Capítulo III

Aplicaciones lineales

El objetivo principal de este capítulo es el estudio de las propiedades de las aplicaciones de espacios lineales. En los parágrafos 8 y 10, así como en los puntos 9.1, 11.1 y 11.2, el cuerpo principal no se somete a ninguna restricción. En los puntos 9.2, 9.3, 11.3 y 11.4 y también en el § 12 se supone que el cuerpo principal es un cuerpo conmutativo. En el § 7, de carácter de introducción, se establece una serie de conceptos y de propiedades relacionados con aplicaciones de conjuntos absolutamente arbitrarios.

§ 7. Aplicaciones de conjuntos arbitrarios

7.1. Producto de aplicaciones. Consideremos un conjunto $\mathfrak{M}$ de entes arbitrarios. Este conjunto puede estar compuesto tanto por un número finito como infinito de elementos. Se llama *aplicación* del conjunto $\mathfrak{M}$ toda ley que permite a partir de cualquier elemento del conjunto $\mathfrak{M}$ encontrar de nuevo un elemento de $\mathfrak{M}$. Convendremos en indicar las aplicaciones por las letras $\mathcal{A}$, $\mathcal{B}$, ... Si m es un elemento de $\mathfrak{M}$, indicaremos por $m\mathcal{A}$ aquel elemento del conjunto $\mathfrak{M}$ que se obtiene del elemento m mediante la aplicación $\mathcal{A}$. El elemento $m\mathcal{A}$ se denomina *imagen* del elemento m en la aplicación $\mathcal{A}$, mientras que m se denomina *imagen recíproca* del elemento $m\mathcal{A}$.

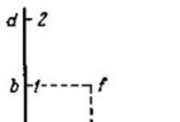

Fig. 2.

Consideremos, a título de ejemplo, el conjunto de todos los puntos del plano y sea $\mathfrak{U}$ la aplicación que consiste en el giro de los puntos de este plano alrededor de uno de sus puntos O en 90° en contra del movimiento de las agujas del reloj (fig. 2).

Recurriendo a la figura, vemos que

$$a\mathfrak{U}=b \quad \text{y} \quad c\mathfrak{U}=d.$$

Análogamente, si $\mathcal{D}$ significa el traslado de los puntos en una unidad paralelamente al eje Oa, se tiene.

$$a\mathcal{D}=c \quad \text{y} \quad b\mathcal{D}=f.$$

Dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ del conjunto $\mathfrak{M}$ se llaman *iguales*, si para todo elemento m de $\mathfrak{M}$ se tiene.

$$m\mathcal{A}=m\mathcal{B}.$$

Uno de los conceptos principales de la teoría de aplicaciones es el de producto de aplicaciones que se introduce del modo siguiente. Sean $\mathcal{A}$ y $\mathcal{B}$ dos aplicaciones del conjunto $\mathfrak{M}$. La primera transforma un elemento arbitrario m del conjunto $\mathfrak{M}$ en $m\mathcal{A}$. Si este nuevo elemento se somete a la aplicación $\mathcal{B}$, se obtendrá el elemento $(m\mathcal{A})\mathcal{B}$. La aplicación que transforma el elemento m directamente en $(m\mathcal{A})\mathcal{B}$ se llama *producto* de $\mathcal{A}$ por $\mathcal{B}$ y se designa $\mathcal{A}\mathcal{B}$. Es decir, se toma por definición que

$$m(\mathcal{A}\mathcal{B})=(m\mathcal{A})\mathcal{B}.$$

Si al elemento m se aplica primero la aplicación $\mathcal{B}$ y después la aplicación $\mathcal{A}$, se obtendrá el elemento $(m\mathcal{B})\mathcal{A}$ que puede no coincidir con $(m\mathcal{A})\mathcal{B}$. Efectivamente, en el ejemplo con el plano, considerado anteriormente, se tiene

$$a(\mathcal{U}\mathcal{D})=(a\mathcal{U})\mathcal{D}=b\mathcal{D}=f,$$
$$a(\mathcal{D}\mathcal{U})=(a\mathcal{D})\mathcal{U}=c\mathcal{U}=d$$

y, por consiguiente, $\mathcal{U}\mathcal{D}\neq\mathcal{D}\mathcal{U}$. Por lo tanto, *el producto de aplicaciones depende, en general, del orden de los factores*. Vemos, pues, que una de las leyes principales que se verifica para el producto de los números no se cumple para las aplicaciones. Sin embargo, la otra ley principal—*la asociatividad de la multiplicación*—se conserva para las aplicaciones. En efecto, sean $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$ aplicaciones arbitrarias del conjunto $\mathfrak{M}$ y sea m uno de sus elementos. Por definición, tenemos

$$m((\mathcal{A}\mathcal{B})\mathcal{C})=(m(\mathcal{A}\mathcal{B}))\mathcal{C}=((m\mathcal{A})\mathcal{B})\mathcal{C},$$
$$m(\mathcal{A}(\mathcal{B}\mathcal{C}))=(m\mathcal{A})(\mathcal{B}\mathcal{C})=((m\mathcal{A})\mathcal{B})\mathcal{C},$$

de donde

$$(\mathcal{A}\mathcal{B})\mathcal{C}=\mathcal{A}(\mathcal{B}\mathcal{C}).$$

Empleando esta ley es fácil deducir que *el producto de cualquier número finito de aplicaciones, tomadas en un orden determinado, no depende de la disposición de los paréntesis*. Así, por ejemplo,

$$(\mathcal{A}\mathcal{B})(\mathcal{C}\mathcal{D})=((\mathcal{A}\mathcal{B})\mathcal{C})\mathcal{D}=(\mathcal{A}(\mathcal{B}\mathcal{C}))\mathcal{D}=\mathcal{A}((\mathcal{B}\mathcal{C})\mathcal{D}).$$

Por esto, en los productos que contienen varios factores se pueden omitir los paréntesis y se puede hablar simplemente del producto de dos, de tres y de un número mayor de aplicaciones. El producto de n factores iguales a $\mathcal{A}$ se llama n-ésima potencia de la aplicación $\mathcal{A}$ y se indica por $\mathcal{A}^n$. Las operaciones con las potencias se realizan siguiendo las reglas corrientes

$$\mathcal{A}^m\mathcal{A}^n = \mathcal{A}^{m+n}, \tag{1}$$
$$(\mathcal{A}^m)^n = \mathcal{A}^{mn}; \tag{2}$$

la demostración de las mismas es evidente.

Si el producto de dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ no depende del orden de los factores, se dice que $\mathcal{A}$ y $\mathcal{B}$ son *permutables* o que *conmutan*. La fórmula (1) señala que *las potencias de una misma aplicación son permutables.*

Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ conmutan, se tiene

$$(\mathcal{A}\mathcal{B})^2 = \mathcal{A}\mathcal{B}\mathcal{A}\mathcal{B} = \mathcal{A}\mathcal{A}\mathcal{B}\mathcal{B} = \mathcal{A}^2\mathcal{B}^2$$

y, en general,

$$(\mathcal{A}\mathcal{B})^n = \mathcal{A}^n\mathcal{B}^n. \tag{3}$$

En cambio, si $\mathcal{A}$ y $\mathcal{B}$ no conmutan, la fórmula (3) puede no tener lugar.

7.2. Las aplicaciones idéntica e inversa. Entre todas las aplicaciones de un conjunto $\mathfrak{M}$ desempeña un papel especial la aplicación que pone en correspondencia a todo elemento m de este conjunto el propio elemento m. Esta aplicación lleva el nombre de aplicación *unidad* o *idéntica* y será indicada en adelante por $\mathcal{E}$. Es decir, para todo m se tiene

$$m\mathcal{E} = m.$$

Sea $\mathcal{A}$ una aplicación arbitraria del conjunto $\mathfrak{M}$. Puesto que

$$m(\mathcal{E}\mathcal{A}) = (m\mathcal{E})\mathcal{A} = m\mathcal{A}$$

y

$$m(\mathcal{A}\mathcal{E}) = (m\mathcal{A})\mathcal{E} = m\mathcal{A},$$

se tiene

$$\mathcal{E}\mathcal{A} = \mathcal{A}\mathcal{E} = \mathcal{A}.$$

Si dada la aplicación $\mathcal{A}$ se puede hallar una aplicación $\mathcal{B}$ tal que

$$\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{E}, \tag{4}$$

se dice que $\mathcal{B}$ es la *inversa* de $\mathcal{A}$, mientras que $\mathcal{A}$ se denomina *invertible*. Es fácil ver que *toda aplicación invertible tiene sólo una inversa*. Efectivamente, si $\mathcal{A}$ posee dos aplicaciones inversas $\mathcal{B}$ y $\mathcal{C}$, entonces, multiplicando la relación

$$\mathcal{A}\mathcal{C} = \mathcal{E}$$

a la izquierda por $\mathcal{B}$ y empleando las igualdades (4) y la asociatividad de la multiplicación de las aplicaciones, obtenemos $\mathcal{E}\mathcal{C}=\mathcal{B}\mathcal{E}$, es decir, $\mathcal{C}=\mathcal{B}$.

La aplicación inversa de $\mathcal{A}$ se indica por $\mathcal{A}^{-1}$. Las relaciones (4) son simétricas respecto a $\mathcal{A}$ y $\mathcal{B}$; por esto, si $\mathcal{B}$ es la inversa de $\mathcal{A}$ resulta que $\mathcal{A}$ es la inversa de $\mathcal{B}$, es decir,

$$(\mathcal{A}^{-1})^{-1}=\mathcal{A}. \tag{5}$$

Tomemos por definición

$$\mathcal{A}^0=\mathcal{E} \quad \text{y} \quad \mathcal{A}^{-n}=(\mathcal{A}^{-1})^n \quad (n=1,\ 2,\ 3,\ \ldots).$$

De (4) y (5) se deduce fácilmente que las fórmulas (1) y (2) son válidas no sólo para exponentes enteros positivos, sino para todos los exponentes enteros. En particular,

$$(\mathcal{A}^n)^{-1}=(\mathcal{A}^{-1})^n=\mathcal{A}^{-n}.$$

Además, la relación

$$\mathcal{A}\mathcal{B}\cdot\mathcal{B}^{-1}\mathcal{A}^{-1}=\mathcal{E}$$

implica que

$$(\mathcal{A}\mathcal{B})^{-1}=\mathcal{B}^{-1}\mathcal{A}^{-1}.$$

7.3. Aplicaciones biyectivas. No toda aplicación es invertible. El teorema que sigue ofrece un criterio simple de aplicaciones invertibles.

TEOREMA *Para que una aplicación $\mathcal{A}$ de un conjunto $\mathfrak{M}$ sea invertible es necesario y suficiente que $\mathcal{A}$ sea una aplicación biyectiva del conjunto $\mathfrak{M}$ sobre sí mismo, es decir, que para todo elemento de $\mathfrak{M}$ exista en $\mathfrak{M}$ su imagen recíproca y que distintos elementos de $\mathfrak{M}$ se transformen por la aplicación $\mathcal{A}$ en distintos elementos.*

Demostremos primero la necesidad. Supongamos que $\mathcal{A}$ posee la aplicación inversa $\mathcal{B}$, de manera que

$$\mathcal{A}\mathcal{B}=\mathcal{B}\mathcal{A}=\mathcal{E}.$$

Tomemos en $\mathfrak{M}$ un elemento cualquiera m y sea $m\mathcal{B}=n$. Multiplicando esta relación por $\mathcal{A}$ y sustituyendo $\mathcal{B}\mathcal{A}$ por $\mathcal{E}$, obtendremos $m=n\mathcal{A}$, es decir, todo elemento m de $\mathfrak{M}$ es la imagen de un elemento n de $\mathfrak{M}$. Por otro lado, si dos elementos m_1 y m_2 son transformados por la aplicación $\mathcal{A}$ en un mismo elemento

$$m_1\mathcal{A}=m_2\mathcal{A},$$

entonces, multiplicando esta relación por $\mathcal{B}$, obtenemos $m_1=m_2$. Por consiguiente, todo elemento de $\mathfrak{M}$ tiene en $\mathfrak{M}$ sólo una imagen recíproca.

Demostremos ahora la suficiencia de las condiciones. Según éstas, para todo elemento m del conjunto $\mathfrak{M}$ existe un elemento, y sólo uno, n tal que

$$n\mathcal{A}=m \tag{6}$$

Indiquemos por $\mathcal{B}$ la aplicación que transforma m en n. Es decir,

$$m\mathcal{B}=n. \tag{7}$$

Multiplicando (7) por $\mathcal{A}$ y empleando la igualdad (6), obtenemos

$$m(\mathcal{B}\mathcal{A})=n\mathcal{A}=m.$$

Puesto que m es un elemento arbitrario, de aquí se tiene $\mathcal{B}\mathcal{A}=\mathcal{E}$. Análogamente, multiplicando (6) por $\mathcal{B}$ y empleando (7), obtenemos $\mathcal{A}\mathcal{B}=\mathcal{E}$. Por consiguiente, $\mathcal{B}$ es la aplicación inversa deseada.

7.4. Sustituciones. Las aplicaciones de conjuntos finitos se representan, generalmente, mediante tablas, colocando en la primera fila de las mismas los símbolos de los elementos del conjunto dado en cierto orden y debajo de ellos los símbolos de los elementos que les *corresponden*. Por ejemplo,

$$\sigma=\begin{pmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{pmatrix}$$

es aquella aplicación del conjunto de los números 1, 2, y 3 en la que el 1 se transforma en el 3, el 2 en el 1 y el 3 en el 2 o, empleando la notación aceptada anteriormente,

$$1\sigma=3,\quad 2\sigma=1\quad \text{y}\quad 3\sigma=2.$$

Las aplicaciones invertibles de conjuntos finitos se llaman *sustituciones*. Para que una aplicación representada en forma de una tabla sea una sustitución es necesario y suficiente que tanto en la fila superior como en la inferior aparezcan los símbolos de todos los elementos del conjunto, con la particularidad de que todo elemento figure sólo una vez. Para representar la sustitución inversa σ^{-1} basta, evidentemente, convertir la fila inferior de la tabla de σ en la superior y la superior en la inferior.

Sea $F(x_1, \ldots, x_n)$ una función de las variables $x_1, \ldots, x_n$ y sea σ una sustitución de los números $1, \ldots, n$. El resultado de la sustitución σ aplicada a las variables $x_1, \ldots, x_n$ de la función F es, por definición, la expresión

$$F\sigma=F(x_{1\sigma}, \ldots, x_{n\sigma}).$$

De esta definición se deduce directamente que para cualesquiera sustituciones ρ y σ y para cualesquiera funciones F y G de $x_1, \ldots, x_n$ se tiene

$$F(\rho\sigma)=(F\rho)\,\sigma, \tag{8}$$

$$(FG)\,\sigma=F\sigma\cdot G\sigma\quad \text{y}\quad (F+G)\,\sigma=F\sigma+G\sigma. \tag{9}$$

Tomemos ahora para la función F la expresión

$$\Delta=\prod_{i<j}(x_i-x_j)=(x_1-x_2)(x_1-x_3)\ldots(x_1-x_n)(x_2-x_3)\ldots$$
$$\ldots(x_{n-1}-x_n). \tag{10}$$

Está claro que para cualquier sustitución σ se tiene $\Delta\sigma = \pm\Delta$. Si $\Delta\sigma = \Delta$, se dice que σ es una sustitución *par* y, si $\Delta\sigma = -\Delta$, se dice que σ es *impar*. Para toda aplicación no invertible σ, tenemos $\Delta\sigma = 0$. Por esto, para cualquier aplicación σ del conjunto de los números $1, \ldots, n$ tenemos

$$\Delta\sigma = \varepsilon_\sigma \Delta,$$

donde $\varepsilon_\sigma = +1$, si σ es una sustitución par, $\varepsilon_\sigma = -1$, si σ es una sustitución impar, y $\varepsilon_\sigma = 0$, si σ es una aplicación no invertible. El valor del símbolo ε_σ se denomina también *signatura* de la aplicación σ.

