya que es un cuerpo **algebraicamente cerrado**. Esto quiere decir que todo polinomio de grado n en $\mathbb{C}$ tiene exactamente n raíces (contando multiplicidades). Por tanto, si consideramos $K = \mathbb{C}$, toda matriz admite una forma canónica de Jordan, ya que su ecuación característica tendrá n raíces. Sin embargo, esto no ocurre para $\mathbb{Q}$ y $\mathbb{R}$.

Vamos a definir ya cómo son las matrices de Jordan. Comenzamos con una pieza básica para construir estas matrices:

Bloque de Jordan: Dado un escalar $\lambda \in K$, llamamos **bloque de Jordan** de orden m asociado a λ, a la matriz $m \times m$ siguiente:

$$J(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}.$$

Es decir, para todo i, la entrada (i,i) es λ, y la entrada $(i,i+1)$ es 1. Todas las demás entradas son nulas.

Usando estos bloques de Jordan, podemos definir una matriz de Jordan:

Matriz de Jordan: Diremos que una matriz $J \in \mathcal{M}_{n\times n}$ es una **matriz de Jordan**, si existen unos bloques de Jordan, $J(\lambda_1), \ldots, J(\lambda_r)$ (no necesariamente del mismo tamaño), tales que J es *diagonal por bloques*, de la siguiente forma:

$$J = \begin{pmatrix} \boxed{J(\lambda_1)} & & & \\ & \boxed{J(\lambda_2)} & & \\ & & \ddots & \\ & & & \boxed{J(\lambda_r)} \end{pmatrix},$$

donde todas las entradas de J fuera de los bloques referidos son nulas.

Observemos que si J es una matriz de Jordan, entonces las únicas entradas que pueden ser no nulas son aquellas de la forma (i,i) o $(i,i+1)$, y estas últimas sólo pueden tomar los valores 1 o 0. Por tanto, una matriz de Jordan es "casi" una matriz diagonal.

Queremos demostrar, entonces, que toda matriz $A \in \mathcal{M}_{n\times n}$ que tenga n autovalores (contando multiplicidades), es semejante a una matriz de Jordan. Supongamos que A tiene p autovalores distintos, $\lambda_1, \ldots, \lambda_p$, con multiplicidades geométricas $g_1, \ldots, g_p$, y multiplicidades algebraicas $m_1, \ldots, m_p$. Sabemos, por la sección anterior, que existen $g_1 + \cdots + g_p$ autovectores linealmente independientes. Si este número es igual a n, entonces la matriz es diagonalizable, y como toda matriz diagonal es de Jordan (con bloques de orden 1), este caso ya está probado. Vamos a suponer entonces que existe algún autovalor λ_i con $g_i < m_i$. Necesitaríamos entonces $m_i - g_i$ vectores más, asociados a λ_i, para intentar completar una base que contenga a los autovectores.

La idea es la siguiente. Los autovectores asociados a λ_i son aquellos $\mathbf{v} \in V$ tales que $(A - \lambda_i I)\mathbf{v} = \mathbf{0}$. Es decir, son las preimágenes de $\mathbf{0}$ por la aplicación g asociada a la matriz $A - \lambda_i I$. Si no tenemos suficientes autovectores independientes, consideraremos las preimágenes por g de los autovectores. Si aún no tenemos suficientes, consideraremos las preimágenes por g de estos últimos, y así sucesivamente hasta que obtengamos m_i vectores linealmente independientes. Estos serán los vectores que usaremos para completar la base de autovectores.

Subespacios propios generalizados: Sea $A \in \mathcal{M}_{n\times n}$, y sea λ un autovalor de A. Consideremos la aplicación lineal g determinada por la matriz $A-\lambda I$. Entonces, para todo $i \geq 1$, llamamos **subespacios propios generalizados** asociados a λ, a los subespacios:

$$V_i = \ker(g^i).$$

Es decir:

$$V_i = \{\mathbf{v} \in V| \quad (A - \lambda I)^i\, \mathbf{v} = \mathbf{0}\}.$$

Algunas propiedades importantes de estos subespacios son las siguientes:

Proposición 5.9 *Con las notaciones anteriores, donde f y g son las aplicaciones lineales representadas por A y $A - \lambda I$ respectivamente, se tiene:*

1. $f(V_i) \subset V_i$ *para todo* $i \geq 1$.
2. $g(V_i) \subset V_{i-1}$ *para todo* $i \geq 2$, *y* $\quad g(V_1) = \{0\}$.
3. $V_1 \subset V_2 \subset V_3 \subset V_4 \subset \cdots$
4. *Sea s el menor número tal que $V_s = V_{s+1}$. Entonces $V_s = V_t$ para cualquier $t > s$.*

Demostración: La primera propiedad se demuestra observando que las matrices A y $A - \lambda I$ conmutan:

$$A(A - \lambda I) = A^2 - \lambda A = (A - \lambda I)A.$$

Por tanto, las matrices A y $(A - \lambda I)^i$ también conmutan. Pero entonces, dado un vector $\mathbf{v} \in V_i$, es decir, un vector tal que $(A-\lambda I)^i\mathbf{v} = 0$, tendremos, $(A-\lambda I)^i A\mathbf{v} = A(A - \lambda I)^i\mathbf{v} = A\mathbf{0} = \mathbf{0}$. Por tanto $A\mathbf{v}$, es decir, $f(\mathbf{v})$, pertenece a $\ker(g^i)$. Hemos demostrado entonces que $f(V_i) \subset V_i$, o dicho de otra forma, que V_i es un subespacio invariante para f.

La segunda propiedad se obtiene directamente a partir de la definición. Si $\mathbf{v} \in V_i$ para $i \geq 2$, es decir si $g^i(\mathbf{v}) = \mathbf{0}$, entonces $g^{i-1}(g(\mathbf{v})) = \mathbf{0}$, por lo que $g(\mathbf{v}) \in V_{i-1}$. Por otra parte $g(V_1) = \{0\}$ también por definición.

Probemos entonces la tercera propiedad. Si $\mathbf{v} \in V_i$, entonces $g^i(\mathbf{v}) = \mathbf{0}$. Aplicando g de nuevo, se tiene $g^{i+1}(\mathbf{v}) = \mathbf{0}$, por lo que $\mathbf{v} \in V_{i+1}$. Por tanto $V_i \subset V_{i+1}$ para todo $i \geq 1$, como queríamos demostrar.

Por último, veamos que la cuarta propiedad es cierta, es decir, que la cadena ascendente de subespacios $V_1 \subset V_2 \subset V_3 \subset \cdots$ se estabiliza en cuanto aparece una repetición. En primer lugar, debe existir un s tal que $V_s = V_{s+1}$, ya que todos estos espacios están contenidos en V, que tiene dimensión finita, y con cada inclusión estricta aumenta la dimensión del subespacio correspondiente. Por tanto, como máximo $s = n$. Sea entonces s el mínimo entero tal que $V_s = V_{s+1}$. Esto quiere decir que si $\mathbf{v}'$ es un vector tal que $g^{s+1}(\mathbf{v}') = \mathbf{0}$, entonces $g^s(\mathbf{v}') = \mathbf{0}$. Sea entonces $\mathbf{v} \in V_t$, con $t > s$. Se tiene

$$g^t(\mathbf{v}) = \mathbf{0} \quad \Rightarrow \quad g^{s+1}(g^{t-s-1}(\mathbf{v})) = \mathbf{0}.$$

Como $t - s - 1 \geq 0$, podemos considerar el vector $\mathbf{v}' = g^{t-s-1}(\mathbf{v})$. Tenemos entonces

$$g^{s+1}(\mathbf{v}') = \mathbf{0} \quad \Rightarrow \quad g^s(\mathbf{v}') = \mathbf{0} \quad \Rightarrow \quad g^{t-1}(\mathbf{v}) = g^s(g^{t-s-1}(\mathbf{v})) = \mathbf{0}.$$

Luego $V_t = V_{t-1}$, para todo $t > s$. Por inducción, concluimos que $V_t = V_s$ para todo $t > s$.
□

Sea $A \in \mathcal{M}_{n\times n}$, y sea λ un autovalor de A. Sean V_i, para $i \geq 1$, los subespacios propios generalizados asociados a λ, y sea s el menor entero tal que $V_s = V_{s+1}$. Entonces V_s se llama **subespacio propio generalizado maximal** asociado a λ, y lo denotamos V_{max}.

5.4. Cálculo de la base de Jordan.

La importancia del espacio V_{max} es que vamos a poder obtener de él los vectores que buscamos, para obtener una base de V que contenga a los autovectores. Además, al cambiar a esta nueva base, la matriz A se transformará en una matriz de Jordan. El resultado que necesitamos es el siguiente.

Proposición 5.10 *Con las notaciones anteriores, existe una base B de V_{max} tal que, para todo $\mathbf{v} \in B$, o bien $g(\mathbf{v}) = \mathbf{0}$ (es decir, $\mathbf{v} \in V_1$ es un autovector asociado a λ), o bien $g(\mathbf{v}) \in B$.*

Nota: El enunciado quiere decir que existe una base de V_{max} cuyos vectores se pueden distribuir formando una “escalera”, como en el siguiente ejemplo:

$$\begin{tikzcd}
\boxed{\mathbf{v}_1^{(4)}} \arrow[d, "g"] & \boxed{\mathbf{v}_2^{(4)}} \arrow[d, "g"] & & & \\
\boxed{\mathbf{v}_1^{(3)}} \arrow[d, "g"] & \boxed{\mathbf{v}_2^{(3)}} \arrow[d, "g"] & \boxed{\mathbf{v}_3^{(3)}} \arrow[d, "g"] & & \\
\boxed{\mathbf{v}_1^{(2)}} \arrow[d, "g"] & \boxed{\mathbf{v}_2^{(2)}} \arrow[d, "g"] & \boxed{\mathbf{v}_3^{(2)}} \arrow[d, "g"] & & \\
\boxed{\mathbf{v}_1} \arrow[d, "g"] & \boxed{\mathbf{v}_2} \arrow[d, "g"] & \boxed{\mathbf{v}_3} \arrow[d, "g"] & \boxed{\mathbf{v}_4} \arrow[d, "g"] & \boxed{\mathbf{v}_5} \arrow[d, "g"] \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{tikzcd}$$

Demostración: Comenzaremos por definir una base conveniente de $V_1 = V_1(\lambda)$. En este subespacio propio, formado por los autovectores asociados a λ, puede que haya autovectores que pertenezcan a $\mathrm{Im}(g)$. También puede haber autovectores contenidos en $\mathrm{Im}(g^2), \mathrm{Im}(g^3), \ldots, \mathrm{Im}(g^{s-1})$. No necesitamos ir más allá, ya que $\mathrm{Im}(g^s) \cap V_1 = \{\mathbf{0}\}$. En efecto, si un vector $\mathbf{v}$ pertenece a $\mathrm{Im}(g^s) \cap V_1$, es decir, si existe $\mathbf{u}$ tal que $g^s(\mathbf{u}) = \mathbf{v} \in V_1$, entonces $g^{s+1}(\mathbf{u}) = g(\mathbf{v}) = \mathbf{0}$. Pero en ese caso, como $V_{s+1} = V_s$, se tiene $g^s(\mathbf{u}) = \mathbf{0}$, es decir, $\mathbf{v} = \mathbf{0}$.

Por tanto, tenemos una sucesión ascendente de subespacios de V_1:

$$\{\mathbf{0}\} \subset \big(\mathrm{Im}(g^{s-1}) \cap V_1\big) \subset \big(\mathrm{Im}(g^{s-2}) \cap V_1\big) \subset \cdots \subset \ (\mathrm{Im}(g) \cap V_1) \subset V_1.$$

Denotaremos las dimensiones de estos subespacios $p_s, p_{s-1}, \ldots, p_1$, respectivamente. (En la “escalera” anterior, p_i es el número de columnas de tamaño mayor o igual a i, por tanto, $p_4 = 2$, $p_3 = 3$, $p_2 = 3$, $p_1 = 5$. Observemos que p_i es precisamente el tamaño de la fila i, si contamos las filas comenzando por abajo.) Tendremos entonces $p_s \leq p_{s-1} \leq \cdots \leq p_1$. Consideremos entonces una base B_s de $\mathrm{Im}(g^{s-1}) \cap V_1$. La podemos ampliar a una base B_{s-1} de $\mathrm{Im}(g^{s-2}) \cap V_1$, y así sucesivamente, hasta una base $B_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_{p_1}\}$ de V_1 (que corresponde a la fila inferior de la “escalera”).

Vamos ahora a ampliar la base B_1, formada por autovectores, usando los vectores de $V_2, V_3, \ldots, V_s$. Lo haremos de la siguiente manera: calcularemos el primer vector de cada “columna de la escalera”, y le aplicaremos g repetidas veces, hasta completar la “columna”. Esto se puede hacer, ya que dado $\mathbf{v}_i \in B$ cuya “columna” tenga tamaño r, es decir, $\mathbf{v}_i \in \mathrm{Im}(g^{r-1})$, debe existir un vector $\mathbf{v}_i^{(r)}$ tal que $g^{r-1}(\mathbf{v}_i^{(r)}) = \mathbf{v}_i$. Este vector puede ser cualquier solución del sistema $(A - \lambda I)^{r-1}\mathbf{x} = \mathbf{v}_i$. Una vez hallado $\mathbf{v}_i^{(r)}$, se calculan

$\mathbf{v}_i^{(r-1)}, \mathbf{v}_i^{(r-2)}, \ldots, \mathbf{v}_i^{(2)}$, aplicando g repetidas veces. Es decir, $\mathbf{v}_i{}^{(r-j)} = g^j(\mathbf{v}_i^{(r)})$. Esto "rellena" la columna i de la "escalera".

Hemos construido entonces un sistema de vectores $B \subset V_{max}$ que satisface las condiciones del enunciado. Es decir, para todo $\mathbf{v} \in B$, o bien $g(\mathbf{v}) = \mathbf{0}$ o bien $g(\mathbf{v}) \in B$. Queda por demostrar todavía que B es una base de V_{max}.

Para ello, llamemos T_i al conjunto formado por los vectores de la "fila" i de la "escalera". Es decir, $T_1 = B_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_{p_1}\}$ y $T_i = \{\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{p_i}^{(i)}\}$ para $i = 2, \ldots, s$. Observemos que $B = T_1 \cup \cdots \cup T_s$. Vamos a probar por inducción en j, que el $T_1 \cup \cdots \cup T_j$ es base de V_j, y con esto habremos probado que B es base de $V_s = V_i^{max}$.

Para $j = 1$, tenemos $T_1 = B_1$, que es base de V_1. Podemos entonces suponer que $T_1 \cup \cdots \cup T_{j-1}$ es base de V_{j-1}, y probaremos el resultado para j. Primero veamos que $T_1 \cup \cdots \cup T_j$ sistema de generadores: Dado un vector $\mathbf{v} \in V_j$, sabemos que $g^{j-1}(\mathbf{v}) \in \mathrm{Im}(g^{j-1}) \cap V_1$. Como sabemos que B_j es una base de este subespacio, podemos escribir $g^{j-1}(\mathbf{v})$ como combinación lineal de los vectores de B_j. Tendremos:

$$g^{j-1}(\mathbf{v}) = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_{p_j} \mathbf{v}_{p_j}.$$

Consideremos ahora el vector $\mathbf{v}' = \alpha_1 \mathbf{v}_1^{(j)} + \cdots + \alpha_{p_j} \mathbf{v}_{p_j}^{(j)}$. Este vector es combinación lineal de los vectores de T_j, pero además, como $g^{j-1}(\mathbf{v}_r^{(j)}) = \mathbf{v}_r$ para todo r, se tiene $g^{j-1}(\mathbf{v}') = g^{j-1}(\mathbf{v})$. Es decir, $g^{j-1}(\mathbf{v} - \mathbf{v}') = \mathbf{0}$. Pero entonces $\mathbf{v} - \mathbf{v}' \in V_{j-1}$, y podemos escribir este vector como combinación lineal de los vectores de $T_1 \cup \cdots \cup T_{j-1}$. Por tanto, $\mathbf{v} = (\mathbf{v} - \mathbf{v}') + \mathbf{v}'$ se puede escribir como combinación lineal de los vectores de $T_1 \cup \cdots \cup T_j$, luego este sistema genera V_j, como queríamos demostrar.

Ahora veamos que $T_1 \cup \cdots \cup T_j$ es linealmente independiente. Supongamos que tenemos una combinación

$$\sum_{r=1}^{j} \left(\alpha_1^{(r)} \mathbf{v}_1^{(r)} + \cdots + \alpha_{p_r}^{(r)} \mathbf{v}_{p_r}^{(r)} \right) = \mathbf{0},$$

donde hemos escrito $\mathbf{v}_i^{(1)} = \mathbf{v}_i$ por comodidad. Aplicando g a toda la igualdad, y recordando que $g(\mathbf{v}_i^{(1)}) = \mathbf{0}$ para todo i, queda

$$\sum_{r=2}^{j} \left(\alpha_1^{(r)} \mathbf{v}_1^{(r-1)} + \cdots + \alpha_{p_r}^{(r)} \mathbf{v}_{p_r}^{(r-1)} \right) = \mathbf{0}.$$

Pero esta es una combinación lineal de elementos de $T_1 \cup \cdots \cup T_{j-1}$, que es un sistema linealmente independiente por hipótesis de inducción. Por tanto, los coeficientes $\alpha_i^{(r)} = 0$ para todo $r > 1$. Nos queda entonces la igualdad

$$\alpha_1^{(1)} \mathbf{v}_1^{(1)} + \cdots + \alpha_{p_1}^{(1)} \mathbf{v}_{p_1}^{(1)} = \mathbf{0},$$

pero como los vectores implicados son los de la base B_1, todos los coeficientes deben ser nulos, y por tanto, $T_1 \cup \cdots \cup T_j$ es linealmente independiente, como queríamos demostrar.

Hemos probado, por tanto, que B es una base de V_{max} que satisface las condiciones del enunciado, lo que termina la demostración. $\square$

Cálculo de la forma de la "escalera": A la hora de calcular la base B de V_{max}, comenzamos calculando las matrices de $g, g^2, g^3, \ldots$, es decir, si llamamos $G = A - \lambda I$, calculamos $G, G^2, G^3, \ldots$, hasta obtener G^s tal que $\mathrm{rg}(G^s) = n - m$, donde $m = \dim(V_{max})$ (como veremos más adelante, m es la multiplicidad del autovalor estudiado). En ese momento ya sabemos que el **número de filas de la escalera** es s.

Más aún, como hemos demostrado que $T_1 \cup \cdots \cup T_j$ es base de V_j, y $V_j = \ker(g^j) = n - \mathrm{rg}(G^j)$, se tiene que el número de vectores en las j filas inferiores de la escalera es precisamente $n - \mathrm{rg}(G^j)$. Por tanto, el **tamaño de la fila j** (contando las filas desde abajo) es $\mathrm{rg}(G^{j-1}) - \mathrm{rg}(G^j)$. Esto nos da la forma exacta de la escalera.

Nota: El cálculo de la base B_s de $\mathrm{Im}(g^{s-1}) \cap V_1$, con la que comienza el cálculo de la primera fila de la escalera, puede simplificarse mucho en el caso siguiente: Si f tiene sólo un autovalor de multiplicidad n, entonces $G^s = 0$, y por tanto el tamaño de la fila s es $\mathrm{rg}(G^{s-1})$, que es igual a $\dim(\mathrm{Im}(g^{s-1}))$. Como el tamaño de la fila s también es igual a $\dim(\mathrm{Im}(g^{s-1}) \cap V_1)$, esto implica que $\mathrm{Im}(g^{s-1}) \cap V_1 = \mathrm{Im}(g^{s-1})$, y los vectores que necesitamos (vectores de B_s) los podemos tomar simplemente del conjunto de columnas de la matriz G^{s-1}.

5.5. Base de Jordan y forma canónica de Jordan.

La base B de $V_{max}(\lambda_i)$ construida en la proposición anterior, llamada base de Jordan, es muy importante para hallar la forma canónica de Jordan de una matriz. Pero necesitamos ordenar sus vectores de la siguiente manera: para cada autovector $\mathbf{v}_j$, sea $S_{\mathbf{v}_j} = \{\mathbf{v}_j, \mathbf{v}_j^{(2)}, \ldots, \mathbf{v}_j^{(r)}\}$ el sistema formado por la columna j de la escalera, leída de abajo a arriba. Entonces tenemos:

$$B = S_{\mathbf{v}_1} \cup S_{\mathbf{v}_2} \cup \cdots \cup S_{\mathbf{v}_{p_1}}.$$

Esta es la base de V_i^{max} que usaremos para transformar la matriz A en una matriz de Jordan.