Para dos aplicaciones arbitrarias ρ y σ tenemos de las fórmulas (8) y (9)

$$\varepsilon_{\rho\sigma}\Delta = \Delta(\rho\sigma) = (\Delta\rho)\sigma = \varepsilon_\rho\varepsilon_\sigma\Delta,$$

de donde

$$\varepsilon_{\rho\sigma} = \varepsilon_\rho\varepsilon_\sigma,$$

es decir, *la signatura del producto de aplicaciones es igual al producto de las signaturas de los factores.* Puesto que la signatura de la aplicación idéntica es, evidentemente, igual a $+1$, resulta que *la signatura de la sustitución inversa siempre coincide con la signatura de la sustitución dada.*

Se llama *sustitución cíclica*, o *ciclo*, $(i_1 i_2 \ldots i_m)$ la sustitución σ en la que $i_1\sigma = i_2$, $i_2\sigma = i_3$, $\ldots$, $i_{m-1}\sigma = i_m$, $i_m\sigma = i_1$ e $i\sigma = i$ para los demás elementos i del conjunto, si es que existen. En particular, se llama *ciclo doble*, o *trasposición*, (ij) la sustitución que cambie los elementos i y j y no altere los demás elementos. Cambiando los índices 1 y 2 en la expresión (10) para Δ, veremos fácilmente que $\Delta(1\ 2) = -\Delta$, es decir, que el ciclo doble $(1\ 2)$ es una sustitución impar. Por otro lado, mediante cálculo directo se comprueba que

$$(1\ i)(1\ 2)(1\ i) = (i\ 2) \quad \text{y} \quad (2\ j)(i\ 2)(2\ j) = (i\ j)$$

y puesto que, además, para cualesquiera ρ y σ siempre $\varepsilon_{\rho\sigma\rho} = \varepsilon_\sigma$, obtenemos

$$\varepsilon_{(ij)} = \varepsilon_{(2j)(i\,2)(2\,j)} = \varepsilon_{(i\,2)} = \varepsilon_{(1\,2)},$$

es decir, *todo ciclo doble es una sustitución impar.*

De aquí se deduce, además, que el producto de un número impar de ciclos dobles es una sustitución impar y que el producto de un número par de ciclos dobles es una sustitución par. En particular, de la fórmula

$$(1\,2\,3\ldots m) = (1\ 2)(1\ 3)\ldots(1\ m),$$

que se comprueba fácilmente, se deduce que un ciclo $(i_1\, i_2 \ldots i_m)$ de longitud par m es una sustitución impar y que un ciclo de longitud impar es una sustitución par.

Es claro que los ciclos sin elementos comunes representan sustituciones que conmutan. Al mismo tiempo, cualquier sustitución puede ser descompuesta fácilmente en un producto de ciclos sin elementos comunes. Para ello se toma un elemento cualquiera i_1 del conjunto y se mira en que elemento i_2 lo transforma la sustitución. Si resulta que $i_1 = i_2$, obtenemos para el primer factor el ciclo unidad (i_1), es decir, la sustitución idéntica. Si resulta que $i_1 \neq i_2$, tomamos la imágen i_3 del elemento i_2. Para $i_3 = i_1$ obtenemos el factor en forma del ciclo $(i_1\ i_2)$; si $i_3 \neq i_1$, pasamos a considerar la imagen i_4 del elemento i_3. Si resulta que $i_4 = i_1$, el ciclo se cierra y se obtiene el factor $(i_1\ i_2\ i_3)$, etc.; considerando uno tras otro todos los elementos del conjunto, obtenemos la descomposición de la sustitución en ciclos sin elementos comunes. Por ejemplo, tenemos

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix} = (1\ \ 3\ \ 5)\,(2\ \ 4).$$

Multiplicando las signaturas de los ciclos obtenemos que $\varepsilon_\rho = -1$.

Existe también otra forma para determinar la signatura de una sustitución: mediante el cálculo del número de las así llamadas inversiones. Al representar las sustituciones en forma de tablas hemos convenido en que los elementos del conjunto que figuran en la fila superior pueden ser dispuestos en un orden arbitrario. Sin embargo, en el caso en que el conjunto considerado es la colección de los números enteros, podemos convenir en escribir los números de la primera fila en el orden de crecimiento. En este caso la sustitución quedará plenamente determinada al indicar solamente la fila inferior de la tabla, es decir, al indicar *una permutación* de números. La correspondencia entre las sustituciones y las permutaciones, obtenida de esta forma, resulta biyectiva y obtenemos la posibilidad de representar las sustituciones en forma de tablas de una fila en lugar de tablas de dos filas. Sea σ una sustitución que corresponde a la permutación $i_1, i_2, \ldots, i_n$ de los números $1, 2, \ldots, n$. Consideremos todos los pares i_k, i_l $(k < l)$; diremos que el par i_k, i_l forma *una inversión*, si $i_k > i_l$. De la relación

$$\Delta\sigma = \prod_{k<l} (x_{k\sigma} - x_{l\sigma}) = \prod (x_{i_k} - x_{i_l})$$

se ve que $\varepsilon_\sigma = 1$, si es par el número de factores $x_{i_k} - x_{i_l}$ para los cuales $i_k > i_l$, y que $\varepsilon_\sigma = -1$, si el número de estos factores es impar. Por consiguiente, la paridad de la sustitución σ coincide con la paridad del número de inversiones en la permutación $i_1, i_2, \ldots, i_n$.

Ejemplos y problemas

1. Tomemos en el espacio corriente un sistema rectangular de coordenadas $OXYZ$. Sea $\mathscr{A}$ el giro del espacio alrededor del eje OX en 90° en dirección de OY hacia OZ, sea $\mathscr{B}$ el giro en 90° alrededor del eje OY en dirección de OZ hacia OX y sea $\mathscr{C}$ el giro en 90° alrededor del eje OZ en dirección de OX hacia OY. Sea m el punto de coordenadas $(1, 0, 1)$. Calcúlense las coordenadas de los puntos $m\mathscr{A}$, $m(\mathscr{B}\mathscr{C})$ y $m(\mathscr{C}\mathscr{B})$. Demuéstrese que

$$\mathscr{A}^4=\mathscr{B}^4=\mathscr{C}^4=\mathscr{E},\ \mathscr{A}\mathscr{B}\neq\mathscr{B}\mathscr{A}\ \text{y}\ \mathscr{A}^2\mathscr{B}^2=\mathscr{B}^2\mathscr{A}^2.$$

2. Consideremos dos aplicaciones $\mathscr{A}$ y $\mathscr{C}$ de un conjunto arbitrario $\mathfrak{M}$ aceptando que $\mathscr{C}$ es invertible. Tomemos un elemento cualquiera u de $\mathfrak{M}$ y sea $v=u\mathscr{A}$. La aplicación $\mathscr{C}$ «traslada» los elementos u y v en unos elementos x e y. Demuéstrese que en estas condiciones la aplicación $\mathscr{C}^{-1}\mathscr{A}\mathscr{C}$ transforma x en y (fig. 3).

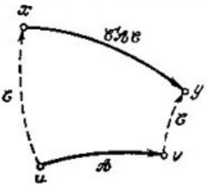

Fig. 3.

3. Sea O un punto arbitrario del plano, sea $\mathscr{A}$ el giro del plano alrededor del punto O en un ángulo determinado α y sea $\mathscr{B}$ el traslado paralelo del plano a una distancia a en una dirección determinada. Demuéstrese que $\mathscr{B}^{-1}\mathscr{A}\mathscr{B}$ es el giro en el ángulo α alrededor del punto $O\mathscr{B}$ y que $\mathscr{A}^{-1}\mathscr{B}\mathscr{A}$ es el traslado del plano a la distancia a en la dirección que se obtiene girando en el ángulo α la dirección inicial.

§ 8. Aplicaciones lineales y sus matrices

8.1. Propiedades elementales. Se llama *aplicación* de un espacio lineal $\mathfrak{L}$ una ley que a todo vector de $\mathfrak{L}$ pone en correspondencia de nuevo un vector determinado de $\mathfrak{L}$. Una aplicación se llama *lineal*, si transforma el producto de un número por un vector en el producto del mismo número por el vector correspondiente y la suma de vectores en la suma de los vectores correspondientes. Abreviando, una aplicación $\mathscr{A}$ se llama lineal, si para cualesquiera vectores x e y del espacio $\mathfrak{L}$ y para cualquier número α del cuerpo de coeficientes tienen lugar las igualdades

$$(\alpha x)\mathscr{A}=\alpha(x\mathscr{A}), \tag{1}$$

$$(x+y)\mathscr{A}=x\mathscr{A}+y\mathscr{A}. \tag{2}$$

Tomando $\alpha=0$ en (1), obtenemos

$$o\mathscr{A}=o,$$

es decir, *toda aplicación lineal transforma el vector nulo en el vector nulo.*

De (1) y (2) también se desprende directamente la siguiente propiedad principal de las aplicaciones lineales: *si $\mathscr{A}$ es una aplicación lineal, se tiene*

$$(\alpha_1x_1+\alpha_2x_2+\ldots+\alpha_mx_m)\mathscr{A}=\alpha_1(x_1\mathscr{A})+\alpha_2(x_2\mathscr{A})+\ldots+\alpha_m(x_m\mathscr{A}), \tag{3}$$

donde $x_1, x_2, \ldots, x_m$ *son vectores arbitrarios de* $\mathfrak{L}$ y $\alpha_1, \alpha_2, \ldots, \alpha_m$ *son números arbitrarios del cuerpo de coeficientes.*

Para la demostración basta aplicar la inducción según m. Para $m=1$ la fórmula (3) coincide con (1). Supongamos ahora que (3) es válida en el caso de $m-1$ sumandos. Entonces

$$\begin{aligned}(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m)\mathcal{A} &= (\alpha_1 x_1 + (\alpha_2 x_2 + \ldots + \alpha_m x_m))\mathcal{A} = \\ &= (\alpha_1 x_1)\mathcal{A} + (\alpha_2 x_2 + \ldots + \alpha_m x_m)\mathcal{A} = \\ &= \alpha_1 (x_1\mathcal{A}) + \alpha_2 (x_2\mathcal{A}) + \ldots + \alpha_m (x_m\mathcal{A}),\end{aligned}$$

que es lo que se quería demostrar. Para $m=2$, la fórmula (3) se convierte en

$$(\alpha x + \beta y)\mathcal{A} = \alpha (x\mathcal{A}) + \beta (y\mathcal{A}), \tag{4}$$

de donde para $\beta=0$ y $\alpha=\beta=1$ se obtiene de nuevo (1) y (2). Por consiguiente, la propiedad (4) caracteriza totalmente las aplicaciones lineales y puede servir de definición de las mismas.

Hemos convenido anteriormente en llamar aplicación idéntica $\mathcal{E}$ aquella aplicación que transforma todo elemento en sí mismo. Por esto, si el conjunto considerado es la colección de los vectores de un espacio lineal $\mathfrak{L}$, se tiene

$$(\alpha x + \beta y)\mathcal{E} = \alpha x + \beta y = \alpha (x\mathcal{E}) + \beta (y\mathcal{E}).$$

Por consiguiente, *la aplicación idéntica de un espacio vectorial es lineal.* La aplicación que transforma todo vector en el vector nulo se llama *nula* y se indica por $\mathcal{O}$. Está claro que la aplicación nula también es lineal.

Consideremos dos ejemplos concretos. Sea $\mathfrak{R}$ el espacio vectorial corriente, es decir, el conjunto de segmentos orientados que parten de un punto fijo O. Consideremos un plano que pasa por O e indiquemos mediante $x\mathcal{A}$ la proyección del segmento x sobre este plano. Entonces las conocidas propiedades de las proyecciones — 1) la proyección de una suma de segmentos es igual a la suma de sus proyecciones y 2) si un segmento es aumentado en α veces, su proyección también se aumenta en α veces — pueden ser representadas en forma de las igualdades

$$(x+y)\mathcal{A} = x\mathcal{A} + y\mathcal{A} \quad \text{y} \quad (\alpha x)\mathcal{A} = \alpha (x\mathcal{A}),$$

de las cuales se desprende que la operación de proyección es una aplicación lineal.

Como segundo ejemplo consideremos el conjunto de todos los polinomios en la variable λ de orden no mayor que n. Estos polinomios forman, respecto a las operaciones corrientes de adición y de multiplicación por número, un espacio lineal de dimensión $n+1$. Pongamos en correspondencia a todo polinomio su derivada. Puesto que la derivada de una suma es igual a la suma de las derivadas de los sumandos y puesto que un factor constante puede ser extraído del signo de la derivada, resulta que la operación de diferenciación es una aplicación lineal del espacio de polinomios.

Veamos ahora mediante qué elementos puede ser definida una aplicación lineal.

TEOREMA *Sea $a_1, a_2, \ldots, a_n$ una base de un espacio lineal $\mathfrak{L}$. Tomemos en $\mathfrak{L}$ unos vectores absolutamente arbitrarios $b_1, b_2, \ldots, b_n$. Entonces existe una aplicación lineal del espacio $\mathfrak{L}$, y sólo una, que transforma los vectores $a_1, a_2, \ldots, a_n$ en los vectores $b_1, b_2, \ldots, b_n$ respectivamente.*

Para construir la aplicación deseada, tomemos un vector arbitrario x y representémoslo linealmente en términos de la base $a_1, a_2, \ldots, a_n$. Sea

$$x = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_n a_n. \tag{5}$$

Consideremos el vector

$$x' = \xi_1 b_1 + \xi_2 b_2 + \ldots + \xi_n b_n.$$

Indiquemos por $\mathcal{A}$ la aplicación que transforma x en x'. Por consiguiente, si x está representado por (5), se tiene

$$x\mathcal{A} = \xi_1 b_1 + \xi_2 b_2 + \ldots + \xi_n b_n. \tag{6}$$

Tomando aquí $\xi_i = 1$ y $\xi_j = 0$ ($j \neq i$, $j = 1, 2, \ldots, n$), obtenemos $a_i\mathcal{A} = b_i$ ($i = 1, 2, \ldots, n$), de modo que la aplicación $\mathcal{A}$ transforma los vectores $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$. Demostremos que $\mathcal{A}$ es *lineal*. Multiplicando (5) por un número cualquiera α, tendremos

$$\alpha x = (\alpha\xi_1) a_1 + (\alpha\xi_2) a_2 + \ldots + (\alpha\xi_n) a_n.$$

Comparando este resultado con (6), obtenemos

$$(\alpha x)\mathcal{A} = (\alpha\xi_1) b_1 + (\alpha\xi_2) b_2 + \ldots + (\alpha\xi_n) b_n,$$

es decir,

$$(\alpha x)\mathcal{A} = \alpha (x\mathcal{A}). \tag{7}$$

Sea

$$y = \eta_1 a_1 + \eta_2 a_2 + \ldots + \eta_n a_n$$

otro vector de $\mathfrak{L}$. Entonces se tiene

$$x + y = (\xi_1 + \eta_1) a_1 + (\xi_2 + \eta_2) a_2 + \ldots + (\xi_n + \eta_n) a_n$$

y, por consiguiente,

$$y\mathcal{A} = \eta_1 b_1 + \eta_2 b_2 + \ldots + \eta_n b_n,$$
$$(x + y)\mathcal{A} = (\xi_2 + \eta_1) b_1 + \ldots + (\xi_n + \eta_n) b_n.$$

De la última igualdad resulta

$$(x + y)\mathcal{A} = \xi_1 b_1 + \ldots + \xi_n b_n + \eta_1 b_1 + \ldots + \eta_n b_n = x\mathcal{A} + y\mathcal{A}. \tag{8}$$

Las propiedades (7) y (8) significan que $\mathcal{A}$ es lineal.

Resta probar que toda aplicación lineal $\mathcal{B}$ que transforma $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$ coincide con $\mathcal{A}$. Por hipótesis, $a_i\mathcal{B}=b_i$ $(i=1, 2, \ldots, n)$. Luego, si el vector x está representado por (5), se tiene

$$x\mathcal{B}=\xi_1(a_1\mathcal{B})+\ldots+\xi_n(a_n\mathcal{B})=\xi_1b_1+\ldots+\xi_nb_n=x\mathcal{A},$$

es decir, $\mathcal{B}=\mathcal{A}$. Hemos demostrado el teorema.

8.2. Matriz de una aplicación lineal. Veremos ahora cómo se puede definir una aplicación lineal mediante números. Tomemos en el espacio $\mathfrak{L}$ un sistema cualquiera de coordenadas $a_1, a_2, \ldots, a_n$ y supongamos dada una aplicación lineal $\mathcal{A}$ de este espacio. La aplicación $\mathcal{A}$ transforma los vectores $a_1, \ldots, a_n$ en unos vectores $a_1\mathcal{A}, a_2\mathcal{A}, \ldots, a_n\mathcal{A}$ que pueden ser expresados linealmente en términos de $a_1, a_2, \ldots, a_n$. Sean

$$\begin{aligned}
a_1\mathcal{A}&=\alpha_{11}a_1+\alpha_{12}a_2+\ldots+\alpha_{1n}a_n,\\
a_2\mathcal{A}&=\alpha_{21}a_1+\alpha_{22}a_2+\ldots+\alpha_{2n}a_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
a_n\mathcal{A}&=\alpha_{n1}a_1+\alpha_{n2}a_2+\ldots+\alpha_{nn}a_n
\end{aligned}$$

estas expresiones. La matriz

$$A=\begin{bmatrix}\alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n}\\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n}\\ \cdot & \cdot & \cdot & \cdot\\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn}\end{bmatrix},$$

formada por las filas coordenadas de los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$, se llama *matriz de la aplicación* $\mathcal{A}$ en el sistema de coordenadas $a_1, a_2, \ldots, a_n$. Por consiguiente, dado un sistema de coordenadas, a toda aplicación lineal corresponde una matriz determinada. Surge la pregunta de si esta correspondencia es biyectiva. La respuesta es positiva ya que conociendo la matriz A podemos encontrar primero los vectores

$$\begin{aligned}
b_1&=\alpha_{11}a_1+\alpha_{12}a_2+\ldots+\alpha_{1n}a_n,\\
b_2&=\alpha_{21}a_1+\alpha_{22}a_2+\ldots+\alpha_{2n}a_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
b_n&=\alpha_{n1}a_1+\alpha_{n2}a_2+\ldots+\alpha_{nn}a_n
\end{aligned}$$

y después construir, de acuerdo con el teorema del punto anterior, la aplicación lineal $\mathcal{A}$ que transforma $a_1, \ldots, a_n$ en $b_1, \ldots, b_n$ respectivamente. Esta aplicación es única y su matriz coincide, obviamente, con la matriz A que es lo que se quería demostrar.

Veamos qué matrices corresponden a las aplicaciones idéntica y nula del espacio $\mathfrak{L}$. Sea $a_1, a_2, \ldots, a_n$ un sistema arbitrario de coordenadas de $\mathfrak{L}$. Entonces tendremos

$$\begin{array}{lcl} a_1\mathcal{O} = 0\cdot a_1 + \ldots + 0\cdot a_n, & & a_1\mathcal{E} = 1\cdot a_1 + \ldots + 0\cdot a_n, \\ \ldots\ldots\ldots\ldots & \text{y} & \ldots\ldots\ldots\ldots \\ a_n\mathcal{O} = 0\cdot a_1 + \ldots + 0\cdot a_n & & a_n\mathcal{E} = 0\cdot a_1 + \ldots + 1\cdot a_n. \end{array}$$

Por consiguiente, *la aplicación nula tiene la matriz nula y la aplicación idéntica tiene la matriz unidad.*

Planteémonos el problema: ¿cómo conociendo la matriz de una aplicación lineal $\mathcal{A}$ y las coordenadas de un vector x determinar las coordenadas del vector $x\mathcal{A}$?

Sea $a_1, a_2, \ldots, a_n$ el sistema de coordenadas escogido y sean $\xi_1, \xi_2, \ldots, \xi_n$ las coordenadas del vector x en el mismo. Indiquemos por α_{ij} $(i, j = 1, \ldots, n)$ los elementos de la matriz A de la aplicación $\mathcal{A}$ calculados en este mismo sistema de coordenadas. Tenemos

$$x = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_n a_n,$$
$$x\mathcal{A} = \xi_1 (a_1\mathcal{A}) + \xi_2 (a_2\mathcal{A}) + \ldots + \xi_n (a_n\mathcal{A}).$$

Según la definición de la matriz de una aplicación se tiene

$$a_i\mathcal{A} = \alpha_{i1}a_1 + \alpha_{i2}a_2 + \ldots + \alpha_{in}a_n \quad (i = 1, \ldots, n).$$

Introduciendo estos valores en la igualdad anterior, obtenemos

$$x\mathcal{A} = (\xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1})\, a_1 + \ldots$$
$$\ldots + (\xi_1\alpha_{1n} + \xi_2\alpha_{2n} + \ldots + \xi_n\alpha_{nn})\, a_n.$$

Por consiguiente, la fila de coordenadas del nuevo vector $x\mathcal{A}$ es

$$[x\mathcal{A}] = [\xi_1\alpha_{11} + \ldots + \xi_n\alpha_{n1}, \ \ldots, \ \xi_1\alpha_{1n} + \ldots + \xi_n\alpha_{nn}] = [x]\, A,$$

es decir, *la fila de coordenadas del vector nuevo es igual a la fila de coordenadas del vector antiguo multiplicada por la matriz de la aplicación lineal:*

$$[x\mathcal{A}] = [x]\, A.$$

8.3. Transformación de coordenadas. En el punto anterior ha sido establecida una correspondencia biyectiva entre las aplicaciones lineales de un espacio vectorial $\mathfrak{L}$ de n dimensiones y las matrices cuadradas de orden n. Sin embargo, para ello ha sido necesario escoger primero en $\mathfrak{L}$ un determinado sistema de coordenadas. Si cambiamos éste, cambiamos la correspondencia. Tendremos como resultado que a una misma aplicación lineal $\mathcal{A}$ corresponderán en los sistemas antiguo y nuevo de coordenadas diferentes matrices A y A_1. Hallemos la relación entre las mismas.

Sean $a_1, a_2, \ldots, a_n$ y $a_1', a_2', \ldots, a_n'$ el antiguo y el nuevo sistemas de coordenadas y sea T la matriz del cambio (véase el p. 5.1). Indiquemos por $[x]$ y $[x]_1$ las filas de coordenadas del vector x en el antiguo y en el nuevo sistemas de coordenadas. De acuerdo con la regla de una aplicación lineal, tenemos en el sistema antiguo de coordenadas

$$[x\mathcal{A}] = [x] A.$$

En el nuevo sistema de coordenadas esta misma regla da

$$[x\mathcal{A}]_1 = [x]_1 A_1.$$

Sin embargo, la regla de transformación de coordenadas del p. 5.1 muestra que

$$[x] = [x]_1 T \quad \text{y} \quad [x\mathcal{A}] = [x\mathcal{A}]_1 T.$$

Introduciendo aquí los valores de $[x\mathcal{A}]$, $[x\mathcal{A}]_1$ y $[x]$ de las igualdades anteriores, obtenemos

$$[x]_1 TA = [x]_1 A_1 T.$$

Puesto que el vector x es arbitrario, tenemos, aplicando el lema del p. 5.1, $TA = A_1 T$, o

$$A_1 = TAT^{-1}.$$

Es decir, *la matriz de una aplicación lineal en el sistema nuevo de coordenadas es igual a la matriz de esta misma aplicación lineal en el sistema antiguo transformada por la matriz del cambio recíproco.*

Para concluir consideremos el problema de cómo puede ser interpretada, desde el punto de vista de la teoría de aplicaciones lineales, la matriz del cambio de un sistema de coordenadas por otro. Sean $a_1, a_2, \ldots, a_n$ y $a'_1, a'_2, \ldots, a'_n$ dos sistemas de coordenadas dados de un espacio $\mathfrak{L}$. Según el teorema del p. 8.1, existe una aplicación lineal $\mathcal{T}$, determinada unívocamente, que transforma el sistema antiguo de vectores coordenados en el sistema nuevo, es decir, que posee la propiedad

$$a_i\mathcal{T} = a'_i \quad (i = 1, 2, \ldots, n).$$

Para escribir ahora, según la regla del p. 8.2, la matriz de la aplicación $\mathcal{T}$ en el sistema antiguo, tenemos que expresar todos los vectores a'_i en términos de $a_1, \ldots, a_n$. Pero lo mismo tenemos que hacer para obtener la matriz del cambio. Por consiguiente, *la matriz del cambio es la matriz de la aplicación lineal que transforma el sistema antiguo de coordenadas en el sistema nuevo*, calculada en el sistema antiguo de coordenadas. Es más, la última indicación es innecesaria, ya que al calcular la matriz de la aplicación $\mathcal{T}$ en el sistema nuevo de coordenadas obtendremos la misma matriz del cambio T. En efecto, si la matriz de la aplicación $\mathcal{T}$ en el sistema antiguo de coordenadas es T, la matriz del cambio también será T. Por lo tanto, la matriz de $\mathcal{T}$ en el nuevo sistema será igual a $TTT^{-1} = T$, que es lo que se quería demostrar.

Ejemplos y problemas

1. Consideremos el espacio de los vectores que pertenecen a un plano y parten de un punto O. Demuéstrese que la aplicación, consistente en el giro de todos los vectores en un ángulo α alrededor del punto O, es lineal y que su matriz es igual a $\begin{bmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$, siempre que el sistema de coordenadas esté formado por dos vectores perpendiculares de longitud 1.

2. Sea $\mathfrak{R}$ el espacio corriente de los segmentos orientados que parten de un punto O. Tomemos en $\mathfrak{R}$ un sistema de coordenadas formado por tres vectores perpendiculares e_1, e_2 y e_3 de longitud 1. Demuéstrese que la matriz de la aplicación lineal $\mathcal{P}$, consistente en la proyección de los vectores sobre el eje e_1, es igual a $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, mientras que la matriz de la proyección sobre el plano e_1e_2 es igual a $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

3. ¿Cómo cambia la matriz de una aplicación lineal $\mathcal{A}$, si en el sistema de coordenadas $e_1, e_2, \ldots, e_n$ se cambian entre si dos vectores cualesquiera, por ejemplo, e_1 y e_2?

4. Toda aplicación lineal de un espacio de dimensión uno consiste en la multiplicación de todos sus vectores por un mismo número.

5. Sea $\mathfrak{L}$ el espacio de todas las matrices cuadradas de segundo orden. Demuéstrese que la aplicación $\mathcal{A}$, consistente en la multiplicación a la derecha de todas las matrices de $\mathfrak{L}$ por la matriz $\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$, es lineal. Hállese la matriz de la aplicación $\mathcal{A}$ si como sistema de coordenadas de $\mathfrak{L}$ se toma el sistema

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

6. Una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ de cuatro dimensiones tiene en el sistema de coordenadas e_1, e_2, e_3 y e_4 la matriz

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

¿Cuál será la matriz de esta aplicación, si para el nuevo sistema de coordenadas se toma el sistema 1) e_1, e_3, e_2 y e_4; 2) $e_1, e_1+e_2, e_1+e_2+e_3$ y $e_1+e_2+e_3+e_4$?

§ 9. Operaciones con aplicaciones lineales

9.1. Multiplicación de aplicaciones lineales. Sean $\mathcal{A}$ y $\mathcal{B}$ dos aplicaciones lineales definidas en un espacio lineal $\mathfrak{L}$. Aplicando a cualquier vector x de $\mathfrak{L}$ primero la aplicación $\mathcal{A}$ y después la aplicación $\mathcal{B}$, obtendremos un vector

$$y = (x\mathcal{A})\,\mathcal{B}.$$

La aplicación que transforma x directamente en y ha sido llamada en el p. 7.2 producto de $\mathcal{A}$ por $\mathcal{B}$. Por consiguiente,

$$x(\mathcal{A}\mathcal{B}) = (x\mathcal{A})\,\mathcal{B}. \tag{1}$$

Demostremos que *el producto de dos aplicaciones lineales es una aplicación lineal*. Para ello es suficiente demostrar, según el p. 8.1, que tiene lugar la igualdad

$$(\alpha x+\beta y)(\mathcal{A}\mathcal{B})=\alpha\cdot x(\mathcal{A}\mathcal{B})+\beta\cdot y(\mathcal{A}\mathcal{B}).$$

En virtud de (1) se tiene

$$(\alpha x+\beta y)(\mathcal{A}\mathcal{B})=((\alpha x+\beta y)\mathcal{A})\mathcal{B}.$$

Puesto que $\mathcal{A}$ y $\mathcal{B}$ son lineales, resulta

$$((\alpha x+\beta y)\mathcal{A})\mathcal{B}=(\alpha(x\mathcal{A})+\beta(y\mathcal{A}))\mathcal{B}=\alpha\cdot(x\mathcal{A})\mathcal{B}+\beta\cdot(y\mathcal{A})\mathcal{B},$$

es decir,

$$(\alpha x+\beta y)(\mathcal{A}\mathcal{B})=\alpha\cdot x(\mathcal{A}\mathcal{B})+\beta\cdot y(\mathcal{A}\mathcal{B})$$

que es lo que se quería demostrar.

Tomemos en el espacio $\mathfrak{L}$ un sistema de coordenadas e indiquemos por A y B las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$. ¿Cómo hallar la matriz de la aplicación $\mathcal{A}\mathcal{B}$? Indiquemos esta matriz incógnita por C. Sea x un vector cualquiera del espacio $\mathfrak{L}$ y sea $[x]$ su fila de coordenadas. Según la regla de aplicación lineal, se tiene

$$[x(\mathcal{A}\mathcal{B})]=[x]C.$$

Por otro lado, tenemos

$$[x(\mathcal{A}\mathcal{B})]=[(x\mathcal{A})\mathcal{B}]=[x\mathcal{A}]B=[x]AB.$$

Comparando ambos resultados vemos que

$$[x]C=[x]AB$$

cualquiera que sea x. De acuerdo con el lema del p. 5.1 de aquí se deduce que

$$C=AB,$$

es decir, *la matriz de un producto de aplicaciones lineales es igual al producto de las matrices de estas aplicaciones*.

Consideremos una aplicación lineal *biyectiva* $\mathcal{A}$. Si $\mathcal{A}$ transforma el vector x en y, la aplicación que transforma y en x será la aplicación *inversa* $\mathcal{A}^{-1}$ (véase el p. 7.3). Demostremos que *siendo $\mathcal{A}$ lineal, también $\mathcal{A}^{-1}$ es lineal*. En efecto, sea

$$u\mathcal{A}^{-1}=x \quad \text{y} \quad v\mathcal{A}^{-1}=y,$$

donde u y v son vectores arbitrarios de $\mathfrak{L}$. Puesto que la aplicación $\mathcal{A}$ es lineal, se tiene

$$(\alpha x+\beta y)\mathcal{A}=\alpha\cdot x\mathcal{A}+\beta\cdot y\mathcal{A}=\alpha u+\beta v,$$

de donde resulta

$$(\alpha u+\beta v)\mathcal{A}^{-1}=\alpha x+\beta y=\alpha\cdot u\mathcal{A}^{-1}+\beta\cdot v\mathcal{A}^{-1}$$

que es lo que se quería demostrar.

Siendo A la matriz de la aplicación $\mathcal{A}$ y X la matriz de la aplicación $\mathcal{A}^{-1}$, de las relaciones

$$\mathcal{A}\mathcal{A}^{-1}=\mathcal{A}^{-1}\mathcal{A}=\mathcal{E}$$

se deduce que

$$AX=XA=E,$$

es decir, $X=A^{-1}$. Por consiguiente, *la aplicación inversa tiene la matriz inversa*. En particular, *para que una aplicación lineal sea invertible es necesario y suficiente que sea invertible su matriz.*

De nuestros resultados también se desprende directamente la siguiente regla: *la matriz de la aplicación $\mathcal{A}^n$ es igual a A^n, donde A es la matriz de la aplicación $\mathcal{A}$.*

9.2. Multiplicación por número y adición. En este punto y en el que sigue se supone que el cuerpo principal K es un cuerpo conmutativo. La aplicación que resulta al aplicar a un vector primero la aplicación $\mathcal{A}$ y al multiplicar después el vector nuevo por un número α se llama *producto* de α por $\mathcal{A}$ y se indica por $\alpha\mathcal{A}$. Esta definición se puede expresar mediante la fórmula

$$x(\alpha\mathcal{A})=\alpha(x\mathcal{A}).$$

Razonando igual que en el caso de la multiplicación de aplicaciones lineales, veremos fácilmente que *siendo $\mathcal{A}$ una aplicación lineal, también $\alpha\mathcal{A}$ es una aplicación lineal*. Determinemos su matriz. Supongamos que en un sistema de coordenadas fijado la aplicación $\mathcal{A}$ tiene la matriz A y la aplicación $\alpha\mathcal{A}$ tiene la matriz B. Según la regla de una aplicación lineal, tenemos

$$[x(\alpha\mathcal{A})]=[x]B \quad \text{y}$$
$$[x(\alpha\mathcal{A})]=[\alpha(x\mathcal{A})]=\alpha[x\mathcal{A}]=\alpha[x]A=[x](\alpha A),$$

de donde resulta que $[x]B=[x](\alpha A)$, es decir, $B=\alpha A$.