Proposición 5.11 *Sea f un endomorfismo de V, sea λ_i un autovalor de f, $\mathbf{v}$ un autovector asociado a λ_i, y $S_{\mathbf{v}} = \{\mathbf{v}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(r)}\}$ el sistema de vectores definido anteriormente. Entonces se tiene $f(S_{\mathbf{v}}) \subset \langle S_{\mathbf{v}} \rangle$, y la matriz de la restricción $f_{|\langle S_{\mathbf{v}}\rangle}$ respecto de la base $S_{\mathbf{v}}$ es un bloque de Jordan $J(\lambda_i)$.*

Demostración: Por simplificar la notación, llamaremos J a la matriz del endomorfismo $f_{|\langle S_{\mathbf{v}}\rangle}$. Recordemos que $S_{\mathbf{v}} = \{\mathbf{v}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(r)}\}$, para un cierto r, y que las columnas de J serán las coordenadas, respecto de esta base, de las imágenes por f de los elementos de la base.

Apliquemos f a cada elemento de $S_{\mathbf{v}}$. En primer lugar, como $\mathbf{v}$ es un autovector, se tiene $f(\mathbf{v}) = \lambda_i \mathbf{v}$. Por tanto, $f(\mathbf{v}) \in \langle S_{\mathbf{v}} \rangle$, y la primera columna de J será $(\lambda_i, 0, \ldots, 0)$. Ahora, para todo $r > 1$, tendremos $g(\mathbf{v}^{(r)}) = \mathbf{v}^{(r-1)}$, es decir, $f(\mathbf{v}^{(r)}) - \lambda_i \mathbf{v}^{(r)} = \mathbf{v}^{(r-1)}$. Por tanto, $f(\mathbf{v}^{(r)}) = \mathbf{v}^{(r-1)} + \lambda_i \mathbf{v}^{(r)}$, luego $f(\mathbf{v}^{(r)}) \in \langle S_{\mathbf{v}} \rangle$, y la columna correspondiente de la matriz J será $(0, \ldots, 0, \overset{(r-1)}{1}, \overset{(r)}{\lambda_i}, 0 \ldots, 0)$. Por tanto, tendremos

$$J = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix} = J(\lambda_i),$$

como queríamos demostrar. $\square$

Corolario 5.12 *Con las condiciones anteriores, $f(V_{max}) \subset V_{max}$, y la matriz de $f_{|V_{max}}$ respecto de la base $B = S_{\mathbf{v}_1} \cup \cdots \cup S_{\mathbf{v}_{p_1}}$ es una matriz de Jordan.*

Demostración: Basta aplicar el resultado anterior a cada uno de los sistemas $S_{\mathbf{v}_j}$, y obtendremos que la matriz $M(f_{|V_{max}})$ es diagonal por bloques de Jordan, todos ellos asociados al autovalor λ_i. $\square$

Corolario 5.13 *Con las condiciones anteriores, si λ_i tiene multiplicidad algebraica m_i, entonces* $\dim(V_{max}) = m_i$.

Demostración: Llamemos d a la dimensión de V_{max}, y consideremos la base B de V_{max} definida anteriormente. Ampliemos B hasta una base B' de todo V, y llamemos M a la matriz de f respecto de la base B'. Ya sabemos cómo son las d primeras columnas de M, luego esta matriz será de la forma:

$$M = \left(\begin{array}{c|c} J & P \\ \hline \mathcal{O} & Q \end{array}\right),$$

donde J es una matriz de Jordan formada por bloques asociados a λ_i, y $\mathcal{O}$ es la matriz nula. Como J es una matriz triangular superior, y los elementos de su diagonal principal son todos iguales a λ_i, se tiene:

$$|M - \lambda I| = (\lambda_i - \lambda)^d |Q - \lambda I|.$$

Por tanto, $d \leq m_i$, y se tendrá la igualdad si λ_i no es raíz de $|Q - \lambda I|$.

Supongamos entonces que λ_i es raíz de $|Q - \lambda I|$, es decir, que λ_i es autovalor de la matriz $Q \in \mathcal{M}_{(n-d)\times(n-d)}$. En ese caso, la matriz Q admitirá un autovector $\mathbf{v} = (v_{d+1}, \ldots, v_n)$ asociado a λ_i, es decir, tal que $Q\mathbf{v} = \lambda_i \mathbf{v} = (\lambda_i v_{d+1}, \cdots, \lambda_i v_n)$.

Consideremos entonces el vector $\mathbf{v}' = (0, \ldots, 0, v_{d+1}, \ldots, v_n)$. Claramente $\mathbf{v}' \notin V_{max}$. Si le aplicamos f, obtendremos el vector $f(\mathbf{v}') = (w_1, \ldots, w_d, \lambda_i v_{d+1}, \ldots, \lambda_i v_n)$, para unas ciertas coordenadas $w_1, \ldots, w_d$. Pero entonces

$$g(\mathbf{v}') = f(\mathbf{v}') - \lambda_i \mathbf{v}' = M\mathbf{v}' - \lambda_i \mathbf{v}' = (w_1, \ldots, w_d, 0, \ldots, 0).$$

Esto es, como las d primeras coordenadas corresponden a los vectores de la base B, hemos demostrado que $g(\mathbf{v}') \in V_{max}$, para un vector $\mathbf{v}' \notin V_{max}$. Pero esto es imposible, ya que si $g(\mathbf{v}') \in V_j$ para un cierto j, entonces $\mathbf{v}' \in V_{j+1} \subset V_{max}$. Por tanto, la matriz Q no puede tener a λ_i como autovalor, luego $d = m_i$, como queríamos probar. $\square$

Ahora sólo nos queda demostrar el siguiente resultado, para ver que la matriz de f se puede transformar en una matriz de Jordan:

Proposición 5.14 *Sea f un endomorfismo de V que admite n autovalores, contando multiplicidades. Sean $\lambda_1, \ldots, \lambda_p$ los autovalores (distintos) de f, y sean $V_{max}(\lambda_1), \ldots, V_{max}(\lambda_p)$ sus espacios propios generalizados maximales. Entonces $V = V_{max}(\lambda_1) \oplus \cdots \oplus V_{max}(\lambda_p)$.*

Demostración: Gracias al resultado anterior, sabemos que $\dim(V_{max}(\lambda_1)) + \cdots + \dim(V_{max}(\lambda_p)) = m_1 + \cdots + m_p = n$. Por tanto, lo único que tenemos que probar es que la suma $V_{max}(\lambda_1) + \cdots + V_{max}(\lambda_p)$ es directa.

Procedamos por inducción, probando que la suma $V_{max}(\lambda_1) + \cdots + V_{max}(\lambda_i)$ es directa. Para $i = 1$, no hay nada que probar. Supongamos que $i > 1$, y que el resultado es cierto para

$i-1$. Tenemos que demostrar que si $\mathbf{v}_1 + \cdots + \mathbf{v}_i = \mathbf{0}$, con $\mathbf{v}_j \in V_{max}(\lambda_j)$ para $j = 1, \ldots, i$, entonces $\mathbf{v}_j = \mathbf{0}$ para todo j.

Sea s tal que $V_{max}(\lambda_i) = V_s(\lambda_i)$, y sea $g = f - \lambda_i \mathrm{id}$. Sabemos que $g^s(\mathbf{v}) = \mathbf{0}$ para todo $\mathbf{v} \in V_{max}(\lambda_i)$. Entonces aplicamos g^s a la suma anterior y obtenemos:

$$g^s(\mathbf{v}_1 + \cdots + \mathbf{v}_{i-1} + \mathbf{v}_i) = g^s(\mathbf{v}_1) + \cdots + g^s(\mathbf{v}_{i-1}) + g^s(\mathbf{v}_i) = g^s(\mathbf{v}_1) + \cdots + g^s(\mathbf{v}_{i-1}) = \mathbf{0}.$$

Ahora veamos que para todo $j \neq i$, si un vector $\mathbf{v}$ pertenece a $V_r(\lambda_j)$ pero no pertenece a $V_{r-1}(\lambda_j)$, entonces $g^s(\mathbf{v})$ satisface la misma propiedad. En efecto, se tiene:

$$g(\mathbf{v}) = f(\mathbf{v}) - \lambda_i \mathbf{v} = f(\mathbf{v}) - \lambda_j \mathbf{v} + \lambda_j \mathbf{v} - \lambda_i \mathbf{v} = (f - \lambda_j \mathrm{id})(\mathbf{v}) + (\lambda_j - \lambda_i)\mathbf{v}.$$

Como $\mathbf{v} \in V_r(\lambda_j)$, tendremos $(f - \lambda_j \mathrm{id})(\mathbf{v}) \in V_{r-1}(\lambda_j)$. Y como $\lambda_j - \lambda_i \neq 0$, entonces $(\lambda_j - \lambda_i)\mathbf{v}$ es un múltiplo no nulo de $\mathbf{v}$, luego pertenece a $V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$. Por tanto, la suma de los dos vectores, es decir, $g(\mathbf{v})$ pertenece a $V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$. Pero esto implica que, si volvemos a aplicar g, volveremos a obtener un vector de $V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$. Y así sucesivamente, hasta llegar a $g^s(\mathbf{v}) \in V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$, como queríamos probar.

Supongamos entonces que algún vector $\mathbf{v}_j$ de la suma anterior es no nulo. Tendremos $\mathbf{v}_j \in V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$ para un cierto $r > 0$, luego tendríamos $g^s(\mathbf{v}_j) \in V_r(\lambda_j)\backslash V_{r-1}(\lambda_j)$, es decir, $g^s(\mathbf{v}_j) \neq 0$. Pero sabemos que $g^s(\mathbf{v}_1) + \cdots + g^s(\mathbf{v}_{i-1}) = \mathbf{0}$, donde $g^s(\mathbf{v}_l) \in V_{max}(\lambda_l)$ para todo l, por la propiedad que acabamos de probar. La hipótesis de inducción nos dice entonces que $g^s(\mathbf{v}_l) = \mathbf{0}$, para todo $l = 1, \ldots, i-1$, lo que lleva a una contradicción con $g^s(\mathbf{v}_j) \neq \mathbf{0}$. Por tanto, necesariamente $\mathbf{v}_1 = \cdots = \mathbf{v}_{i-1} = \mathbf{0}$. La suma inicial quedará entonces: $\mathbf{0} + \cdots + \mathbf{0} + \mathbf{v}_i = \mathbf{0}$, luego $\mathbf{v}_1 = \cdots = \mathbf{v}_{i-1} = \mathbf{v}_i = \mathbf{0}$, lo que demuestra el resultado. $\square$

5.6. Teorema de Jordan.

Reuniendo todos los resultados anteriores, obtenemos por fin el teorema que buscábamos:

Teorema 5.15 *Sea V un espacio vectorial de dimensión n, y sea $f \in \mathrm{End}(V)$. Si f admite n autovalores (contando multiplicidades), entonces existe una base de V respecto de la cual la matriz de f es una matriz de Jordan.*

Demostración: Si $\lambda_1, \ldots, \lambda_p$ son los autovalores (distintos) de f, consideramos los subespacios propios generalizados maximales $V_{max}(\lambda_1), \ldots, V_{max}(\lambda_p)$, y construimos las bases $B_1, \ldots, B_p$ de cada uno de ellos, como anteriormente. Por el resultado anterior, el sistema $B = B_1 \cup \cdots \cup B_p$ es una base de V, y la matriz de f respecto de B está formada por bloques de Jordan, luego es una matriz de Jordan. $\square$

Nota: De los resultados anteriores también podemos deducir cuántos bloques de Jordan tendrá la matriz, y qué dimensiones tendrán. En efecto, sea λ_i un autovalor de f, con $V_{max}(\lambda_i) = V_s(\lambda_i)$. Recordemos que $p_j = \dim(\mathrm{Im}(g^{j-1}) \cap V_{\lambda_i})$ es el tamaño de la fila j en la "escalera" correspondiente a λ_i. Si $\mathbf{v}$ es un autovector de la base B, que pertenece a $\mathrm{Im}(g^{j-1})$, pero no pertenece a $\mathrm{Im}(g^j)$ (es decir, cuya columna correspondiente en la "escalera" tiene tamaño j) entonces el sistema $S_{\mathbf{v}}$ consta de j vectores, y da lugar a un bloque de Jordan $J(\lambda_i)$ de tamaño j. Dicho de otra forma, cada columna de la "escalera" de tamaño j da lugar a un bloque de Jordan de tamaño j.

Por tanto, asociados al autovalor λ_i habrá $p_1 - p_2$ bloques de tamaño 1, habrá $p_2 - p_3$ bloques de tamaño 2, etc. En general, para $j = 1, \ldots, k$, habrá $p_j - p_{j+1}$ bloques de tamaño j. En otras palabras, hay tantos bloques de orden j como columnas de la escalera de tamaño j.

Forma canónica de Jordan: Dado $f \in \mathrm{End}(V)$, hemos demostrado que la matriz $M(f)$ es semejante a una matriz J de Jordan. A esta matriz J se le llama **forma canónica de Jordan** de f.

Proposición 5.16 *La forma canónica de Jordan de un endomorfismo f es única salvo permutación de los bloques de Jordan.*

Demostración: Sea J una forma canónica de Jordan de f. Sabemos que J es la matriz de f respecto de una cierta base $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, y sus columnas corresponden a $f(\mathbf{u}_1), \ldots, f(\mathbf{u}_n)$. Entonces, dado un bloque de Jordan $J(\lambda_i)$ de J, su primera columna corresponde a un autovector de f, su segunda columna corresponde a un vector de $V_2(\lambda_i)$, y así sucesivamente: su columna j corresponde a un vector de $V_j(\lambda_i)$.

Por tanto, a la vista de la matriz J podemos deducir los siguientes datos sobre f: El número de bloques de Jordan es igual al número de autovalores de f. El número de bloques asociados a λ_i es igual a la dimensión de $V_1(\lambda_i)$. De estos bloques, el número de ellos de tamaño menor o igual a j es igual a $\dim(V_j(\lambda_i)) - \dim(V_{j-1}(\lambda_i))$. Como estas dimensiones no dependen de la base respecto de la cual f está representada, se sigue que cualquier otra forma de Jordan de f tiene exactamente los mismos bloques, aunque tal vez cambiados de orden (esto equivale a una reordenación de los elementos de la base). $\square$

Gracias a lo estudiado en este tema, tenemos un método para determinar si dos matrices $n \times n$ son semejantes, es decir, si son matrices de un mismo endomorfismo de V respecto de dos bases distintas. Pero recordemos que este resultado sólo es válido para matrices con n autovalores (contando multiplicidades). O más generalmente, es válido para todas las matrices sobre un cuerpo algebraicamente cerrado (digamos $\mathbb{C}$).

Teorema 5.17 (Teorema de Jordan) *Dos matrices cuadradas sobre un cuerpo algebraicamente cerrado son semejantes si y sólo si tienen la misma forma canónica de Jordan (salvo permutación de sus bloques).*

Tema 6. Espacios vectoriales euclídeos

6.1. Formas bilineales.

Terminaremos esta asignatura con un tema que tiene mucho que ver con la asignatura de Geometría. Se trata de otra forma distinta de usar las matrices, los vectores, y las aplicaciones entre espacios vectoriales. Terminaremos definiendo, de manera muy general, lo que es un *producto escalar* y sus principales propiedades.

Cuando estudiamos las aplicaciones lineales entre espacios vectoriales, vimos que podían representarse mediante una matriz, y así el vector $f(\mathbf{v})$ era igual al vector $A\mathbf{v}$. Pues bien, hay otro tipo de aplicaciones entre espacios vectoriales, en los que se pueden usar matrices: dados dos vectores $\mathbf{u}, \mathbf{v}$, de un espacio vectorial V (de dimensión n) sobre K, y una matriz $A \in \mathcal{M}_{n\times n}(K)$, podemos definir el escalar $\mathbf{u}^t A\mathbf{v}$. Es decir:

$$\mathbf{u}^t\, A\, \mathbf{v} = (u_1\ u_2\ \cdots\ u_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Esto se puede considerar como una aplicación del espacio vectorial $V \times V$ en el cuerpo K, que podemos denotar f. Así, tendremos una aplicación $f : V \times V \to K$.

Proposición 6.1 *Dada una matriz $A \in \mathcal{M}_{n\times n}(K)$, la aplicación $f : V\times V \to K$, definida por $f(\mathbf{u}, \mathbf{v}) = \mathbf{u}^t\, A\, \mathbf{v}$, satisface las siguientes propiedades, para todo $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, y todo $\alpha \in K$:*

1. $f(\mathbf{u} + \mathbf{v}, \mathbf{w}) = f(\mathbf{u}, \mathbf{w}) + f(\mathbf{v}, \mathbf{w})$.
2. $f(\alpha\mathbf{u}, \mathbf{v}) = \alpha f(\mathbf{u}, \mathbf{v})$.
3. $f(\mathbf{u}, \mathbf{v} + \mathbf{w}) = f(\mathbf{u}, \mathbf{v}) + f(\mathbf{u}, \mathbf{w})$.
4. $f(\mathbf{u}, \alpha\mathbf{v}) = \alpha f(\mathbf{u}, \mathbf{v})$.

DEMOSTRACIÓN: Directa. $\square$

Aplicación bilineal: Si una aplicación $f : V \times V \to K$ satisface las cuatro propiedades anteriores, se llama **aplicación bilineal**, o **forma bilineal** sobre V.

La correspondencia entre aplicaciones bilineales y matrices es una correspondencia biunívoca:

Proposición 6.2 *Sea V un espacio vectorial sobre K de dimensión n, y sea B una base de V. Dada una aplicación bilineal f sobre V, existe una única matriz $A \in \mathcal{M}_{n\times n}(K)$, tal que $f(\mathbf{u},\mathbf{v}) = \mathbf{u}^t A\mathbf{v}$, donde los vectores $\mathbf{u},\mathbf{v} \in V$ están expresados con respecto a B.*

Demostración: Si $B = \{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$, la matriz A viene dada por: $a_{ij} = f(\mathbf{e}_i,\mathbf{e}_j)$. Es decir,

$$A = \begin{pmatrix} f(\mathbf{e}_1,\mathbf{e}_1) & f(\mathbf{e}_1,\mathbf{e}_2) & \cdots & f(\mathbf{e}_1,\mathbf{e}_n) \\ f(\mathbf{e}_2,\mathbf{e}_1) & f(\mathbf{e}_2,\mathbf{e}_2) & \cdots & f(\mathbf{e}_2,\mathbf{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{e}_n,\mathbf{e}_1) & f(\mathbf{e}_n,\mathbf{e}_2) & \cdots & f(\mathbf{e}_n,\mathbf{e}_n) \end{pmatrix}$$

Se demuestra de forma directa que $f(\mathbf{u},\mathbf{v}) = \mathbf{u}^t A\mathbf{v}$, y la unicidad se tiene ya que la matriz está definida de forma unívoca a partir de f. $\square$

Como ya hemos dicho, este tipo de funciones se usarán, entre otras cosas, para definir productos escalares entre dos vectores. Pero antes veremos cómo afecta a la matriz de f un cambio de la base de V.

Proposición 6.3 *Sea f una aplicación bilineal sobre V. Sean B y B' dos bases de V, y sean A y A' las matrices de f respecto de las bases B y B'. Si $M_{B',B}$ es la matriz del cambio de base, entonces $A' = M_{B',B}^t A M_{B',B}$.*

Demostración: Sabemos que para todo $\mathbf{v} \in V$, se tiene $\mathbf{v}_B = M_{B',B}\mathbf{v}_{B'}$. Por tanto, $f(\mathbf{u},\mathbf{v}) = \mathbf{u}_B^t A\mathbf{v}_B = (\mathbf{u}_{B'}^t M_{B',B}^t)A(M_{B',B}\mathbf{v}_{B'})$. Pero por otro lado, $f(\mathbf{u},\mathbf{v}) = \mathbf{u}_{B'}A'\mathbf{v}_{B'}$, de donde se deduce la igualdad propuesta. $\square$

Al igual que dos matrices que definían el mismo endomorfismo de V se decían semejantes, existe un término para denotar a las matrices que definen una misma aplicación bilineal:

Matrices congruentes: Se dice que dos matrices $A, A' \in \mathcal{M}_{n\times n}(K)$ son **congruentes**, si existe una matriz no singular P tal que $A' = P^t AP$.

Por el resultado anterior, se tiene que dos matrices son congruentes si y sólo si son las matrices de una misma aplicación bilineal, respecto de bases distintas.

De entre todas las posibles aplicaciones (o formas) bilineales, nos interesan especialmente un tipo concreto:

Formas bilineales simétricas: Una forma bilineal $f : V \times V \to K$ se dice **simétrica** si $f(\mathbf{u}, \mathbf{v}) = f(\mathbf{v}, \mathbf{u})$, para todo $\mathbf{u}, \mathbf{v} \in V$.