Luego, *la matriz del producto de un número por una aplicación lineal es igual al producto de este número por la matriz de la aplicación.*

Indiquemos, finalmente, las fórmulas

$$\alpha(\beta\mathcal{A})=(\alpha\beta)\mathcal{A},$$
$$0\cdot\mathcal{A}=\mathcal{O},$$
$$1\cdot\mathcal{A}=\mathcal{A},$$
$$\alpha(\mathcal{A}\mathcal{B})=(\alpha\mathcal{A})\mathcal{B}=\mathcal{A}(\alpha\mathcal{B}),$$

análogas plenamente a las fórmulas correspondientes del cálculo de matrices. La demostración queda a cargo del lector.

Tomemos ahora dos aplicaciones $\mathcal{A}$ y $\mathcal{B}$ de un espacio lineal $\mathfrak{L}$ y pongamos en correspondencia a todo vector x de $\mathfrak{L}$ el vector $x\mathcal{A}+x\mathcal{B}$. La aplicación que transforma x en $x\mathcal{A}+x\mathcal{B}$ se indica por $\mathcal{A}+\mathcal{B}$ y se denomina *suma* de $\mathcal{A}$ y $\mathcal{B}$. Es decir, por definición,

$$x(\mathcal{A}+\mathcal{B})=x\mathcal{A}+x\mathcal{B}.$$

Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ son lineales, se tiene

$$\begin{gathered}(\alpha x+\beta y)(\mathcal{A}+\mathcal{B})=(\alpha x+\beta y)\mathcal{A}+(\alpha x+\beta y)\mathcal{B}=\\ =\alpha\cdot x\mathcal{A}+\beta\cdot y\mathcal{A}+\alpha\cdot x\mathcal{B}+\beta\cdot y\mathcal{B}=\\ =\alpha(x\mathcal{A}+x\mathcal{B})+\beta(y\mathcal{A}+y\mathcal{B})=\alpha\cdot x(\mathcal{A}+\mathcal{B})+\beta\cdot y(\mathcal{A}+\mathcal{B}).\end{gathered}$$

Por consiguiente, *la suma de dos aplicaciones lineales es una aplicación lineal*. Determinemos su matriz. Sean A y B las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ en un sistema de coordenadas. Sea C la matriz de la aplicación $\mathcal{A}+\mathcal{B}$. Tenemos

$$[x(\mathcal{A}+\mathcal{B})]=[x]C,$$

$$[x(\mathcal{A}+\mathcal{B})]=[x\mathcal{A}+x\mathcal{B}]=[x\mathcal{A}]+[x\mathcal{B}]=[x]A+[x]B=[x](A+B).$$

Por consiguiente,

$$[x]C=[x](A+B) \quad \text{y} \quad C=A+B,$$

es decir, *la matriz de una suma de aplicaciones es igual a la suma de sus matrices.*

Las operaciones con aplicaciones lineales se rigen por las mismas leyes que las operaciones con matrices. Una parte de ellas, las relacionadas solamente con la multiplicación, ya las hemos indicado. Señalemos ahora también aquellas que se refieren o bien a la adición o bien a las relaciones entre la adición y multiplicación:

$$\begin{gathered}\mathcal{A}+\mathcal{B}=\mathcal{B}+\mathcal{A},\\ (\mathcal{A}+\mathcal{B})+\mathcal{C}=\mathcal{A}+(\mathcal{B}+\mathcal{C}),\\ \mathcal{A}+\mathcal{O}=\mathcal{A},\\ \alpha(\mathcal{A}+\mathcal{B})=\alpha\mathcal{A}+\alpha\mathcal{B},\\ (\alpha+\beta)\mathcal{A}=\alpha\mathcal{A}+\beta\mathcal{A},\\ \mathcal{A}(\mathcal{B}+\mathcal{C})=\mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C},\\ (\mathcal{A}+\mathcal{B})\mathcal{C}=\mathcal{A}\mathcal{C}+\mathcal{B}\mathcal{C}.\end{gathered}$$

Todas estas igualdades se demuestran siguiendo un mismo método: se toma un vector arbitrario x y se demuestra que la aplicación que figura en el primer miembro transforma x en el mismo vector en el que lo transforma la aplicación que figura en el segundo miembro. Por ejemplo,

$$\begin{gathered}x\mathcal{A}(\mathcal{B}+\mathcal{C})=(x\mathcal{A})(\mathcal{B}+\mathcal{C})=(x\mathcal{A})\mathcal{B}+(x\mathcal{A})\mathcal{C},\\ x(\mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C})=x(\mathcal{A}\mathcal{B})+x(\mathcal{A}\mathcal{C})=(x\mathcal{A})\mathcal{B}+(x\mathcal{A})\mathcal{C},\end{gathered}$$

de donde resulta que $\mathcal{A}(\mathcal{B}+\mathcal{C})=\mathcal{A}\mathcal{B}+\mathcal{A}\mathcal{C}$.

Todas estas leyes también se pueden deducir directamente de las fórmulas correspondientes del cálculo de matrices. En efecto, existe una correspondencia biyectiva entre las matrices cuadradas de orden n y las aplicaciones lineales de un espacio lineal de n dimensiones. Esta correspondencia posee la propiedad de transformar la suma en la suma y el producto en el producto. Por esto, toda identidad entre matrices implica una identidad análoga entre aplicaciones lineales, que es lo que se quería demostrar.

9.3. Polinomios en aplicaciones lineales. Sea

$$f(\lambda)=\alpha_0+\alpha_1\lambda+\ldots+\alpha_m\lambda^m$$

un polinomio en la variable λ. La expresión

$$f(\mathcal{A})=\alpha_0\mathcal{E}+\alpha_1\mathcal{A}+\ldots+\alpha_m\mathcal{A}^m,$$

donde $\mathcal{E}$ es la aplicación idéntica y $\mathcal{A}$ es una aplicación cualquiera, se llama valor del polinomio $f(\lambda)$ para $\lambda=\mathcal{A}$ o simplemente *polinomio en $\mathcal{A}$. Si la matriz de la aplicación $\mathcal{A}$ en un sistema de coordenadas es igual a A, la matriz de la aplicación $f(\mathcal{A})$ en este mismo sistema de coordenadas es*

$$f(A)=\alpha_0E+\alpha_1A+\ldots+\alpha_mA^m.$$

Efectivamente, $f(\mathcal{A})$ se obtiene de $\mathcal{A}$ mediante las operaciones de multiplicación, de multiplicación por número y de adición. De los resultados señalados anteriormente se ve que realizando estas mismas operaciones con la matriz A se obtiene la matriz de la aplicación $f(\mathcal{A})$.

Todas las reglas de operaciones con polinomios en una variable tienen lugar también para los polinomios en una aplicación lineal. Por esto, si en alguna identidad entre polinomios en λ se sustituye λ por una aplicación lineal, se obtiene una relación verídica. Por ejemplo, de las identidades

$$\lambda^2-1=(\lambda-1)(\lambda+1) \quad \text{y} \quad (\lambda+1)^2+(\lambda-1)^2-2\lambda^2=2$$

se obtiene realizando la sustitución $\lambda=\mathcal{A}$

$$\mathcal{A}^2-\mathcal{E}=(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E}) \text{ y}$$
$$(\mathcal{A}+\mathcal{E})^2+(\mathcal{A}-\mathcal{E})^2-2\mathcal{A}^2=2\mathcal{E}.$$

En particular, la identidad

$$f(\lambda)g(\lambda)=g(\lambda)f(\lambda)$$

implica la relación

$$f(\mathcal{A})g(\mathcal{A})=g(\mathcal{A})f(\mathcal{A})$$

que significa que los polinomios en una misma aplicación lineal siempre conmutan.

La situación resulta diferente en el caso de polinomios en varias variables. Esto se debe a que en los cálculos con los polinomios se acepta que las variables conmutan. Por esto tenemos, por ejemplo, las igualdades $\lambda\mu\lambda = \lambda^2\mu$, $\lambda\mu^2\lambda\mu = \lambda^2\mu^3$, etc.

Sustituyendo en estas igualdades las variables λ y μ por unas aplicaciones lineales arbitrarias $\mathcal{A}$ y $\mathcal{B}$, obtenemos las relaciones

$$\mathcal{A}\mathcal{B}\mathcal{A} = \mathcal{A}^2\mathcal{B} \quad \text{y} \quad \mathcal{A}\mathcal{B}^2\mathcal{A}\mathcal{B} = \mathcal{A}^2\mathcal{B}^3$$

que pueden resultar falsas para determinadas aplicaciones lineales. Está claro que estas dificultades desaparecen, si las aplicaciones consideradas conmutan. Por consiguiente, en toda identidad entre polinomios en varias incógnitas se pueden sustituir estas incógnitas por aplicaciones lineales arbitrarias que conmutan, obteniéndose como resultado una relación verídica entre aplicaciones lineales.

Hemos convenido en que todos los espacios que se consideran en este libro son de dimensión finita. Sin embargo, esto no excluye el hecho de que una parte de definiciones y de teoremas tenga lugar también para los espacios de dimensión infinita. A título de ejemplo, podemos señalar las definiciones de aplicaciones lineales y de las operaciones con las mismas y aquellas propiedades de las aplicaciones lineales, expuestas en este parágrafo, que no están relacionadas con matrices.

Ejemplos y problemas

1. Sea $\mathfrak{L}$ el conjunto de todos los polinomios en λ de grado $\leqslant n$. Sea $\mathcal{D}$ la aplicación que transforma todo polinomio $f(\lambda)$ en su derivada $f'(\lambda)$. Demuéstrese que $\mathcal{D}^{n+1} = \mathcal{O}$. Hállese la matriz de $\mathcal{D}$ en el sistema de coordenadas $1, \lambda, \ldots, \lambda^n$.

2. Sea $\mathfrak{L}$ el espacio de dimensión infinita de todos los polinomios en λ. Indiquemos por $\mathcal{D}$ la operación de diferenciación y por $\mathcal{C}$ la operación de multiplicación de polinomios por λ. Demuéstrese que ambas operaciones son lineales y que están ligadas por las relaciones

$$\mathcal{C}^n\mathcal{D} - \mathcal{D}\mathcal{C}^n = n\mathcal{C}^{n-1} \quad (n = 1, 2, \ldots).$$

3. ¿Por qué no se puede considerar la aplicación $\mathcal{C}$, indicada en el problema anterior, en el espacio de polinomios de grado no mayor de n?

4. Las aplicaciones lineales de un espacio $\mathfrak{L}$ forman, respecto a las operaciones de adición y de multiplicación por número, un espacio lineal. ¿Cuál es la dimensión de este espacio, si la dimensión de $\mathfrak{L}$ es igual a n?

§ 10. Rango y defecto de una aplicación lineal

Hasta el momento hemos considerado aquellas propiedades de las aplicaciones lineales que se refieren principalmente a las reglas de las operaciones con las mismas. Estudiemos ahora algunas propiedades de carácter más bien geométrico.

10.1. Núcleo y dominio de valores. Sea $\mathfrak{M}$ un conjunto de vectores de un espacio lineal $\mathfrak{L}$ y sea $\mathcal{A}$ una aplicación lineal cualquiera del último. Todo vector a de $\mathfrak{M}$ se transforma por la aplicación lineal $\mathcal{A}$ en un nuevo vector $a\mathcal{A}$ que es la imagen del vector a. En general, esta imagen no pertenecerá a $\mathfrak{M}$. El conjunto de las imágenes de todos los vectores de $\mathfrak{M}$ se llamará *imagen* de $\mathfrak{M}$ respecto de $\mathcal{A}$ y se indicará por $\mathfrak{M}\mathcal{A}$. Convendremos en llamar *imagen recíproca* del conjunto $\mathfrak{M}$ el conjunto de todos los vectores de $\mathfrak{L}$ cuyas imágenes pertenezcan a $\mathfrak{M}$.

TEOREMA 1. *Las imágenes y las imágenes recíprocas de los subespacios lineales de un espacio $\mathfrak{L}$ respecto a una aplicación lineal cualquiera $\mathcal{A}$ son también subespacios lineales.*

En efecto, sea $\mathfrak{A}$ un subespacio lineal de $\mathfrak{L}$. Mostremos que $\mathfrak{A}\mathcal{A}$ es también un subespacio lineal. Tomemos en $\mathfrak{A}\mathcal{A}$ unos vectores a y b cualesquiera. Estos vectores son las imágenes de unos vectores x e y de $\mathfrak{A}$, es decir, $a = x\mathcal{A}$ y $b = y\mathcal{A}$. Puesto que $\mathfrak{A}$ es un subespacio lineal, el vector $\alpha x + \beta y$ pertenece a $\mathfrak{A}$ cualesquiera que sean α y β. Por ello, el vector $(\alpha x + \beta y)\mathcal{A}$ también pertenece a $\mathfrak{A}\mathcal{A}$. Pero tenemos

$$(\alpha x + \beta y)\mathcal{A} = \alpha \cdot x\mathcal{A} + \beta \cdot y\mathcal{A} = \alpha a + \beta b,$$

es decir, $\mathfrak{A}\mathcal{A}$ contiene el vector $\alpha a + \beta b$ y, por consiguiente, es un subespacio lineal. De forma análoga se demuestra también que la imagen recíproca de un subespacio lineal $\mathfrak{A}$ es un subespacio lineal.

Se llama *núcleo* de una aplicación lineal $\mathcal{A}$ el conjunto de todos los vectores de $\mathfrak{L}$ que se transforman por la aplicación $\mathcal{A}$ en el vector nulo y se llama *dominic de valores* de $\mathcal{A}$ el conjunto de las imágenes de todos los vectores de $\mathfrak{L}$. La dimensión del dominio de valores se llama *rango* de la aplicación y la dimensión del núcleo se llama *defecto* de la aplicación.