Es muy fácil ver si una aplicación bilineal es simétrica, simplemente observando su matriz:

Proposición 6.4 *Sea $A \in \mathcal{M}_{n\times n}(K)$ la matriz de una forma bilineal, respecto de una base cualquiera de V. Entonces f es simétrica si y sólo si A es una matriz simétrica.*

Demostración: Directa. $\square$

6.2. Ortogonalidad.

Al igual que hicimos con las aplicaciones lineales, vamos a intentar encontrar una base de V respecto de la cual la matriz de una aplicación bilineal sea lo más sencilla posible: A ser posible, diagonal. Nos centraremos en las aplicaciones bilineales simétricas. Primero definiremos la ortogonalidad respecto de una forma bilineal:

Vectores ortogonales: Sea $f : V \times V \to K$ una forma bilineal simétrica. Diremos que dos vectores $\mathbf{u}$ y $\mathbf{v}$ son **ortogonales** respecto de f, si $f(\mathbf{u}, \mathbf{v}) = 0$.

Nota: Observemos que dos vectores ortogonales no tienen por qué ser *perpendiculares*. Esto ocurrirá si la matriz de f es la matriz identidad. En este caso, $f(\mathbf{u}, \mathbf{v}) = \mathbf{u}^t\, \mathbf{v}$ es el *producto escalar* usual de los vectores $\mathbf{u}$ y $\mathbf{v}$. En este caso particular, *ortogonal* y *perpendicular* son palabras equivalentes.

Recordemos ahora que las entradas de la matriz de f son los elementos $f(\mathbf{v}_i, \mathbf{v}_j)$, donde $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ son los elementos de la base de V que hayamos fijado. Por tanto, si queremos que la matriz de f sea diagonal, es necesario que $f(\mathbf{v}_i, \mathbf{v}_j) = 0$, para todo $i \neq j$. Es decir, que los elementos de la base sean ortogonales dos a dos.

Base ortogonal: Dada una aplicación bilineal f, diremos que una base B de V es **ortogonal** si sus vectores son ortogonales dos a dos, respecto de f. Es decir, si $f(\mathbf{u}, \mathbf{v}) = 0$, para cualesquiera $\mathbf{u}, \mathbf{v} \in B$, $\mathbf{u} \neq \mathbf{v}$.

Afortunadamente, toda aplicación bilineal *simétrica* es diagonalizable, es decir, para toda aplicación bilineal f existe una base de V que es ortogonal. Para probar esto, definiremos primero la *variedad ortogonal* a una variedad lineal.

Variedad ortogonal: Sea L una variedad lineal de un espacio vectorial V, y fijemos una forma bilineal simétrica f sobre V. Se define la **variedad ortogonal** a L, que denotamos $L^{\perp}$, como el conjunto de los vectores ortogonales a todos los de L. Es decir:

$$L^{\perp} = \{\mathbf{v} \mid \quad f(\mathbf{u}, \mathbf{v}) = 0, \ \forall \mathbf{u} \in L\}.$$

Proposición 6.5 *En las condiciones anteriores, fijemos una base B de V. Supongamos que $L = \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle$, donde las coordenadas de $\mathbf{u}_i$ respecto de B son $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,n})$, y sea A la matriz de f respecto de B. Entonces $L^{\perp}$ viene definida por las siguientes ecuaciones implícitas:*

$$\left(\begin{array}{cccc} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ \hline u_{2,1} & u_{2,2} & \cdots & u_{2,n} \\ \hline \vdots & \vdots & & \vdots \\ \hline u_{r,1} & u_{r,2} & \cdots & u_{r,n} \end{array}\right) A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Es decir,

$$\left(\begin{array}{c} \mathbf{u}_1^t A \\ \hline \mathbf{u}_2^t A \\ \hline \vdots \\ \hline \mathbf{u}_r^t A \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Corolario 6.6 *Con las condiciones anteriores, si* $\dim(L) = r$ *entonces* $\dim(L^{\perp}) \geq n - r$.

Demostración: Por el resultado anterior sabemos que $L^{\perp}$ viene definida por r ecuaciones implícitas, que no necesariamente serán independientes. Si hay m ecuaciones independientes ($m \leq r$), entonces $\dim(L^{\perp}) = n - m \geq n - r$. $\square$

Veamos ahora que hay un caso particular (e importante) en el que $\dim(L) = r$ y $\dim(L^{\perp}) = n - r$. Más aún, en este caso particular las dos variedades van a ser complementarias.

Proposición 6.7 *Con las condiciones anteriores, sea* $L = \langle \mathbf{u}_1, \dots, \mathbf{u}_r \rangle$ *tal que los vectores* $\mathbf{u}_i$ *son linealmente independientes, ortogonales entre sí (es decir,* $f(\mathbf{u}_i, \mathbf{u}_j) = 0$ *para* $i \neq j$*) y no son ortogonales a sí mismos (es decir,* $f(\mathbf{u}_i, \mathbf{u}_i) = b_i \neq 0$ *para todo* i*). Entonces* $V = L \oplus L^\perp$.

Demostración: En primer lugar, veamos que $L \cap L^\perp = \{\mathbf{0}\}$. En efecto, todo vector $\mathbf{v} \in L$ se puede escribir $\mathbf{v} = \alpha_1 \mathbf{u}_1 + \cdots \alpha_r \mathbf{u}_r$. Si también tuviéramos $\mathbf{v} \in L^\perp$, entonces para todo $i = 1, \dots, r$ tendríamos $f(\mathbf{u}_i, \mathbf{v}) = 0$, donde $f(\mathbf{u}_i, \mathbf{v}) = f(\mathbf{u}_i, \alpha_1 \mathbf{u}_1 + \cdots + \alpha_r \mathbf{u}_r) = \alpha_1 f(\mathbf{u}_i, \mathbf{u}_1) + \cdots + \alpha_r f(\mathbf{u}_i, \mathbf{u}_r)$. Como los vectores $\mathbf{u}_1, \dots, \mathbf{u}_r$ son ortogonales entre sí, nos queda $f(\mathbf{u}_i, \mathbf{v}) = \alpha_i f(\mathbf{u}_i, \mathbf{u}_i) = \alpha_i b_i = 0$. Pero como $b_i \neq 0$, esto implica necesariamente $\alpha_i = 0$ para todo i, es decir, $\mathbf{v} = \mathbf{0}$. Por tanto, $L \cap L^\perp = \{\mathbf{0}\}$.

Sabemos entonces que la suma $L + L^\perp$ es directa. Ahora sólo hay que demostrar que $\dim(L) + \dim(L^\perp) = n$. Sabemos que $\dim(L) = r$, luego queda probar que $\dim(L^\perp) = n - r$. Para ello hay que probar que las r ecuaciones que definen $L^\perp$ son independientes, es decir, que los vectores $\mathbf{u}_1^t A, \mathbf{u}_2^t A, \dots, \mathbf{u}_r^t A$ son linealmente independientes. Vamos a demostrarlo por inducción en r. Si $r = 1$ el resultado es cierto, puesto que $\mathbf{u}_1^t A$ no es un vector nulo (si lo fuera, tendríamos $0 = \mathbf{u}_1^t A \mathbf{u}_1 = f(\mathbf{u}_1, \mathbf{u}_1) = b_1 \neq 0$, una contradicción).

Supongamos entonces que los vectores $\mathbf{u}_1^t A, \mathbf{u}_2^t A, \dots, \mathbf{u}_{r-1}^t A$ son independientes. Para demostrar que al añadir $\mathbf{u}_r^t A$ siguen siendo independientes consideremos el sistema:

$$\left(\begin{array}{c} \mathbf{u}_1^t A \\ \hline \mathbf{u}_2^t A \\ \hline \vdots \\ \hline \mathbf{u}_r^t A \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_r \end{pmatrix}.$$

Este sistema es compatible, puesto que el vector $\mathbf{u}_r$ es una solución. En efecto:

$$\left(\begin{array}{c} \mathbf{u}_1^t A \\ \hline \mathbf{u}_2^t A \\ \hline \vdots \\ \hline \mathbf{u}_r^t A \end{array}\right) \left(\begin{array}{c} \\ \mathbf{u}_r \\ \\ \end{array}\right) = \left(\begin{array}{c} \mathbf{u}_1^t A \mathbf{u}_r \\ \hline \mathbf{u}_2^t A \mathbf{u}_r \\ \hline \vdots \\ \hline \mathbf{u}_r^t A \mathbf{u}_r \end{array}\right) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b_r \end{pmatrix}.$$

Esto quiere decir, según el Teorema de Rouché-Forbenius, que el rango de la matriz de coeficientes coincide con el rango de la ampliada. La matriz ampliada es:

$$A' = \left(\begin{array}{c|c} \mathbf{u}_1^t A & 0 \\ \hline \vdots & \vdots \\ \hline \mathbf{u}_{r-1}^t A & 0 \\ \hline \mathbf{u}_r^t A & b_r \end{array}\right).$$

Las r primeras filas son linealmente independientes por hipótesis de inducción, y la última fila es independiente de las anteriores ya que $b_r \neq 0$. Por tanto, el rango de la matriz ampliada es r, luego el rango de la matriz de coeficientes también es r. Es decir, las filas de la matriz de coeficientes, $\mathbf{u}_1^t A, \ldots, \mathbf{u}_r^t A$ son linealmente independientes, como queríamos demostrar. $\square$

6.3. Diagonalización de formas bilineales simétricas.

Ya podemos demostrar que toda forma bilineal simétrica admite una base ortogonal, es decir, es diagonalizable.

Teorema 6.8 *Dada una aplicación bilineal simétrica $f : V \times V \to K$, existe una base B de V ortogonal respecto de f. Por tanto, la matriz de f respecto de B será diagonal.*

Demostración: Demostraremos el resultado dando un método para encontrar una base ortogonal para f. Buscamos primero un vector $\mathbf{u}_1$ tal que $f(\mathbf{u}_1, \mathbf{u}_1) = b_1 \neq 0$. Si no existe, significa que f es la aplicación nula, por la siguiente razón. Si consideramos cualquier base de V, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ tendremos $f(\mathbf{e}_i, \mathbf{e}_i) = 0$. Pero también $0 = f(\mathbf{e}_i + \mathbf{e}_j, \mathbf{e}_i + \mathbf{e}_j) = f(\mathbf{e}_i, \mathbf{e}_i) + 2f(\mathbf{e}_i, \mathbf{e}_j) + f(\mathbf{e}_j, \mathbf{e}_j) = f(\mathbf{e}_i, \mathbf{e}_j)$ para todo $i \neq j$. Por tanto, la matriz de f es la matriz nula. Podemos tomar entonces cualquier base, y será una base ortogonal.

Supongamos entonces que existe $\mathbf{u}_1 \in V$ tal que $f(\mathbf{u}_1, \mathbf{u}_1) = 0$. Tomaremos $\mathbf{u}_1$ como el primer vector de la base que buscamos. Los demás vectores deben ser, por tanto, ortogonales a $\mathbf{u}_1$. Consideramos entonces $L_1 = \langle \mathbf{u}_1 \rangle$, y buscaremos el resto de los vectores en $L_1^\perp$. Observemos que, por el resultado anterior, $\dim(L_1^\perp) = n - 1$.

Busquemos ahora un vector $\mathbf{u}_2 \in L_1^\perp$ tal que $f(\mathbf{u}_2, \mathbf{u}_2) \neq 0$. Si no existe, entonces podemos usar el razonamiento anterior para demostrar que, tomando cualquier base $\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$ de $L_1^\perp$, se tiene $f(\mathbf{v}_i, \mathbf{v}_j) = 0$ para cualquier i, j. Como además $V = L_1 \oplus L_1^\perp$, tendremos que $\{\mathbf{u}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ es base de V, y que la matriz de f respecto de esta base será

$$\begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Si por el contrario existe un vector $\mathbf{u}_2 \in L_1^\perp$ tal que $f(\mathbf{u}_2, \mathbf{u}_2) = b_2 \neq 0$, tomamos $\mathbf{u}_2$ como segundo elemento de la base buscada, consideramos $L_2 = \langle \mathbf{u}_1, \mathbf{u}_2 \rangle$ y seguimos buscando vectores en $L_2^\perp$.

Seguimos este proceso. Mientras tengamos $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$, tales que $\mathbf{u}_i \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_{i-1}\rangle^\perp$, y $f(\mathbf{u}_i, \mathbf{u}_i) = b_i \neq 0$ para $i = 1, \ldots, r$, definimos $L_r = \langle \mathbf{u}_1, \ldots, \mathbf{u}_r\rangle$ y buscamos $\mathbf{u}_{r+1} \in L_r^\perp$ tal que $f(\mathbf{u}_{r+1}, \mathbf{u}_{r+1}) = b_{r+1} \neq 0$. Si existe, continuamos el proceso. Si no existe, tomamos una base cualquiera B' de $L_r^\perp$, y tendremos que $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \cup B'$ será una base de V tal que la matriz de f respecto de B es de la forma

$$\begin{pmatrix} b_1 & & & & & \\ & \ddots & & & & \\ & & b_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

En cualquier caso, el proceso termina como máximo en n pasos, y se obtiene una base ortogonal para f. $\square$

Nota: En el proceso anterior, no está claro cómo podemos estar seguros de si en $L_r^\perp$ existe un vector $\mathbf{v}$ tal que $f(\mathbf{v}, \mathbf{v}) \neq 0$. Una forma de saberlo es la siguiente: Tomamos cualquier base $B' = \{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ de $L_r^\perp$, y comprobamos los valores $f(\mathbf{v}_i, \mathbf{v}_i)$ y los valores $f(\mathbf{v}_i + \mathbf{v}_j, \mathbf{v}_i + \mathbf{v}_j)$. Si hay alguno no nulo, ya hemos encontrado el vector deseado. Si todos son nulos, entonces tendremos $f(\mathbf{v}_i, \mathbf{v}_j) = 0$ para todo $i, j \in \{r+1, \ldots, n\}$, usando el mismo razonamiento que en la demostración del resultado anterior. Por tanto, para cualquier vector $\mathbf{v} = \alpha_{r+1}\mathbf{v}_{r+1} + \cdots \alpha_n \mathbf{v}_n \in L_r^\perp$, tendremos

$$f(\mathbf{v}, \mathbf{v}) = \sum_{i=r+1}^{n} \sum_{j=r+1}^{n} \alpha_i \alpha_j f(\mathbf{v}_i, \mathbf{v}_j) = 0.$$

Por tanto, en este caso todos los vectores de $L_r^\perp$ son ortogonales a sí mismos. Hemos visto entonces que, para encontrar un vector de $L^\perp$ que no sea ortogonal a sí mismo, basta buscarlo entre los vectores $\mathbf{v}_i$ o los vectores $\mathbf{v}_i + \mathbf{v}_j$, donde $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ es una base cualquiera de $L_r^\perp$.

6.4. Teorema de Sylvester.

Una vez que sabemos que toda matriz de una forma bilineal simétrica es diagonalizable, podemos intentar simplificarla todavía más. Pero para eso necesitamos conocer cuál es el cuerpo K. Trabajaremos con los dos casos más usuales, $\mathbb{C}$ y $\mathbb{R}$.

Si $K = \mathbb{C}$, entonces todo elemento $\alpha \in \mathbb{C}$ admite una raíz cuadrada (ya que la ecuación $x^2 - \alpha = 0$ admite solución en $\mathbb{C}$). Llamaremos $\sqrt{\alpha}$ a cualquiera de las dos raíces cuadradas de α. Supongamos entonces que tenemos una matriz A de una forma bilineal simétrica f. Ya hemos demostrado que, respecto de una cierta base $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, la matriz de f es diagonal, es decir, A es congruente a una matriz de la forma

$$D = \begin{pmatrix} d_{11} & & & \\ & d_{22} & & \\ & & \ddots & \\ & & & d_{nn} \end{pmatrix},$$

donde $d_{ii} = f(\mathbf{u}_i, \mathbf{u}_i)$. Supongamos que $d_{11}, \ldots, d_{rr}$ son no nulos, y que $d_{r+1,r+1} = \cdots = d_{nn} = 0$ (esto es siempre posible si reordenamos la base B de manera que los vectores ortogonales a sí mismos sean los últimos). Entonces podemos considerar la base $B' = \{\frac{\mathbf{u}_1}{\sqrt{d_{11}}}, \ldots, \frac{\mathbf{u}_r}{\sqrt{d_{rr}}}, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$. Observemos que esta base sigue siendo ortogonal, pero además ahora se tiene, para todo $i = 1, \ldots, r$,

$$f\left(\frac{\mathbf{u}_i}{\sqrt{d_{ii}}}, \frac{\mathbf{u}_i}{\sqrt{d_{ii}}}\right) = \frac{1}{d_{ii}}\, f(\mathbf{u}_i, \mathbf{u}_i) = \frac{d_{ii}}{d_{ii}} = 1,$$

mientras que para todo $i > r$, $f(\mathbf{u}_i, \mathbf{u}_i) = d_{ii} = 0$. Por tanto, la matriz de f respecto de B' es de la forma:

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}.$$

El número de unos de esta matriz, r, coincide con el número de elementos distintos de cero de cualquier matriz diagonal congruente con A. Además, este número coincide con el rango de A. Por tanto, lo llamaremos **rango** de la aplicación bilineal f asociada a A. Y lo denotaremos $\operatorname{rg}(f)$.

Por otra parte, si $K = \mathbb{R}$, no todo elemento de $\mathbb{R}$ admite una raíz cuadrada. Sólo los elementos positivos. Por tanto, si los elementos $d_{11}, \ldots, d_{ss}$ son positivos, los elementos $d_{s+1,s+1}, \ldots, d_{rr}$ son negativos, y los elementos $d_{r+1,r+1}, \ldots, d_{nn}$ son nulos, entonces consideramos la base $B' = \{\frac{\mathbf{u}_1}{\sqrt{d_{11}}}, \ldots, \frac{\mathbf{u}_s}{\sqrt{d_{ss}}}, \frac{\mathbf{u}_{s+1}}{\sqrt{|d_{s+1,s+1}|}}, \ldots, \frac{\mathbf{u}_r}{\sqrt{|d_{rr}|}}, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$. En este caso tendremos, para $1 \leq i \leq s$,

$$f\left(\frac{\mathbf{u}_i}{\sqrt{d_{ii}}}, \frac{\mathbf{u}_i}{\sqrt{d_{ii}}}\right) = \frac{1}{d_{ii}}\, f(\mathbf{u}_i, \mathbf{u}_i) = \frac{d_{ii}}{d_{ii}} = 1,$$

para $s + 1 \leq i \leq r$,

$$f\left(\frac{\mathbf{u}_i}{\sqrt{|d_{ii}|}}, \frac{\mathbf{u}_i}{\sqrt{|d_{ii}|}}\right) = \frac{1}{|d_{ii}|}\, f(\mathbf{u}_i, \mathbf{u}_i) = \frac{d_{ii}}{|d_{ii}|} = -1,$$

y para $i > r$, $f(\mathbf{u}_i, \mathbf{u}_i) = d_{ii} = 0$. Por tanto, si $K = \mathbb{R}$, la matriz de f respecto de B' queda de la forma:

$$\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix}.$$

En este caso, el número de unos de esta diagonal, s, que coincide con el número de elementos positivos en cualquier diagonal congruente con A, se llama **signatura** de A, o de f, y se denota $\mathrm{sig}(A)$ o $\mathrm{sig}(f)$. Al igual que antes, el número de elementos distintos de cero es el **rango** de f, denotado $\mathrm{rg}(f)$, que coincide con $\mathrm{rg}(A)$.

De este razonamiento se deduce el siguiente resultado:

Teorema 6.9 (Teorema de Sylvester) *Sean $A, B \in \mathcal{M}_{n\times n}(K)$.*
Si $K = \mathbb{C}$, entonces A y B son congruentes si y sólo si $rg(A) = rg(B)$.
Si $K = \mathbb{R}$, entonces A y B son congruentes si y sólo si $rg(A) = rg(B)$ y $sig(A) = sig(B)$.

Demostración: El caso $K = \mathbb{C}$ es evidente: Dos matrices congruentes deben tener el mismo rango, ya que se pasa de la una a la otra multiplicándolas por matrices no singulares. Además, como toda matriz sobre $\mathbb{C}$ es congruente a una que sólo tenga unos y ceros en su diagonal, donde el número de unos es el rango de la matriz, se obtiene el resultado.

Para el caso $K = \mathbb{R}$, debemos probar que la signatura de una matriz está bien definida. Es decir, que una matriz A sólo puede ser congruente a una única matriz diagonal cuyos elementos sean $1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0$. Supongamos que A es congruente a dos matrices D y D' de esta forma, con $p = \mathrm{sig}(D)$ y $p' = \mathrm{sig}(D')$. Debemos probar que $p = p'$.