TEOREMA 2. *La suma del rango y del defecto de una aplicación lineal $\mathcal{A}$ es igual a la dimensión del espacio $\mathfrak{L}$.*

Indiquemos por $\mathfrak{N}$ el núcleo de la aplicación $\mathcal{A}$ y supongamos que d es la dimensión del núcleo $\mathfrak{N}$. Indiquemos por r la dimensión del dominio de valores $\mathfrak{L}\mathcal{A}$. Por definición, d y r son, respectivamente, el defecto y el rango de la aplicación $\mathcal{A}$. Tomemos en el dominio de valores $\mathfrak{L}\mathcal{A}$ una base $a_1, a_2, \ldots, a_r$ y sean $b_1, b_2, \ldots, b_r$ unos elementos del espacio $\mathfrak{L}$ que se transforman por la aplicación $\mathcal{A}$ en $a_1, a_2, \ldots, a_r$, respectivamente. Los vectores $b_1, b_2, \ldots, b_r$ son linealmente independientes, ya que de la relación

$$\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r = o$$

se deduce que

$$(\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r)\mathcal{A} = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_r a_r = o,$$

y, puesto que $a_1, a_2, \ldots, a_r$ son linealmente independientes, se tiene

$$\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0.$$

Consideremos el subespacio $\mathfrak{M}$ tendido sobre los vectores $b_1, b_2, \ldots, b_r$. El sistema $b_1, \ldots, b_r$ es *una base* de $\mathfrak{M}$ y por ello la dimensión del subespacio $\mathfrak{M}$ es igual a r, es decir, al rango de la aplicación $\mathcal{A}$. Demostremos que el espacio $\mathfrak{L}$ es la suma directa de los subespacios $\mathfrak{M}$ y $\mathfrak{N}$. Para ello es suficiente demostrar, según el p. 6.2, que $\mathfrak{M} \cap \mathfrak{N} = \{o\}$ y que $\mathfrak{L} = \mathfrak{M} + \mathfrak{N}$. Demostremos lo primero. Todo vector de $\mathfrak{M}$ es de la forma

$$b = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r.$$

Si b pertenece a $\mathfrak{N}$, se tiene $b\mathcal{A} = o$, es decir,

$$(\alpha_1 b_1 + \ldots + \alpha_r b_r)\,\mathcal{A} = \alpha_1 a_1 + \ldots + \alpha_r a_r = o.$$

Pero los vectores $a_1, \ldots, a_r$ son linealmente independientes y por esto $\alpha_1 = \ldots = \alpha_r = 0$, de modo que $b = o$ que es lo que se quería demostrar. Resta probar que $\mathfrak{L} = \mathfrak{M} + \mathfrak{N}$. Tomemos un vector cualquiera a de $\mathfrak{L}$. Su imagen $a\mathcal{A}$ pertenece a $\mathfrak{L}\mathcal{A}$ y, por consiguiente, se expresa linealmente en términos de $a_1, \ldots, a_r$:

$$a\mathcal{A} = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_r a_r.$$

Sea

$$b = \alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_r b_r \quad \text{y} \quad a - b = c.$$

Puesto que

$$b\mathcal{A} = \alpha_1 \cdot b_1\mathcal{A} + \ldots + \alpha_r \cdot b_r\mathcal{A} = \alpha_1 a_1 + \ldots + \alpha_r a_r = a\mathcal{A},$$

se tiene

$$c\mathcal{A} = (a - b)\,\mathcal{A} = a\mathcal{A} - b\mathcal{A} = o.$$

Por consiguiente, c pertenece a $\mathfrak{N}$. Es decir, tenemos

$$a = b + c \quad (b \in \mathfrak{M}, \quad c \in \mathfrak{N});$$

pero esto significa precisamente que

$$\mathfrak{L} = \mathfrak{M} + \mathfrak{N}.$$

Como que esta suma es directa, la dimensión del espacio $\mathfrak{L}$ es igual a la suma de las dimensiones de los subespacios $\mathfrak{M}$ y $\mathfrak{N}$, es decir, es igual a la suma del rango y del defecto de la aplicación $\mathcal{A}$. Hemos demostrado el teorema.

Consideremos un ejemplo. Sea $\mathfrak{R}$ el espacio corriente de los segmentos orientados que parten de un punto O. Tomemos un plano cualquiera $\mathfrak{P}$ que pase por el punto O e indiquemos por $\mathcal{P}$ la operación de proyección ortogonal sobre $\mathfrak{P}$. La aplicación $\mathcal{P}$ transforma todo el espacio $\mathfrak{R}$ en el plano $\mathfrak{P}$. Es decir, $\mathfrak{P}$ es el dominio de valores de la aplicación $\mathcal{P}$ y el rango de $\mathcal{P}$ es igual a 2. El núcleo de la aplicación $\mathcal{P}$ está compuesto por los vectores que pertenecen a la recta que pasa por el punto O y que es perpendicular al plano $\mathfrak{P}$, ya que solamente estos vectores son transformados por la aplicación $\mathcal{P}$ en el nulo. Por consiguiente el defecto de $\mathcal{P}$ es igual a 1. La suma del rango y del defecto de la aplicación $\mathcal{P}$ es igual a 3 tal y como debe ser según el teorema 2.

10.2. Aplicaciones singulares y regulares. Más arriba (en el p. 9.1) hemos señalado que no toda aplicación es invertible. En lo sucesivo las aplicaciones lineales invertibles serán llamadas *regulares*, mientras que las aplicaciones no invertibles se llamarán *singulares*. En el p. 9.1 hemos encontrado en forma matricial las condiciones que garantizan que una aplicación lineal sea regular. Queremos dar ahora a estas condiciones un carácter geométrico.

TEOREMA 3. *Para que una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ sea regular es necesario y suficiente que el núcleo de esta aplicación sea nulo, es decir, que el defecto de $\mathcal{A}$ sea igual a cero.*

DEMOSTRACION. Si la aplicación dada $\mathcal{A}$ es regular, todo vector debe tener sólo una imagen recíproca; en particular, sólo una imagen recíproca debe tener el vector nulo o. Puesto que o siempre es una imagen recíproca del vector nulo y puesto que en este caso el núcleo está compuesto sólo de un vector, éste será precisamente el vector o.

Viceversa, sea el defecto de $\mathcal{A}$ igual a cero. En virtud del teorema 2, de aquí se deduce que el rango de $\mathcal{A}$ es igual a la dimensión de $\mathfrak{L}$, es decir, que la dimensión del dominio de valores $\mathfrak{L}\mathcal{A}$ es igual a la dimensión de $\mathfrak{L}$. Por consiguiente, $\mathfrak{L}\mathcal{A}=\mathfrak{L}$; vemos, pues, que todo vector de $\mathfrak{L}$ es la imagen de un vector de $\mathfrak{L}$. Si demostramos que la aplicación $\mathcal{A}$ transforma diferentes vectores a y b en diferentes vectores, esto significará precisamente que la aplicación $\mathcal{A}$ es invertible (p. 7.3). Pero de $a\mathcal{A}=b\mathcal{A}$ se deduce que $(a-b)\mathcal{A}=o$. Puesto que, por hipótesis, el núcleo de $\mathcal{A}$ es nulo, tenemos $a-b=o$, es decir, $a=b$ que es lo que se necesitaba. Hemos demostrado el teorema.

La igualdad a cero del defecto de $\mathcal{A}$ equivale a la coincidencia del rango de $\mathcal{A}$ y de la dimensión del espacio $\mathfrak{L}$. Por ello, el teorema 3 puede ser enunciado también en la forma siguiente:

TEOREMA 4. *Para que una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ sea regular es necesario y suficiente que el dominio de valores de $\mathcal{A}$ coincida con $\mathfrak{L}$, es decir, que el rango de $\mathcal{A}$ sea igual a la dimensión de $\mathfrak{L}$.*

Una aplicación biyectiva de un espacio lineal sobre otro se denomina *isomorfismo*, si transforma una suma de vectores del primer espacio en la suma de los vectores correspondientes del segundo espacio y si transforma, además, el producto de un número por un vector del primer espacio en el producto del mismo número por el vector correspondiente del segundo espacio (p. 4.3). Si ambos espacios lineales coinciden, obtenemos una aplicación isomorfa de un espacio lineal sobre sí mismo. Toda aplicación de este tipo se llama *automorfismo* de un espacio lineal. La definición de automorfismo coincide, obviamente, con la definición de una aplicación lineal regular. Por consiguiente, las aplicaciones lineales regulares de un espacio $\mathfrak{L}$ pueden ser consideradas como automorfismos de este espacio. De la definición misma de automorfismo resulta que los automorfismos de un espacio $\mathfrak{L}$ son aquellas superposiciones del espacio $\mathfrak{L}$ sobre sí mismo que conservan todas sus propiedades geométricas, es decir, las propiedades que se enuncian en términos de las operaciones de adición y de multiplicación por número.

Consideremos dos aplicaciones lineales arbitrarias $\mathcal{A}$ y $\mathcal{B}$ de un espacio $\mathfrak{L}$. Convendremos en llamar estas aplicaciones *isomorfas* o *semejantes*, si existe un automorfismo $\mathcal{C}$ del espacio $\mathfrak{L}$ que transforma una aplicación en la otra.

Sea u un vector de $\mathfrak{L}$ y sea $v=u\mathcal{A}$. El automorfismo $\mathcal{C}$ transforma u en un vector x y v en un vector y. Se dice que $\mathcal{C}$ transforma la aplicación $\mathcal{A}$ en $\mathcal{B}$, si $x\mathcal{B}=y$ (véase la fig. 3). Puesto que $x=u\mathcal{C}$ e $y=v\mathcal{C}$, de la igualdad $x\mathcal{B}=y$ resulta

$$u\mathcal{C}\mathcal{B}=v\mathcal{C} \quad \text{y} \quad u\mathcal{C}\mathcal{B}\mathcal{C}^{-1}=v=u\mathcal{A}.$$

De aquí

$$\mathcal{C}\mathcal{B}\mathcal{C}^{-1}=\mathcal{A} \quad \text{y} \quad \mathcal{B}=\mathcal{C}^{-1}\mathcal{A}\mathcal{C}. \tag{1}$$

Por consiguiente, *una aplicación $\mathcal{B}$ es isomorfa a una aplicación $\mathcal{A}$ si, y sólo si, se obtiene transformando $\mathcal{A}$ por un automorfismo del espacio $\mathfrak{L}$, es decir, por una aplicación regular de este espacio.*

Tomemos en $\mathfrak{L}$ un sistema de coordenadas y sean A, B y C las matrices respectivas de las aplicaciones $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$. Entonces la igualdad (1) equivale a la relación matricial

$$B=C^{-1}AC$$

y llegamos a la siguiente conclusión: *para que dos aplicaciones lineales de un espacio $\mathfrak{L}$ sean isomorfas es necesario y suficiente que sus matrices sean semejantes.*

Debemos considerar las aplicaciones isomorfas como aplicaciones que tienen las mismas propiedades geométricas. De aquí la importancia de saber clasificar, salvo un isomorfismo, todas las aplicaciones lineales. Algebraicamente este problema equivale a la clasificación, salvo semejanza, de todas las matrices cuadradas de orden n. El problema de clasificación de todos los espacios lineales sobre un cuerpo dado se resuelve sin dificultad (véase el p. 4.3); en cambio, el problema de la clasificación de las aplicaciones lineales exige para su resolución un estudio más detallado de las propiedades de las aplicaciones lineales. Este problema quedará totalmente resuelto sólo en el capítulo siguiente.

Señalemos, para concluir, una propiedad más de las aplicaciones lineales isomorfas: *para que las aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ sean isomorfas es necesario y suficiente que existan unos sistemas de coor-*

denadas en los que estas aplicaciones se representen por una misma matriz.

En efecto, si $\mathcal{A}$ y $\mathcal{B}$ son isomorfas, sus matrices están ligadas por la relación $A = CBC^{-1}$. Pero esta relación muestra que, si pasamos del sistema de coordenadas dado a otro de manera que la matriz del cambio sea C, la matriz de la aplicación $\mathcal{B}$ en el nuevo sistema de coordenadas será A, es decir, coincidirá con la matriz de la aplicación $\mathcal{A}$ calculada en el sistema antiguo de coordenadas. Viceversa, si la matriz A de la aplicación $\mathcal{A}$ en el sistema de coordenadas $a_1, \ldots, a_n$ coincide con la matriz de la aplicación $\mathcal{B}$ calculada en el sistema de coordenadas $a'_1, \ldots, a'_n$ y si C es la matriz del cambio del primer sistema de coordenadas por el segundo, tendremos que las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ en el segundo sistema de coordenadas serán iguales, respectivamente, a CAC^{-1} y a A, es decir, serán semejantes, que es lo que se quería demostrar.

10.3. Rango de la matriz de una aplicación. En el p. 10.1 hemos introducido el concepto del rango de una aplicación lineal. Por otro lado, conocemos del p. 5.2 el concepto del rango de una matriz que es el número máximo de sus filas linealmente independientes. Surge el problema sobre la relación que existe entre estos conceptos. La respuesta viene dada por el teorema siguiente:

TEOREMA 5. *El rango de una aplicación lineal cualquiera $\mathcal{A}$ de un espacio $\mathfrak{L}$ coincide con el rango de la matriz de esta aplicación.*

Sea $a_1, \ldots, a_n$ un sistema de coordenadas de $\mathfrak{L}$. El dominio de valores de la aplicación $\mathcal{A}$ está compuesto por las imágenes de los vectores del espacio $\mathfrak{L}$, es decir, por vectores de tipo

$$(\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n)\mathcal{A} = \alpha_1 (a_1\mathcal{A}) + \alpha_2 (a_2\mathcal{A}) + \ldots + \alpha_n (a_n\mathcal{A}).$$

De aquí se desprende que el dominio de valores de la aplicación $\mathcal{A}$ es el subespacio tendido sobre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$. El rango de $\mathcal{A}$ es igual a la dimensión del dominio de valores y, por consiguiente, al número máximo de vectores linealmente independientes entre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ (p. 6.1). Sea

$$\begin{aligned} a_1\mathcal{A} &= \alpha_{11}a_1 + \alpha_{12}a_2 + \ldots + \alpha_{1n}a_n, \\ a_2\mathcal{A} &= \alpha_{21}a_1 + \alpha_{22}a_2 + \ldots + \alpha_{2n}a_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ a_n\mathcal{A} &= \alpha_{n1}a_1 + \alpha_{n2}a_2 + \ldots + \alpha_{nn}a_n. \end{aligned}$$

La matriz

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{bmatrix}$$

es la matriz de la aplicación $\mathcal{A}$. Las filas de esta matriz son las filas de coordenadas de los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$. Luego, el número máximo de vectores linealmente independientes entre los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ es igual al número máximo de filas linealmente independientes de la matriz A, es decir, es igual a su rango que es lo que se quería demostrar.

Se llama *defecto* de una matriz cuadrada la diferencia entre su orden y su rango. De los teoremas 2 y 5 se desprende directamente que *el defecto de una aplicación lineal es igual al defecto de su matriz.*

Consideremos a título de ejemplo un sistema de n ecuaciones lineales homogéneas

$$\left.\begin{array}{l} \xi_1\alpha_{11}+\xi_2\alpha_{21}+\ldots+\xi_n\alpha_{n1}=0, \\ \xi_1\alpha_{12}+\xi_2\alpha_{22}+\ldots+\xi_n\alpha_{n2}=0, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \xi_1\alpha_{1n}+\xi_2\alpha_{2n}+\ldots+\xi_n\alpha_{nn}=0, \end{array}\right\} \tag{2}$$

con n incógnitas $\xi_1, \xi_2, \ldots, \xi_n$. Queremos estudiar las soluciones de este sistema. Geométricamente este problema puede ser interpretado de la forma siguiente. Tomemos un espacio lineal cualquiera $\mathfrak{L}$ de n dimensiones y escojamos en él un sistema de coordenadas $a_1, \ldots, a_n$. Convengamos en interpretar las magnitudes incógnitas $\xi_1, \ldots, \xi_n$ como las coordenadas de un vector x de $\mathfrak{L}$. Consideremos la aplicación lineal $\mathcal{A}$ de este espacio cuya matriz A está compuesta por los elementos α_{ij} $(i, j=1, 2, \ldots, n)$. El sistema de ecuaciones (2) puede ser representado en la forma

$$[x]\,A=O$$

o en la forma vectorial

$$x\mathcal{A}=o.$$

De aquí se ve que las soluciones del sistema (2) son las filas de coordenada de los vectores pertenecientes al núcleo de la aplicación $\mathcal{A}$. Puesto que la dimensión del núcleo coincide con el defecto de la aplicación y el defecto de la aplicación es igual al defecto de su matriz, llegamos al conocido teorema de la teoría de los determinantes: *el número máximo de soluciones linealmente independientes de un sistema homogéneo de n ecuaciones lineales con n incógnitas es igual al defecto de la matriz de este sistema.*

Ejemplos y problemas

1. En el espacio de filas de longitud 4 se ha tomado el sistema de coordenadas

$$e_1=[1,\ 0,\ 0,\ 0],\quad e_2=[0,\ 1,\ 0,\ 0],\quad e_3=[0,\ 0,\ 1,\ 0]\quad \text{y}\quad e_4=[0,\ 0,\ 0,\ 1].$$

Hállense el núcleo y el dominio de valores de las aplicaciones lineales definidas por las matrices

$$\begin{bmatrix} 2 & & & \\ & 0 & & \\ & & 3 & \\ & & & 0 \end{bmatrix},\quad \begin{bmatrix} 2 & 1 & 0 & 0 \\ & 2 & 1 & 0 \\ & & 2 & 0 \\ & & & 0 \end{bmatrix},\quad \begin{bmatrix} 3 & 1 & & \\ 2 & 1 & & \\ & & 0 & 0 \\ & & 0 & 0 \end{bmatrix}\ \text{y}\ \begin{bmatrix} 1 & 3 & 5 & 1 \\ 2 & 1 & 3 & 1 \\ 4 & 7 & 13 & 3 \\ 3 & -1 & 1 & 1 \end{bmatrix}.$$

2. Sea $\mathcal{A}$ una aplicación lineal de un espacio $\mathfrak{L}$ y sea $\mathfrak{M}$ un subespacio lineal de $\mathfrak{L}$. Demuéstrese que la dimensión de la imagen del subespacio $\mathfrak{M}$ satisface las desigualdades

$$\dim.\,\mathfrak{M}-\text{def.}\,\mathcal{A}\leqslant\dim.\,\mathfrak{M}\mathcal{A}\leqslant\dim.\,\mathfrak{M}.$$

3. Si $\mathcal{A}_1$ y $\mathcal{A}_2$ son unas aplicaciones lineales arbitraria y regular, respectivamente, se tiene

$$\text{rango } (\mathcal{A}_1\mathcal{A}_2) = \text{rango } (\mathcal{A}_2\mathcal{A}_1) = \text{rango } \mathcal{A}_1.$$

4. Sean $\mathcal{A}_1$ y $\mathcal{A}_2$ unas aplicaciones lineales cualesquiera de un espacio lineal $\mathfrak{L}$. Entonces se tiene

$$\text{rango } (\mathcal{A}_1 + \mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_1 + \text{rango } \mathcal{A}_2,$$
$$\text{def. } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{def. } \mathcal{A}_1 + \text{def. } \mathcal{A}_2,$$
$$\text{rango } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_1 \text{ y rango } (\mathcal{A}_1\mathcal{A}_2) \leqslant \text{rango } \mathcal{A}_2.$$

5. Toda aplicación lineal de rango m puede ser representada en forma de una suma de aplicaciones de rango 1.

6. Para que una matriz A de orden n tenga el rango no mayor que 1 es necesario y suficiente que A pueda ser representada en la forma

$$A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} [\beta_1\beta_2 \ldots \beta_n] = \| \alpha_i\beta_j \|,$$

donde α_i y β_j son números determinados.