Sabemos que las matrices D y D' representan a la misma aplicación bilineal, f, respecto de dos bases distintas, $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ y $B' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$. Consideremos las variedades lineales: $L_1 = \langle \mathbf{v}_1, \ldots, \mathbf{v}_p \rangle$, y $L_2 = \langle \mathbf{v}'_{p'+1}, \ldots, \mathbf{v}'_n \rangle$. Usando la matriz D, sabemos que para todo vector no nulo $\mathbf{v} \in L_1$, de coordenadas $(a_1, \ldots, a_p, 0, \ldots, 0)_B$, se tiene $f(\mathbf{v}, \mathbf{v}) = a_1^2 + \cdots + a_p^2 > 0$. Por otra parte, usando la matriz D', sabemos que para todo vector $\mathbf{v} \in L_2$ de coordenadas $(0, \ldots, 0, a_{p'+1}, \ldots, a_n)_{B'}$, se tiene $f(\mathbf{v}, \mathbf{v}) = -a_{p'+1}^2 - \cdots - a_r^2 \leq 0$, donde r es el rango de A. Por tanto, $L_1 \cap L_2 = \{\mathbf{0}\}$, ya que si tuviéramos un vector no nulo $\mathbf{v} \in L_1 \cap L_2$, tendríamos $f(\mathbf{v}, \mathbf{v}) > 0$ y al mismo tiempo $f(\mathbf{v}, \mathbf{v}) \leq 0$, lo cual es imposible.

Por tanto, si $L_1 \cap L_2 = \{0\}$, por la fórmula de la dimensión se tiene:

$$p + (n - p') = \dim(L_1) + \dim(L_2) = \dim(L_1 + L_2) \leq \dim(V) = n.$$

Es decir, $p - p' \leq 0$. Pero si ahora invertimos los papeles de p y p', y hacemos un razonamiento análogo, obtendremos $p' - p \leq 0$. En definitiva, $p - p' = 0$, con lo que $p = p'$, y la signatura de A está bien definida. Esto implica que la matriz diagonal formada por unos, menos unos, y ceros, congruente a A es única, lo que demuestra el teorema. $\square$

6.5. Espacios vectoriales euclídeos.

Terminaremos esta asignatura aplicando lo aprendido sobre aplicaciones bilineales simétricas, para definir un *producto escalar* en un espacio vectorial. Esto nos va a permitir generalizar, a espacios vectoriales abstractos, conceptos bien conocidos de los espacios vectoriales $\mathbb{R}^2$ o $\mathbb{R}^3$, como son los ángulos entre vectores o la longitud de un vector.

Eso sí, para que todo funcione debidamente, el cuerpo de escalares con el que trataremos será $\mathbb{R}$. Es decir, a partir de ahora $K = \mathbb{R}$. Recordemos que el producto escalar en $\mathbb{R}^2$ o $\mathbb{R}^3$ es una aplicación que a dos vectores $\mathbf{u}$ y $\mathbf{v}$ les hace corresponder un escalar (en este caso un número real), que se suele denotar $\mathbf{u} \cdot \mathbf{v}$. Las propiedades principales de este producto escalar, que nos van a servir para definir el producto escalar en un espacio vectorial cualquiera, son la siguientes:

Producto escalar: Sea V un espacio vectorial sobre $\mathbb{R}$. Una aplicación $(\,\cdot\,) : V \times V \to \mathbb{R}$, que asocia al par $(\mathbf{u}, \mathbf{v})$ el escalar $\mathbf{u}\cdot\mathbf{v}$ es un **producto escalar** si para todo $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ y todo $\alpha \in \mathbb{R}$ se tiene:

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.
2. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.
3. $(\alpha\mathbf{u}) \cdot \mathbf{v} = \alpha(\mathbf{u} \cdot \mathbf{v})$.
4. $\mathbf{u} \cdot \mathbf{u} > 0$ si $\mathbf{u} \neq 0$.

Observemos que si una aplicación satisface las tres primeras propiedades, entonces es una forma bilineal simétrica sobre V. La cuarta propiedad tiene también nombre propio:

Forma bilineal definida positiva: Una forma bilineal $f : V \times V \to \mathbb{R}$ se dice **definida positiva** si $f(\mathbf{u}, \mathbf{u}) > 0$ para todo $\mathbf{u} \in V$, $\mathbf{u} \neq 0$.

Por tanto, tenemos una forma equivalente para definir un producto escalar sobre V:

Producto escalar: Sea V un espacio vectorial sobre $\mathbb{R}$. Un **producto escalar** sobre V es una forma bilineal simétrica definida positiva.

Ejemplo 6.10 *En $\mathbb{R}^n$, si consideramos la matriz identidad I, esta define una forma bilineal simétrica (ya que la matriz I es simétrica). Además, para todo vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, si aplicamos la forma asociada a I al par de vectores $(\mathbf{v}, \mathbf{v})$, obtenemos $\mathbf{v}^t I \mathbf{v} = v_1^2 + \cdots + v_n^2$. Este número es siempre mayor que cero si $\mathbf{v} \neq 0$. Por tanto, la forma bilineal determinada por la matriz I es un producto escalar. De hecho, es el producto escalar usual, que a dos vectores $\mathbf{u} = (u_1, \ldots, u_n)$ y $\mathbf{v} = (v_1, \ldots, v_n)$ asocia el escalar $u_1v_1 + u_2v_2 + \cdots + u_nv_n$.*

El producto escalar definido en este ejemplo se usa en $\mathbb{R}^n$ para determinar el ángulo entre dos vectores, o el tamaño (o módulo) de un vector. De hecho, se tienen las conocidas fórmulas:

1. **Módulo de un vector:** $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$.
2. **Fórmula del coseno:** Si α es el ángulo que forman $\mathbf{u}$ y $\mathbf{v}$, se tiene:

$$\cos \alpha = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}|\,|\mathbf{v}|}.$$

Estas dos ecuaciones se pueden usar, por tanto, para definir el módulo de un vector, o el ángulo entre dos vectores en un espacio vectorial abstracto V, donde hayamos definido un producto escalar. Es por eso que se tiene la siguiente definición:

Un **espacio vectorial euclídeo**, $(V, \cdot)$ es un espacio vectorial V sobre $\mathbb{R}$, dotado de un producto escalar $(\ \cdot\ )$.

Algunas propiedades importantes de un espacio vectorial euclídeo son las siguientes:

Proposición 6.11 *Sea $(V, \cdot)$ un espacio vectorial euclídeo. Para todo $\mathbf{u}, \mathbf{v} \in V$ y todo $\alpha \in \mathbb{R}$, se tiene:*

1. $\mathbf{v} = 0 \Leftrightarrow |\mathbf{v}| = 0$.
2. $|\alpha\mathbf{v}| = |\alpha||\mathbf{v}|$.
3. **Desigualdad de Cauchy-Schwartz:** $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}|\, |\mathbf{v}|$.
4. **Desigualdad triangular:** $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$.

Demostración: Las dos primeras propiedades se demuestran de forma directa. Para probar la desigualdad de Cauchy-Schwartz, como se trata de números reales positivos, probaremos que sus cuadrados satisfacen la desigualdad. Es decir, probaremos que $(\mathbf{u}\cdot\mathbf{v})^2 \leq \mathbf{u}^2\, \mathbf{v}^2$, donde el cuadrado de un vector $\mathbf{v}$ significa $\mathbf{v} \cdot \mathbf{v}$, es decir, $|\mathbf{v}|^2$. Hay que distinguir dos casos. En primer lugar, si $\mathbf{u}$ y $\mathbf{v}$ son linealmente dependientes, es decir, si $\mathbf{u} = \alpha\mathbf{v}$, entonces se tiene:

$$(\mathbf{u} \cdot \mathbf{v})^2 = ((\alpha\mathbf{v}) \cdot \mathbf{v})^2 = (\alpha(\mathbf{v} \cdot \mathbf{v}))^2 = \alpha^2(\mathbf{v} \cdot \mathbf{v})^2 = \alpha^2\mathbf{v}^2\mathbf{v}^2 = (\alpha\mathbf{v})^2\mathbf{v}^2 = \mathbf{u}^2\mathbf{v}^2.$$

Sin embargo, si $\mathbf{u}$ y $\mathbf{v}$ son linealmente independientes, entonces $\mathbf{u} + \alpha\mathbf{v} \neq 0$ para todo $\alpha \in \mathbb{R}$. Por tanto, $(\mathbf{u} + \alpha\mathbf{v})^2 > 0$, con lo que tendremos:

$$(\mathbf{u} + \alpha\mathbf{v})^2 = \mathbf{u}^2 + 2\alpha(\mathbf{u} \cdot \mathbf{v}) + \alpha^2\mathbf{v}^2 > 0,$$

para todo número real α. Esto quiere decir que, si consideramos la expresión anterior como una ecuación de segundo grado con incógnita α, esta ecuación no tiene solución real. Por tanto, el discriminante de esta ecuación debe ser menor que cero, es decir:

$$4(\mathbf{u} \cdot \mathbf{v})^2 - 4\mathbf{u}^2\mathbf{v}^2 < 0 \quad \Rightarrow \quad (\mathbf{u} \cdot \mathbf{v})^2 < \mathbf{u}^2\, \mathbf{v}^2.$$

Por último, debemos demostrar la desigualdad triangular. Como los módulos $|\mathbf{u} + \mathbf{v}|$, $|\mathbf{u}|$ y $|\mathbf{v}|$ son números reales positivos, sólo hay que demostrar que sus cuadrados satisfacen la desigualdad. Se tiene:

$$(|\mathbf{u} + \mathbf{v}|)^2 = (\mathbf{u} + \mathbf{v})^2 = \mathbf{u}^2 + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v}^2 = |\mathbf{u}|^2 \pm 2|\mathbf{u} \cdot \mathbf{v}| + |\mathbf{v}|^2 \leq |\mathbf{u}|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + |\mathbf{v}|^2$$

Por la desigualdad de Cauchy-Schwartz, se tiene:

$$|\mathbf{u}|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + |\mathbf{v}|^2 \leq |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}| + |\mathbf{v}|^2 = (|\mathbf{u}| + |\mathbf{v}|)^2.$$

Por tanto, $(|\mathbf{u} + \mathbf{v}|)^2 \leq (|\mathbf{u}| + |\mathbf{v}|)^2$, luego $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$, como queríamos probar. $\square$

Recordemos ahora que toda forma bilineal simétrica sobre $\mathbb{R}$ se puede diagonalizar, de manera que su matriz, respecto de una base adecuada, sea de la forma:

$$D = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix}.$$

Pero si la forma bilineal es un producto escalar, es decir, si es definida positiva, entonces se tiene $D = I$, ya que si hubiera algún -1 o algún 0 en la diagonal de la matriz, tendríamos algún elemento de la base, $\mathbf{v}$, tal que $\mathbf{v} \cdot \mathbf{v} = -1$ o $\mathbf{v} \cdot \mathbf{v} = 0$, lo que contradice que f sea definida positiva. Por tanto, todo producto escalar se puede considerar, respecto de una base adecuada, como el producto escalar usual de $\mathbb{R}^n$. Este tipo de bases tienen un nombre específico:

Bases ortogonales y ortonormales: Sea $(V, \cdot)$ un espacio vectorial euclídeo. Se dice que una base $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V es **ortogonal** si $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ para todo $i \neq j$. Se dirá además que es una base **ortonormal** si $\mathbf{v}_i \cdot \mathbf{v}_i = 1$, para todo i.

Dicho de otra manera, una base es ortogonal si todos sus vectores son perpendiculares entre sí (el producto escalar es cero), y será ortonormal si todos los vectores, además de ser perpendiculares entre sí, tienen módulo 1. Por cierto, los vectores que tienen módulo 1 se llaman **unitarios**, luego una base ortonormal es una base de vectores unitarios mutuamente ortogonales. Con estas definiciones, y con la diagonalización que conocemos de cualquier producto escalar, se tiene:

Proposición 6.12 *Todo espacio vectorial euclídeo $(V, \cdot)$ admite una base ortonormal.*

Demostración: Sólo hay que diagonalizar la matriz del producto escalar, como en la sección anterior, para obtener la matriz identidad. La matriz de paso nos dará una base ortonormal, ya que las entradas de la matriz I son exactamente los productos escalares de los vectores de la base. $\square$

6.6. Variedades ortogonales. Método de Gram-Schmidt.

Hemos visto cómo, en un espacio vectorial euclídeo V, podemos obtener una base ortonormal. Pero existen muchas bases ortogonales posibles, y puede que nos interese encontrar alguna en particular. Más concretamente, si tenemos una variedad lineal L en V, nos puede interesar encontrar una base ortonormal de L, para completarla hasta una base ortonormal de V. En esta última sección veremos que esto es siempre posible, y además usando este método obtendremos la variedad lineal *ortogonal* a L, es decir, la variedad $L^\perp$ que definimos anteriormente. Con las notaciones del producto escalar, la definición queda:

Variedad ortogonal: Sea L una variedad lineal de un espacio vectorial euclídeo $(V, \cdot)$. Se define la **variedad ortogonal** a L, que denotamos $L^\perp$, como el conjunto de los vectores ortogonales a todos los de L. Es decir:

$$L^\perp = \{\mathbf{v} \mid \quad \mathbf{u} \cdot \mathbf{v} = 0, \ \forall \mathbf{u} \in L\}.$$

Ejemplo 6.13 *En $\mathbb{R}^3$ con el producto vectorial usual, si L es un plano que pasa por el origen, $L^\perp$ será la recta perpendicular a L que pasa por el origen.*

A continuación estudiaremos un método para obtener una base ortonormal de cualquier variedad L, y también una base ortonormal de $L^\perp$. Se trata del método de Gram-Schmidt. Comenzamos viendo un resultado que usaremos repetidamente en el método:

Proposición 6.14 *Sea L una variedad lineal de un espacio vectorial euclídeo, y supongamos que $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ es una base ortogonal de L. Entonces dado $\mathbf{v} \notin L$, existe un vector $\mathbf{v}'$ tal que $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}'\}$ es una base ortogonal de $\langle \mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}\rangle$.*

DEMOSTRACIÓN: Simplemente tomamos $\mathbf{v}' = \mathbf{v} + a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$, donde $a_1, \ldots, a_k$ son unos escalares apropiados tales que $\mathbf{v}' \cdot \mathbf{v}_i = 0$ para todo $i = 1, \ldots, k$. Más concretamente, $a_i = -\dfrac{\mathbf{v} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$. Se demuestra directamente que con estos datos el resultado se verifica.
□

Teorema 6.15 (Método de ortonormalización de Gram-Schmidt) *Dada una variedad lineal L de un espacio vectorial euclídeo $(V, \cdot)$, existe una base ortonormal $B_L = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ de L. Además, B_L se puede completar hasta una base ortonormal $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V, donde $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ es una base ortonormal de $L^\perp$.*

Demostración: Probaremos la existencia de B_L por inducción en $k = \dim(L)$. Si $k = 1$, entonces admite una base formada por un sólo vector, $\mathbf{u}_1$. Dividiendo este vector por su módulo, obtenemos el vector $\mathbf{v}_1 = \frac{\mathbf{u}_1}{|\mathbf{u}_1|}$, que es unitario, luego $B_L = \{\mathbf{v}_1\}$.

Si $k > 1$, y suponemos el resultado cierto para variedades lineales de dimensión $k-1$, consideremos una base $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ de L. La variedad lineal generada por $\{\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}\}$ admite, por tanto, una base ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$. Esta base será, en particular, ortogonal, luego podemos aplicar el resultado anterior a esta base y a $\mathbf{u}_k$, y obtendremos un vector $\mathbf{u}'_k$, tal que $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{u}'_k\}$ es una base ortogonal de L. Si ahora dividimos $\mathbf{u}'_k$ por su módulo, obtenemos el vector unitario $\mathbf{v}_k = \frac{\mathbf{u}'_k}{|\mathbf{u}'_k|}$, tal que $B_L = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ es una base ortonormal de L, como queríamos demostrar.

La forma de ampliar la base B_L a una base ortonormal de todo V es exactamente la misma: ampliamos B_L a una base cualquiera de V, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$, y vamos transformando progresivamente cada $\mathbf{u}_j$ por un vector unitario $\mathbf{v}_j$, que será ortogonal a todos los anteriores. De esta forma llegaremos a una base ortonormal B de V.

Por último, los vectores $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ forman una base ortonormal de $L^\perp$, por lo siguiente: son ortonormales, y linealmente independientes ya que pertenecen a una base ortonormal de V; Son vectores de $L^\perp$ ya que cada uno de ellos es ortogonal a todos los $\mathbf{v}_i$, con $i = 1, \ldots, k$; Finalmente, sabemos que $\dim(L^\perp) = n - k$, luego son sistema de generadores. Por tanto, forman una base ortonormal de $L^\perp$, como queríamos probar. $\square$

La forma más eficaz de usar el **método de Gram-Schmidt**, para transformar una base cualquiera $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ de V en una base ortonormal es la siguiente:

1. Llamamos $\mathbf{u}'_1 = \mathbf{u}_1$.
2. Si ya hemos sustituido $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ por $\mathbf{u}'_1, \ldots, \mathbf{u}'_{k-1}$, sustituimos $\mathbf{u}_k$ por
$$\mathbf{u}'_k = \mathbf{u}_k + a_{k1}\mathbf{u}'_1 + \cdots + a_{k\,k-1}\mathbf{u}'_{k-1},$$
donde $a_{kj} = -\frac{\mathbf{u}_k \cdot \mathbf{u}'_j}{\mathbf{u}'_j \cdot \mathbf{u}'_j}$, para todo $j < k$. Cuando $k = n$, habremos conseguido una base ortogonal $B' = \{\mathbf{u}'_1, \ldots, \mathbf{u}'_n\}$.
3. Dividimos cada vector de B' por su módulo, y conseguiremos una base ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ de V.

Además, usando este método, el sistema $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ genera el mismo subespacio que el sistema ortonormal $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, para todo $k = 1, \ldots, n$.

Terminaremos este tema, y por tanto esta asignatura, dando un criterio para determinar cuándo una matriz simétrica determina un producto escalar (es decir, determina una aplicación bilineal definida positiva), sin tener que diagonalizarla. Tomemos por tanto una matriz $A \in \mathcal{M}_{n\times n}(\mathbb{R})$. Para todo $k = 1, \dots, n$, vamos a denotar $A_{(k)}$ a la matriz menor de A formada por las filas $1, \dots, k$ y las columnas $1, \dots, k$. Es decir,

$$A_{(k)} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}.$$

Entonces se tiene:

Proposición 6.16 *Una matriz simétrica $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ define un producto escalar si y sólo si $|A_{(k)}| > 0$ para todo $k = 1, \dots, n$.*

Demostración: Supongamos que A define un producto escalar. En ese caso tendremos $P^tAP = I$, donde P es el cambio de base usado en el método de Gram-Schmidt. Por tanto, P es una matriz triangular superior, no singular. Si llamamos $Q = P^{-1}$, entonces Q también será triangular superior, no singular, y tendremos $A = Q^tIQ = Q^tQ$. Por tanto, A será el producto de una matriz triangular inferior por una triangular superior. Si analizamos el producto Q^tQ, vemos que el menor $A_{(k)}$, para todo k, es precisamente $A_{(k)} = Q^t_{(k)}Q_{(k)}$. Y su determinante será: $|A_{(k)}| = |Q^t_{(k)}||Q_{(k)}| = |Q_{(k)}|^2 > 0$.

Recíprocamente, supongamos que $|A_{(k)}| > 0$ para todo $k = 1, \dots, n$, y llamemos f a la forma bilineal definida por A. Vamos a construir una base ortogonal $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ tal que $f(\mathbf{u}_i, \mathbf{u}_i) > 0$ para todo i, por lo que A define un producto escalar.

En concreto, si llamamos $(A_{(i)})_{jk}$ al adjunto del elemento (j, k) de la matriz $A_{(i)}$, vamos a definir los vectores:

$$\mathbf{u}_i = ((A_{(i)})_{1i}, (A_{(i)})_{2i} \dots, (A_{(i)})_{ii}, 0, \dots, 0).$$

Observemos que la última coordenada no nula de este vector es $(A_{(i)})_{ii} = |A_{(i-1)}| > 0$. Por tanto, los vectores $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ forman una base de V, puesto que forman una matriz triangular cuya diagonal está formada por elementos no nulos.

Tenemos que demostrar entonces que los vectores $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ son ortogonales entre sí, y que $f(\mathbf{u}_i, \mathbf{u}_i) > 0$. Para ello, observemos que

$$\mathbf{u}_i^tA = (0, \dots, 0, |A_{(i)}|, 0, \dots, 0).$$

Esto es debido a que, si $j = i$, el producto de $\mathbf{u}_i^t$ por la columna j de A es igual al desarrollo del determinante de $A_{(i)}$ por la columna i. Pero si $j \neq i$, este producto es el desarrollo del determinante de una matriz con dos columnas iguales, luego es nulo.

Esto implica que $\mathbf{u}_i^t A \mathbf{u}_j = 0$ para todo $j < i$, luego la base es ortogonal. Pero además $\mathbf{u}_i^t A \mathbf{u}_i = |A_{(i)}||A_{(i-1)}|$, que por hipótesis es un número real positivo. Por tanto, $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ satisface las propiedades requeridas, y A define un producto escalar. $\square$