§ 11. Subespacios invariantes

<u>11.1. Aplicación inducida.</u> Se dice que un subespacio $\mathfrak{A}$ de un espacio lineal $\mathfrak{L}$ *es invariante* respecto a una aplicación $\mathcal{A}$ si todo vector de $\mathfrak{A}$ se transforma por la aplicación $\mathcal{A}$ de nuevo en un vector de $\mathfrak{A}$, es decir, si

$$\mathfrak{A}\mathcal{A} \subset \mathfrak{A}.$$

De esta definición se desprende directamente que los subespacios impropios (el subespacio nulo y el propio espacio $\mathfrak{L}$) son invariantes respecto a cualquier aplicación lineal. También directamente se deduce la proposición de que *toda suma y toda intersección de subespacios invariantes es de nuevo un subespacio invariante.*

Observemos además que *siendo un subespacio $\mathfrak{A}$ invariante respecto a una aplicación $\mathcal{A}$, $\mathfrak{A}$ también será invariante respecto a la aplicación $f(\mathcal{A})$, donde $f(\mathcal{A}) = \alpha_0\mathcal{E} + \alpha_1\mathcal{A} + \ldots + \alpha_m\mathcal{A}^m$ es un polinomio arbitrario en $\mathcal{A}$.*

En efecto, si a es un vector de $\mathfrak{A}$, el vector $a\mathcal{A}$ está contenido, por hipótesis, en $\mathfrak{A}$. De aquí se deduce que $(a\mathcal{A})\mathcal{A} = a\mathcal{A}^2$ está contenido en $\mathfrak{A}$, etc. Vemos, por consiguiente, que el vector $a\mathcal{A}^k$ pertenece a $\mathfrak{A}$ cualquiera que sea $k \geqslant 0$. Pero en este caso $\mathfrak{A}$ contiene también todas las combinaciones lineales de estos vectores; en particular, $\mathfrak{A}$ contiene el vector

$$af(\mathcal{A}) = \alpha_0 a + \alpha_1 \cdot a\mathcal{A} + \ldots + \alpha_m \cdot a\mathcal{A}^m,$$

que es lo que se quería demostrar.

Los métodos de determinación de los subespacios invariantes serán considerados en el p. 11.4 y en el § 12; ahora, queremos sólo explicar cómo se pueden aprovechar los subespacios invariantes para simplificar la matriz de una aplicación.

Sea $\mathfrak{A}$ un subespacio invariante no trivial de una aplicación lineal $\mathcal{A}$. Tomemos en $\mathfrak{A}$ una base $a_1, a_2, \ldots, a_m$ complementándola con vectores linealmente independientes $a_{m+1}, \ldots, a_n$ hasta obtener una base de todo el espacio $\mathfrak{L}$. Para hallar la matriz de la aplicación $\mathcal{A}$ en el sistema de coordenadas $a_1, \ldots, a_m, a_{m+1}, \ldots, a_n$ es necesario expresar linealmente los vectores $a_1\mathcal{A}, \ldots, a_n\mathcal{A}$ en términos de los vectores coordenados $a_1, \ldots, a_n$. Pero el subespacio $\mathfrak{A}$ es invariante y, por ello, los vectores $a_1\mathcal{A}, \ldots, a_m\mathcal{A}$ pertenecen de nuevo a $\mathfrak{A}$ y se expresan linealmente en términos de $a_1, \ldots, a_m$. Por consiguiente, tenemos

$$\left.\begin{array}{l} a_1\mathcal{A} = \alpha_{11}a_1 + \ldots + \alpha_{1m}a_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_m\mathcal{A} = \alpha_{m1}a_1 + \ldots + \alpha_{mm}a_m, \\ a_{m+1}\mathcal{A} = \alpha_{m+1,1}a_1 + \ldots + \alpha_{m+1,m}a_m + \ldots + \alpha_{m+1,n}a_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_n\mathcal{A} = \alpha_{n1}a_1 + \ldots + \alpha_{nm}a_m + \ldots + \alpha_{nn}a_n, \end{array}\right\} \quad (1)$$

es decir, la matriz de la aplicación $\mathcal{A}$ es igual a

$$A = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{m1} & 0 & \ldots & 0 \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \alpha_{m1} & \ldots & \alpha_{mm} & 0 & \ldots & 0 \\ \alpha_{m+1,1} & \ldots & \alpha_{m+1,m} & \alpha_{m+1,m+1} & \ldots & \alpha_{m+1,n} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \alpha_{n1} & \ldots & \alpha_{nm} & \alpha_{n,m+1} & \ldots & \alpha_{nn} \end{bmatrix} = \begin{bmatrix} A_1 & O \\ B & A_2 \end{bmatrix}. \quad (2)$$

Resumiendo, *si una aplicación lineal posee un subespacio lineal invariante, su matriz en un sistema de coordenadas adecuado se descompone en cuatro células con la particularidad de que las células diagonales son cuadradas, mientras que la célula superior de la derecha puede resultar rectangular, pero formada íntegramente por ceros.* Las matrices de este tipo han sido llamadas en el p. 2.1 *semidescompuestas*. Recíprocamente, si en un sistema de coordenadas la matriz de una aplicación lineal $\mathcal{A}$ tiene la forma semidescompuesta (2), las igualdades (1) muestran que en este caso el subespacio $\mathfrak{A}$ tendido sobre los m primeros vectores coordenados será invariante respecto de $\mathcal{A}$.

Geométricamente la matriz A_1 puede ser interpretada de modo siguiente. Tenemos que la aplicación $\mathcal{A}$ transforma todo vector de $\mathfrak{A}$ de nuevo en un vector de $\mathfrak{A}$. Por esto $\mathcal{A}$ puede ser considerada también como una aplicación del espacio $\mathfrak{A}$. Indiquemos esta aplicación por $\mathcal{A}_1$ y convengamos en llamar $\mathcal{A}_1$ aplicación *inducida*. Las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ actúan sobre los vectores del subespacio $\mathfrak{A}$ idénticamente: si a es un vector de $\mathfrak{A}$, se tiene $a\mathcal{A}=a\mathcal{A}_1$. La diferencia entre estas aplicaciones consiste en que tienen distintos campos de definición: si a es un vector del espacio principal $\mathfrak{L}$ que no pertenece a $\mathfrak{A}$, la operación $a\mathcal{A}$ tiene sentido, mientras que $a\mathcal{A}_1$ no lo tiene.

Las primeras m igualdades del sistema (1) muestran que A_1 *es la matriz de la aplicación inducida* $\mathcal{A}_1$ en el sistema de coordenadas $a_1, a_2, \ldots, a_m$.

11.2. Suma directa de subespacios invariantes. Hemos considerado el caso en que la aplicación lineal $\mathcal{A}$ tiene sólo un subespacio invariante. Supongamos ahora que $\mathcal{A}$ posee dos subespacios invariantes $\mathfrak{A}_1$ y $\mathfrak{A}_2$ y, es más, supongamos que el espacio $\mathfrak{L}$ es la suma directa de estos subespacios. Tomemos en $\mathfrak{A}_1$ y $\mathfrak{A}_2$ unos sistemas de coordenadas $a_1, \ldots, a_m$ y $a_{m+1}, \ldots, a_n$, respectivamente. Según el p. 6.2, los vectores $a_1, \ldots, a_m, a_{m+1}, \ldots, a_n$ forman un sistema de coordenadas de $\mathfrak{L}$. Veamos la forma que toma en este sistema la matriz de la aplicación $\mathcal{A}$. Por hipótesis, los vectores $a_1\mathcal{A}, \ldots, a_m\mathcal{A}$ pertenecen a $\mathfrak{A}_1$ y los vectores $a_{m+1}\mathcal{A}, \ldots, a_n\mathcal{A}$ pertenecen a $\mathfrak{A}_2$. Luego, tenemos

$$\left.\begin{array}{ll}
a_1\mathcal{A}=\alpha_{11}a_1+\ldots+\alpha_{1m}a_m, & \\
\ldots\ldots\ldots\ldots\ldots & \\
a_m\mathcal{A}=\alpha_{m1}a_1+\ldots+\alpha_{mm}a_m, & \\
a_{m+1}\mathcal{A}= & \alpha_{m+1,\,m+1}a_{m+1}+\ldots+\alpha_{m+1,n}\,a_n, \\
\ldots\ldots & \ldots\ldots\ldots\ldots\ldots\ldots \\
a_n\mathcal{A}= & \alpha_{n,\,m+1}a_{m+1}+\ldots+\alpha_{nn}a_n.
\end{array}\right\} \qquad (3)$$

Por consiguiente, la matriz de la aplicación $\mathcal{A}$ resulta ser igual a

$$A=\begin{bmatrix}
\alpha_{11} & \ldots & \alpha_{1m} & 0 & \ldots & 0 \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\alpha_{m1} & \ldots & \alpha_{mm} & 0 & \ldots & 0 \\
0 & \ldots & 0 & \alpha_{m+1,\,m+1} & \ldots & \alpha_{m+1,\,n} \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & & \cdot \\
0 & \ldots & 0 & \alpha_{n,\,m+1} & \ldots & \alpha_{nn}
\end{bmatrix}=\begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix}. \qquad (4)$$

es decir, resulta ser descompuesta. Sean $\mathcal{A}_1$ y $\mathcal{A}_2$ las aplicaciones inducidas en los subespacios $\mathfrak{A}_1$ y $\mathfrak{A}_2$ por la aplicación $\mathcal{A}$. De las igualdades (3) se deduce que A_1 y A_2 son las matrices de las aplicaciones $\mathcal{A}_1$ y $\mathcal{A}_2$ en los sistemas correspondientes de coordenadas.

Resumiendo, *si un espacio $\mathfrak{L}$ se descompone en una suma directa de subespacios invariantes respecto a una aplicación lineal $\mathcal{A}$, la matriz de la aplicación $\mathcal{A}$ calculada en un sistema adecuado de coordenadas toma la forma celular diagonal y sus células diagonales representan las matrices de las aplicaciones inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes.*

Hemos demostrado esta proposición sólo en el caso de una suma de dos subespacios invariantes. Sin embargo, todos los razonamientos se traspasan sin modificaciones al caso de un número arbitrario de sumandos.

Supongamos ahora lo contrario aceptando que en un sistema de coordenadas la matriz de la aplicación $\mathcal{A}$ toma la forma descompuesta (4). Entonces de las igualdades (3) se desprende que el espacio $\mathfrak{A}_1$ tendido sobre los m primeros vectores coordenados y el espacio $\mathfrak{A}_2$ tendido sobre los restantes vectores coordenados serán invariantes respecto de $\mathcal{A}$. Es obvio que la suma $\mathfrak{A}_1+\mathfrak{A}_2$ es directa y coincide con $\mathfrak{L}$. Por consiguiente, la condición de que la matriz de la aplicación $\mathcal{A}$ se reduce a la forma celular diagonal, además de ser necesaria, es también suficiente para que $\mathfrak{L}$ sea la suma directa de los subespacios invariantes respecto de $\mathcal{A}$.

Examinemos el problema siguiente. Se tiene una descomposición de un espacio $\mathfrak{L}$ en la suma directa de unos subespacios lineales

$$\mathfrak{L}=\mathfrak{A}_1\dot{+}\mathfrak{A}_2\dot{+}\dots\dot{+}\mathfrak{A}_s$$

y en cada subespacio $\mathfrak{A}_i$ se tiene una aplicación lineal $\mathcal{A}_i$. ¿Existe una aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$ respecto a la cual todos los subespacios $\mathfrak{A}_i$ son invariantes y que induce en todo $\mathfrak{A}_i$ la aplicación $\mathcal{A}_i$? ¿Será esta aplicación única? La respuesta es, obviamente, afirmativa. En efecto, tomemos en cada uno de los subespacios $\mathfrak{A}_i$ una base $a_{i1}, a_{i2}, \dots, a_{im_i}$ e indiquemos por A_i la matriz de la aplicación $\mathcal{A}_i$ en este sistema de coordenadas. Consideremos la matriz celular diagonal

$$A=A_1\dot{+}A_2\dot{+}\dots\dot{+}A_s.$$

El sistema formado por los vectores $a_{11}, \dots, a_{1m_1}, \dots, a_{s1}, \dots, a_{sm_s}$ es una base del espacio $\mathfrak{L}$. En esta base a la matriz A le corresponde una aplicación lineal $\mathcal{A}$ del espacio $\mathfrak{L}$. En virtud de lo expuesto anteriormente, la aplicación $\mathcal{A}$ satisface todas las condiciones de nuestro problema. Esta aplicación es única, ya que su matriz en el sistema de coordenadas indicado se determina unívocamente por las condiciones del problema.

11.3. Polinomio característico de una aplicación. En este punto, así como en el siguiente, se supone que el cuerpo principal es un cuerpo conmutativo. Tomemos una aplicación lineal cualquiera $\mathcal{A}$ de un espacio lineal $\mathfrak{L}$ de n dimensiones. Escogiendo en $\mathfrak{L}$ un sistema de coordenadas determinado $a_1, \ldots, a_n$ podemos calcular la matriz A de la aplicación $\mathcal{A}$. El polinomio característico $\varphi(\lambda) = |\lambda E - A|$ de la matriz A se llama *polinomio característico de la aplicación* $\mathcal{A}$. Si tomamos otro sistema de coordenadas $a'_1, \ldots, a'_n$ e indicamos por T la matriz del cambio, la matriz de la aplicación $\mathcal{A}$ en el nuevo sistema de coordenadas será, de acuerdo con el p. 8.3, la matriz

$$A_1 = TAT^{-1},$$

es decir, la matriz semejante de A. Sin embargo, en el p. 3.2 hemos demostrado que las matrices semejantes tienen los mismos polinomios característicos. Por consiguiente, *el polinomio característico de una aplicación* $\mathcal{A}$ *no depende del sistema de coordenadas en el que se calcula.*

El grado del polinomio característico es igual al orden de la matriz A y el orden de la matriz A es igual a la dimensión del espacio $\mathfrak{L}$. Por esto, *el grado del polinomio característico de una aplicación* $\mathcal{A}$ *es igual a la dimensión del espacio en el que actúa esta aplicación.*

La suma de las raíces del polinomio característico es igual a la traza y el producto de las raíces es igual al determinante de la matriz A. Puesto que el polinomio característico de $\mathcal{A}$ y, por consiguiente, también sus raíces no dependen del sistema de coordenadas, tampoco la traza y el determinante de A dependerán del sistema de coordenadas. Por esta razón la traza y el determinante de la matriz de una aplicación $\mathcal{A}$ se llaman *traza y determinante de la aplicación* $\mathcal{A}$.

Si el espacio $\mathfrak{L}$ se descompone en una suma directa de los subespacios $\mathfrak{A}_1$ y $\mathfrak{A}_2$ invariantes respecto de $\mathcal{A}$, la matriz de la aplicación $\mathcal{A}$ toma, en un sistema de coordenadas adecuado, la forma celular diagonal

$$A = \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix}.$$

Según el p. 3.3. el polinomio característico de la matriz A es igual en este caso al producto de los polinomios característicos de las matrices A_1 y A_2. Pero A_1 y A_2 son las matrices de las aplicaciones lineales inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes $\mathfrak{A}_1$ y $\mathfrak{A}_2$. Por lo tanto, *si un espacio* $\mathfrak{L}$ *se descompone en la suma directa de subespacios invariantes respecto a una aplicación lineal* $\mathcal{A}$, *el polinomio característico de la aplicación* $\mathcal{A}$ *es igual al* *producto de los polinomios característicos de las aplicaciones inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes.*

Según el teorema de Hamilton—Cayley (p. 3.2), toda matriz cuadrada A es raíz de su polinomio característico $\varphi(\lambda)$, es decir, $\varphi(A)=O$. Sea $\mathcal{A}$ una aplicación lineal de matriz A. La matriz de la aplicación $\varphi(\mathcal{A})$ es, de acuerdo con el p. 9.3, $\varphi(A)$. Como que esta matriz es nula, tenemos $\varphi(\mathcal{A})=\mathcal{O}$. Por consiguiente, *toda aplicación lineal es raíz de su polinomio característico.*

Se llama *polinomio mínimo de una aplicación lineal* $\mathcal{A}$ el polinomio de menor grado de coeficiente principal igual a 1 para el que la aplicación $\mathcal{A}$ es una raíz. Sea A la matriz de la aplicación $\mathcal{A}$ calculada en un sistema de coordenadas. Puesto que las relaciones $f(A)=O$ y $f(\mathcal{A})=\mathcal{O}$, donde $f(\lambda)$ es un polinomio arbitrario, son equivalentes, resulta que *el polinomio mínimo de una aplicación coincide con el polinomio mínimo de la matriz de esta aplicación.*

Si el espacio $\mathfrak{L}$ se descompone en suma directa de subespacios invariantes respecto a una aplicación $\mathcal{A}$, la matriz de la aplicación $\mathcal{A}$, en un sistema de coordenadas adecuado, se descompone. El polinomio mínimo de una matriz descompuesta es el mínimo común múltiplo de los polinomios mínimos de sus células diagonales (p. 3.3). Por esto, el polinomio mínimo de una aplicación $\mathcal{A}$ será igual al mínimo común múltiplo de los polinomios mínimos de las aplicaciones inducidas por la aplicación $\mathcal{A}$ en los subespacios invariantes.

11.4. Vectores propios y valores propios. Continuemos suponiendo que el cuerpo principal es un cuerpo conmutativo. Queremos ahora estudiar más detalladamente los subespacios invariantes de una dimensión. Introduzcamos primero la definición siguiente. Un número ζ se llama *valor propio* de una aplicación lineal $\mathcal{A}$, si existe en el espacio $\mathfrak{L}$ un vector no nulo a, tal que

$$a\mathcal{A}=\zeta a. \tag{5}$$

Todo vector que satisface esta relación se llama *vector propio* de la aplicación $\mathcal{A}$ correspondiente al valor propio ζ.

La búsqueda de los vectores propios y la búsqueda de los subespacios invariantes de una dimensión son problemas equivalentes. Efectivamente, sea a un vector no nulo propio de una aplicación $\mathcal{A}$ y sea ζ su valor propio correspondiente. Consideremos el subespacio de una dimensión $\mathfrak{A}$ tendido sobre el vector a, es decir, el conjunto de todos los vectores de tipo αa. La relación

$$(\alpha a)\mathcal{A}=\alpha(a\mathcal{A})=\zeta\alpha\cdot a \tag{6}$$

muestra que $\mathfrak{A}$ es invariante respecto de $\mathcal{A}$. Recíprocamente, sea $\mathfrak{A}$ un subespacio de dimensión uno invariante respecto de $\mathcal{A}$. Tomemos en $\mathfrak{A}$ un vector arbitrario no nulo a. Puesto que $\mathfrak{A}$ es de una dimensión, todos los vectores de $\mathfrak{A}$ son de la forma αa. Por hipótesis, $a\mathcal{A}$ pertenece a $\mathfrak{A}$; luego,

$$a\mathcal{A} = \zeta a,$$

es decir, a es un vector propio de la aplicación $\mathcal{A}$ correspondiente al valor propio ζ. La igualdad (6) muestra que todos los demás vectores de $\mathfrak{A}$ también son vectores propios correspondientes al valor propio ζ.

Escojamos en el espacio $\mathfrak{L}$ un sistema de coordenadas $a_1, \ldots, a_n$ y sea $A = \|\alpha_{ij}\|$ la matriz de una aplicación lineal $\mathcal{A}$ en este sistema. Indiquemos por a algún vector propio no nulo de la aplicación $\mathcal{A}$. Sea $[\xi_1, \xi_2, \ldots, \xi_n]$ la fila de coordenadas del vector a y sea ζ el valor propio correspondiente. Pasando en la igualdad (5) a las coordenadas, obtenemos

$$[a]\, A = \zeta\, [a], \tag{7}$$

o

$$\left.\begin{array}{l} \xi_1\alpha_{11} + \xi_2\alpha_{21} + \ldots + \xi_n\alpha_{n1} = \zeta\xi_1, \\ \xi_1\alpha_{12} + \xi_2\alpha_{22} + \ldots + \xi_n\alpha_{n2} = \zeta\xi_2, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \xi_1\alpha_{1n} + \xi_2\alpha_{2n} + \ldots + \xi_n\alpha_{nn} = \zeta\xi_n. \end{array}\right\} \tag{8}$$

Pasando todos los términos a un mismo miembro, obtenemos

$$\left.\begin{array}{r} \xi_1(\zeta - \alpha_{11}) - \xi_2\alpha_{21} - \ldots - \xi_n\alpha_{n1} = 0, \\ -\xi_1\alpha_{12} + \xi_2(\zeta - \alpha_{22}) - \ldots - \xi_n\alpha_{n2} = 0, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ -\xi_1\alpha_{1n} - \xi_2\alpha_{2n} - \ldots + \xi_n(\zeta - \alpha_{nn}) = 0. \end{array}\right\} \tag{9}$$

Este sistema puede ser considerado como un sistema de n ecuaciones lineales homogéneas con n incógnitas $\xi_1, \xi_2, \ldots, \xi_n$. Puesto que las coordenadas del vector *no nulo* propio a satisfacen el sistema (9), tenemos (véase el p. 10.3)

$$\begin{vmatrix} \zeta - \alpha_{11} & -\alpha_{12} & \ldots & -\alpha_{1n} \\ -\alpha_{21} & \zeta - \alpha_{22} & \ldots & -\alpha_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ -\alpha_{n1} & -\alpha_{n2} & \ldots & \zeta - \alpha_{nn} \end{vmatrix} = |\zeta E - A| = 0, \tag{10}$$

donde E es la matriz unidad. Pero $|\lambda E - A|$ es el polinomio característico de la matriz A; por ello, de la igualdad (10) se desprende que *todos los valores propios de una aplicación lineal son raíces de su polinomio característico*. Recíprocamente, *si ζ es una raíz del polinomio característico de una aplicación $\mathcal{A}$ que pertenece al cuerpo conmutativo de coeficientes del espacio lineal, resulta que ζ es un valor propio de la aplicación $\mathcal{A}$*. En efecto, la igualdad (10) muestra que el rango de la matriz del sistema (9) será menor que n; por consiguiente, este sistema tendrá al menos una solución no nula. Indicando esta solución por $[\xi_1, \xi_2, \ldots, \xi_n]$, obtenemos directamente de (8) y de (7) que el vector a de coordenadas $\xi_1, \xi_2, \ldots, \xi_n$ será el vector propio no nulo deseado[1].

Se llama *multiplicidad* de un valor propio ζ de una aplicación lineal $\mathcal{A}$ la multiplicidad que tiene ζ como raíz del polinomio característico de la aplicación $\mathcal{A}$.

Consideremos un ejemplo. Supongamos que en el espacio lineal real de tres dimensiones de base a_1, a_2 y a_3 actúa una aplicación lineal $\mathcal{A}$ de matriz

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

Se necesita hallar los valores propios y los vectores propios de la aplicación $\mathcal{A}$. Calculamos, ante todo, el polinomio característico de la aplicación $\mathcal{A}$:

$$\begin{bmatrix} \lambda-3 & -3 & -2 \\ -1 & \lambda-1 & 2 \\ 3 & 1 & \lambda \end{bmatrix}=(\lambda-4)(\lambda^2+4).$$

Sus raíces son iguales a $\lambda_1=4$, $\lambda_2=2i$ y $\lambda_3=-2i$. Puesto que el cuerpo principal es real, los dos últimos valores deben ser omitidos, mientras que el primer valor $\lambda_1=4$ será el valor propio buscado. Para hallar los vectores propios formamos el sistema (9) que, en nuestro caso, se convierte en

$$\left.\begin{aligned} \xi_1-\xi_2+3\xi_3&=0, \\ -3\xi_1+3\xi_2+\xi_3&=0, \\ -2\xi_1+2\xi_2+4\xi_3&=0. \end{aligned}\right\}$$

Resolviendo este sistema obtenemos

$$\xi_2=\xi_1 \quad \text{y} \quad \xi_3=0,$$

donde ξ_1 es arbitrario. Por consiguiente, el vector $\xi_1 a_1+\xi_1 a_2$ será un vector propio de la aplicación $\mathcal{A}$ cualquiera que sea ξ_1.

[1] Existe una demostración más breve de la última proposición. La condición $a\mathcal{A}=\zeta a$ puede ser representada en la forma $a(\zeta\mathcal{E}-\mathcal{A})=o$. Esto demuestra que los vectores propios de $\mathcal{A}$ son vectores que pertenecen al núcleo de la aplicación $\zeta\mathcal{E}-\mathcal{A}$. Pero, para que el núcleo contenga vectores no nulos es necesario y suficiente que la aplicación sea singular (p. 10.2), es decir, es necesario y suficiente que sea $|\zeta E-A|=0$.

Ejemplos y problemas

1. En un espacio complejo $\mathfrak{L}$ de base a_1, a_2, a_3 y a_4 está definida una aplicación $\mathcal{A}$ de matriz

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

Hállense los vectores propios y los valores propios de la aplicación $\mathcal{A}$. Demuéstrese que el subespacio tendido sobre los vectores $2a_1-a_2$, y $-a_3+a_4$ es invariante respecto de $\mathcal{A}$.

2. Supongamos que un espacio $\mathfrak{L}$ tiene una base formada por los vectores propios de una aplicación $\mathcal{A}$. ¿Cuál será la matriz de la aplicación $\mathcal{A}$ en esta base?

3. Si en un espacio $\mathfrak{L}$ con un sistema de coordenadas $a_1, a_2, \ldots, a_n$ la matriz de una aplicación lineal $\mathcal{A}$ es de forma celular semidescompuesta

$$A=\begin{bmatrix} A_1 & B \\ O & A_2 \end{bmatrix},$$

donde A_2 es una matriz cuadrada de orden m, el subespacio tendido sobre los m últimos vectores coordenados $a_{n-m+1}, \ldots, a_n$ será invariante respecto de $\mathcal{A}$.

4. Si la aplicación $\mathcal{A}$ es regular, todo subespacio invariante respecto de $\mathcal{A}$ también será invariante respecto de $\mathcal{A}^{-1}$.

5. Si un subespacio $\mathfrak{A}$ es invariante respecto a una aplicación lineal $\mathcal{A}$, la imagen y la imagen recíproca del subespacio $\mathfrak{A}$ también serán invariantes respecto de $\mathcal{A}$.

6. En un espacio lineal complejo toda aplicación lineal tiene al menos un vector propio no nulo.

7. Supongamos que en un sistema de coordenadas $a_1, \ldots, a_n$ la matriz de una aplicación $\mathcal{A}$ es de forma diagonal con diferentes elementos diagonales. Hállense todos los subespacios invariantes de la aplicación $\mathcal{A}$ y demuéstrese que el número de los mismos es igual a 2^n.

8. Si una aplicación lineal $\mathcal{A}$ de un espacio $\mathfrak{L}$ de n dimensiones tiene n valores propios diferentes, la matriz de la aplicación $\mathcal{A}$ se reduce, en un sistema de coordenadas adecuado, a la forma diagonal.

§ 12. Aplicaciones de matrices de forma normal

En este parágrafo serán examinadas las propiedades de las aplicaciones lineales, cuyas matrices tienen, en un sistema de coordenadas fijo, la así llamada forma normal de Jordan. Por consiguiente, supondremos de antemano que las matrices de las aplicaciones consideradas pueden ser reducidas a esta forma. Más adelante, en el p. 15.4, veremos que esta reducción es siempre posible en el cuerpo de los números complejos.

En todo este parágrafo se supone que el cuerpo principal es un cuerpo conmutativo.

12.1. Forma diagonal. El teorema que sigue ofrece la característica más simple de las aplicaciones, cuyas matrices pueden ser reducidas a la forma diagonal.

TEOREMA 1 *Si una aplicación lineal de un espacio de n dimensiones tiene n vectores propios linealmente independientes, entonces tomándolos como vectores coordenados reduciremos la matriz de la aplicación a la forma diagonal. Recíprocamente, si la matriz de una aplicación en un sistema de coordenadas es de forma diagonal, los vectores de la base son vectores propios de la aplicación.*

La demostración es evidente. Un problema más sutil consiste en averiguar, a partir de la matriz de una aplicación calculada en un sistema de coordenadas eventual, si la aplicación posee vectores propios que constituyan una base del espacio. Este problema quedará resuelto en el p. 15.4, mientras que ahora estudiaremos sólo un caso particular del mismo.

TEOREMA 2 *Los vectores propios correspondientes a diferentes valores propios de una aplicación lineal son linealmente independientes.*

En efecto, sean $\rho_1, \rho_2, \ldots, \rho_m$ distintos valores propios y sean $a_1, a_2, \ldots, a_m$ los vectores propios que les corresponden de una aplicación lineal $\mathcal{A}$. Por inducción, podemos aceptar que $a_1, \ldots, a_{m-1}$ son linealmente independientes. Supongamos que $a_1, \ldots, a_m$ están ligados por una relación

$$\alpha_1 a_1 + \ldots + \alpha_{m-1} a_{m-1} + \alpha_m a_m = o.$$

Aplicando a ambos miembros de esta igualdad la aplicación $\mathcal{A}$, obtenemos

$$\alpha_1 \rho_1 a_1 + \ldots + \alpha_{m-1} \rho_{m-1} a_{m-1} + \alpha_m \rho_m a_m = o.$$

Eliminando a_m, tendremos

$$\alpha_1 (\rho_m - \rho_1) a_1 + \ldots + \alpha_{m-1} (\rho_m - \rho_{m-1}) a_{m-1} = o.$$

Debido a la independencia lineal de $a_1, \ldots, a_{m-1}$ de aquí resulta $\alpha_1 = \ldots = \alpha_{m-1} = 0$ y, por consiguiente, $\alpha_m = 0$ que es lo que se quería demostrar.

Comparando ambos teoremas demostrados, obtenemos el corolario siguiente: *si el polinomio característico de una aplicación lineal de un espacio de n dimensiones tiene n diferentes raíces, la matriz de la aplicación se reduce, en un sistema de coordenadas adecuado, a la forma diagonal.*

Por ejemplo, el polinomio característico de la matriz

$$A = \begin{bmatrix} 1 & 3 & 1 & 2 \\ & -1 & 1 & 3 \\ & & 2 & 5 \\ & & & -2 \end{bmatrix}$$

tiene las raíces ± 1, ± 2; las filas de coordenadas de los vectores propios correspondientes son $[2, 3, -5, -4]$, $[0, 3, -1, 4]$, $[0, 0, 4, 5]$ y $[0, 0, 0, 1]$. Tomándolos como vectores coordenados reduciremos la matriz A a la forma diagonal con los números 1, -1, 2 y -2 a lo largo de la diagonal principal.

12.2. Células de Jordan. Una matriz de tipo

$$A=\begin{bmatrix} \rho & 1 & 0 & \dots & 0 \\ & \rho & 1 & \dots & 0 \\ & & \rho & \dots & 0 \\ & & \dots & \dots & \dots \\ & & & \rho & 1 \\ & & & & \rho \end{bmatrix} \tag{1}$$

se llama *célula de Jordan*. El polinomio característico de una célula de Jordan es igual a $(\lambda-\rho)^n$, donde n es el orden de la matriz. Por consiguiente, ρ es su único valor propio de multiplicidad n.

Sea $\mathfrak{L}$ un espacio lineal de base $e_1, e_2, \dots, e_n$ y sea $\mathcal{A}$ una aplicación lineal que en esta base tiene una matriz A de tipo (1). En este caso tenemos

$$e_1\mathcal{A}=\rho e_1+e_2, \quad \dots, \quad e_{n-1}\mathcal{A}=\rho e_{n-1}+e_n, \quad e_n\mathcal{A}=\rho e_n \tag{2}$$

y, por consiguiente,

$$e_1(\mathcal{A}-\rho\mathcal{E})=e_2, \quad e_1(\mathcal{A}-\rho\mathcal{E})^2=e_3, \quad \dots, \quad e_1(\mathcal{A}-\rho\mathcal{E})^{n-1}=e_n. \tag{3}$$

Puesto que $\mathcal{A}$ debe ser una raíz de su polinomio característico, se tiene $(\mathcal{A}-\rho\mathcal{E})^n=\mathcal{O}$. El polinomio mínimo de la aplicación $\mathcal{A}$ divide su polinomio característico y, por ello, debe ser de la forma $(\lambda-\rho)^s$, $0<s\leqslant n$. La última de las igualdades (3) muestra que $(\mathcal{A}-\rho\mathcal{E})^{n-1}\neq\mathcal{O}$, de manera que $s=n$, es decir, *el polinomio mínimo de una célula de Jordan coincide con su polinomio característico* $(\lambda-\rho)^n$.

Indiquemos por $\mathfrak{L}_i$ el subespacio tendido sobre los vectores e_i, $e_{i+1}, \dots, e_n$ $(i=1, 2, \dots, n)$ de la base. De las igualdades (2) y de la forma de la matriz A se deduce que todos estos subespacios son invariantes. Empleando las relaciones (3) es fácil comprobar que $\mathfrak{L}_i$ está compuesto por aquellos vectores, y sólo aquéllos, x para los cuales se tiene

$$x(\mathcal{A}-\rho\mathcal{E})^{n-i+1}=o.$$

Esto demuestra que la cadena de subespacios $\mathfrak{L}=\mathfrak{L}_1\supset\mathfrak{L}_2\supset\dots$ $\dots\supset\mathfrak{L}_n\supset o$ no depende de la selección del sistema de coordenadas y está definida por la propia aplicación $\mathcal{A}$.

Demostremos que la aplicación no tiene otros subespacios invariantes. En efecto, sea $\mathfrak{M}$ un subespacio invariante de $\mathcal{A}$ diferente de $\mathfrak{L}_i$. Busquemos un índice i tal que $\mathfrak{M}\supset\mathfrak{L}_i$ y $\mathfrak{M}\not\supset\mathfrak{L}_{i-1}$, aceptando, para generalizar, que $\mathfrak{L}_{n+1}=o$. Mostremos que $\mathfrak{M}=\mathfrak{L}_i$. Consideremos, para ello, un vector arbitrario

$$a=\alpha_j e_j+\alpha_{j+1}e_{j+1}+\dots+\alpha_n e_n \quad (\alpha_j\neq 0) \tag{4}$$

de $\mathfrak{M}$. Para $j \geqslant i$ tenemos $a \in \mathfrak{L}_i$. Supongamos que $j < i$. Aplicando a ambos miembros de (4) la aplicación $(\mathcal{A} - \rho\mathcal{E})^{i-j-1}$, obtenemos

$$a(\mathcal{A} - \rho\mathcal{E})^{i-j-1} = \alpha_j e_{i-1} + \alpha_{j+1} e_i + \ldots + \alpha_{n-i+j+1} e_n \in \mathfrak{M}.$$

Puesto que, por hipótesis, $e_i, \ldots, e_n \in \mathfrak{M}$, resulta que $e_{i-1} \in \mathfrak{M}$ y, por consiguiente, $\mathfrak{L}_{i-1} \subset \mathfrak{M}$ lo que contradice a la elección de i. Por esto tenemos $\mathfrak{M} \subset \mathfrak{L}_i$, de donde $\mathfrak{M} = \mathfrak{L}_i$.

Observemos además que *la matriz de la aplicación $\mathcal{A}$ no se descompone en ningún sistema de coordenadas.*

Efectivamente, la descomposición de la matriz de $\mathcal{A}$ equivale a la descomposición de $\mathfrak{L}$ en una suma directa de subespacios invariantes, lo que es imposible ya que uno de dos subespacios invariantes cualesquiera de la aplicación $\mathcal{A}$ debe estar contenido en el otro, mientras que los sumandos de una suma directa tienen intersección nula.

12.3. Subespacios radicales. Las aplicaciones lineales, cuyas matrices pueden ser reducidas a la forma diagonal o a una célula de Jordan, no abarcan todo el conjunto de matrices. En el caso general, la matriz de cualquier aplicación lineal sobre el cuerpo de los números complejos puede ser reducida a la forma celular diagonal con células de Jordan a lo largo de la diagonal. Se dice que las matrices de este último tipo son *de Jordan* o *que tienen la forma normal de Jordan.*

Supongamos que en una base $e_1, \ldots, e_n$ la matriz de una aplicación lineal $\mathcal{A}$ es de la forma normal de Jordan

$$A = A_1 \dotplus A_2 \dotplus \ldots \dotplus A_s, \tag{5}$$

donde A_i es la célula de Jordan de orden n_i con el valor propio ρ_i $(i = 1, \ldots, s)$ y $n_1 + \ldots + n_s = n$. A la célula A_i le corresponde el subespacio invariante $\mathfrak{L}^{(i)}$ tendido sobre los vectores $e_{p_i+1}, e_{p_i+2}, \ldots \ldots, e_{q_i}$ $(p_i = n_1 + \ldots + n_{i-1},\ q_i = p_i + n_i)$. La aplicación $\mathcal{A}$ induce en el subespacio $\mathfrak{L}^{(i)}$ una aplicación $\mathcal{A}_i$, cuya matriz es precisamente la célula A_i. Según lo expuesto, todos los vectores x de $\mathfrak{L}^{(i)}$ satisfacen la relación

$$x(\mathcal{A}_i - \rho_i\mathcal{E}_i)^{n_i} = o$$

y, por consiguiente, también la relación

$$x(\mathcal{A} - \rho_i\mathcal{E})^{n_i} = o. \tag{6}$$

Sin embargo, ahora ya no se puede afirmar que la relación (6) caracteriza sólo los vectores de $\mathfrak{L}^{(i)}$, ya que entre las células diagonales pueden aparecer otras células con el mismo valor propio. Con el fin de examinar este problema más detalladamente, introduciremos la definición siguiente.

Un vector a se llama *vector radical de altitud h correspondiente a la raíz ρ de una aplicación $\mathcal{A}$*, si

$$a(\rho\mathcal{E}-\mathcal{A})^h=o.$$

El concepto de vector radical es una generalización del concepto de vector propio, ya que los vectores propios son vectores radicales de altitud 1.

El conjunto de todos los vectores radicales correspondientes a una raíz fija ρ de una aplicación $\mathcal{A}$ es un subespacio invariante $\mathfrak{L}_\rho$ llamado subespacio radical de la aplicación $\mathcal{A}$.

Efectivamente, si x e y pertenecen a $\mathfrak{L}_\rho$ y son de altitud h_1 y h_2, tenemos para $h=\max(h_1, h_2)$

$$(\alpha x+\beta y)(\rho\mathcal{E}-\mathcal{A})^h=\alpha x(\rho\mathcal{E}-\mathcal{A})^h+\beta y(\rho\mathcal{E}-\mathcal{A})^h=o,$$
$$x\mathcal{A}(\rho\mathcal{E}-\mathcal{A})^h=x(\rho\mathcal{E}-\mathcal{A})^h\mathcal{A}=o.$$

Los vectores radicales correspondientes a diferentes raíces son necesariamente lienealmente independientes. Es más, tiene lugar un teorema más general.

TEOREMA 3. *Si una suma $x_1+x_2+\ldots+x_m=x$ de vectores radicales correspondientes a diferentes raíces $\rho_1, \ldots, \rho_m$ de la aplicación $\mathcal{A}$ está contenida en un subespacio invariante $\mathfrak{M}$, todo sumando por separado está contenido en $\mathfrak{M}$.*

Pongamos

$$\varphi(\lambda)=(\lambda-\rho_1)^{h_1}(\lambda-\rho_2)^{h_2}\ldots(\lambda-\rho_{m-1})^{h_{m-1}}.$$

Por hipótesis, $x\varphi(\mathcal{A})\in\mathfrak{M}$ y, al mismo tiempo,

$$x_1\varphi(\mathcal{A})=x_2\varphi(\mathcal{A})=\ldots=x_{m-1}\varphi(\mathcal{A})=o.$$

Por consiguiente, $x_m\varphi(\mathcal{A})\in\mathfrak{M}$. Los polinomios $\varphi(\lambda)$ y $(\lambda-\rho_m)^{h_m}$ son primos entre sí. Luego, existen unos polinomios $F(\lambda)$ y $G(\lambda)$ tales que

$$1=\varphi(\lambda)F(\lambda)+(\lambda-\rho_m)^{h_m}G(\lambda),$$

de donde

$$\mathcal{E}=\varphi(\mathcal{A})F(\mathcal{A})+(\mathcal{A}-\rho_m\mathcal{E})^{h_m}G(\mathcal{A})$$

y, por consiguiente,

$$x_m=x_m\varphi(\mathcal{A})F(\mathcal{A})+x_m(\mathcal{A}-\rho_m\mathcal{E})^{h_m}G(\mathcal{A})=x_m\varphi(\mathcal{A})F(\mathcal{A})\in\mathfrak{M}$$

que es lo que se quería demostrar.

La afirmación expuesta anteriormente acerca de la independencia lineal de los vectores $x_1, \ldots, x_m$ se obtiene del teorema demostrado tomando $\mathfrak{M}=o$. Como corolario notemos también que *diferentes subespacios radicales tienen intersección nula.*

Volvamos ahora al caso en el que la matriz de una aplicación $\mathcal{A}$ tiene en una base la forma normal de Jordan (5). Hemos definido más arriba dos series de subespacios: los subespacios radicales $\mathfrak{L}_{\rho_1}, \ldots, \mathfrak{L}_{\rho_m}$ y los subespacios $\mathfrak{L}^{(1)}, \ldots, \mathfrak{L}^{(s)}$ correspondientes a las células diagonales de la matriz A. Para explicar la relación que existe entre ambas series indiquemos por $\mathfrak{M}^{(1)}$ la suma de aquellos subespacios $\mathfrak{L}^{(i)}$ que corresponden a las células con el valor propio ρ_1 y definamos análogamente $\mathfrak{M}^{(2)}, \ldots, \mathfrak{M}^{(m)}$. En virtud de ello, uniendo en la descomposición

$$\mathfrak{L}=\mathfrak{L}^{(1)}+\mathfrak{L}^{(2)}+\ldots+\mathfrak{L}^{(s)}$$

determinados sumandos obtenemos la descomposición

$$\mathfrak{L}=\mathfrak{M}^{(1)}+\mathfrak{M}^{(2)}+\ldots\dot{+}\mathfrak{M}^{(m)}. \tag{7}$$

Está claro que $\mathfrak{M}^{(i)}\subseteq\mathfrak{L}_{\rho_i}$ $(i=1, \ldots, m)$. Por esto, además de (7), tiene lugar también la descomposición

$$\mathfrak{L}=\mathfrak{L}_{\rho_1}+\mathfrak{L}_{\rho_2}+\ldots+\mathfrak{L}_{\rho_m}. \tag{8}$$

Puesto que, en virtud del teorema 3, la suma (8) es directa, obtenemos, comparándola con (7), las igualdades requeridas $\mathfrak{M}^{(i)}=\mathfrak{L}_{\rho_i}$.

Es decir, *si la matriz de una aplicación $\mathcal{A}$ puede ser reducida a la forma normal de Jordan, el espacio $\mathfrak{L}$ es la suma directa de los subespacios radicales de $\mathcal{A}$ y, además, cada uno de los subespacios radicales es, a su vez, la suma directa de los subespacios correspondientes a las células de Jordan con el valor propio dado.*

Los razonamientos expuestos permiten ver también que los subespacios radicales se determinan unívocamente por la propia aplicación $\mathcal{A}$ y no dependen de la selección de la base de coordenadas $e_1, \ldots, e_n$. En cuanto a los subespacios $\mathfrak{L}^{(i)}$, ellos dependen, en general, tanto de $\mathcal{A}$ como de la forma de reducción de la matriz a la forma diagonal.

Ejemplos y problemas

1. Supongamos que la aplicación $\mathcal{A}$ tiene en la base $e_1, \ldots, e_6$ la matriz $A=A_1\dot{+}A_2\dot{+}A_3$, donde $A_1=A_2=\begin{bmatrix}6 & 1\\ 0 & 6\end{bmatrix}$ y $A_3=\begin{bmatrix}7 & 1\\ 0 & 7\end{bmatrix}$. Los subespacios radicales de $\mathcal{A}$ son $\mathfrak{L}_6=Ke_1+Ke_2+Ke_3+Ke_4$ y $\mathfrak{L}_7=Ke_5+Ke_6$ (K es el cuerpo principal), mientras que $\mathfrak{L}^{(1)}=Ke_1+Ke_2$ y $\mathfrak{L}^{(2)}=Ke_3+Ke_4$. En la base nueva

$$e_1'=e_1+e_3,\ e_2'=e_2+e_4,\ e_3'=e_1-e_3,\ e_4'=e_2-e_4,\ e_5'=e_5 \text{ y } e_6'=e_6$$

la matriz de $\mathcal{A}$ será la misma; sin embargo, los subespacios $Ke_1'+Ke_2'$ y $Ke_3'+Ke_4'$ correspondientes a la células de Jordan serán distintos.

2. Hállense los polinomios mínimos de las matrices

$$\begin{bmatrix}2 & 1 & 1\\ & 2 & 1\\ & & 2\end{bmatrix},\quad \begin{bmatrix}2 & 1 & 1\\ & 3 & 1\\ & & 4\end{bmatrix},\quad \begin{bmatrix}2 & 1 & 1\\ & 2 & 1\\ & & 3\end{bmatrix} \quad \text{y} \quad \begin{bmatrix}2 & 1 & 1\\ & 2 & 0\\ & & 2\end{bmatrix}.$$

3. Una matriz sobre el cuerpo de los números complejos puede ser reducida a la forma diagonal si, y sólo si, su polinomio mínimo no tiene raíces múltiples.

4. Si la matriz de una aplicación puede ser reducida a la forma normal de Jordan, todo subespacio invariante suyo es suma directa de sus intersecciones con todos los subespacios radicales de la aplicación.

5. Si una matriz de orden n tiene n *diferentes* números característicos, la aplicación lineal correspondiente tiene un total de 2^n subespacios invariantes incluyendo el subespacio nulo y todo el espacio.

Capítulo IV

Matrices polinomiales

Los elementos de casi todas las matrices que hemos estudiado hasta el momento eran números de un cuerpo principal K. Sin embargo, al introducir el concepto de polinomio característico nos hemos visto obligados a considerar la matriz característica $\lambda E - A$, cuyos elementos no son números de K, sino *polinomios* en λ con coeficientes de K, suponiendo, además, que K es un *cuerpo conmutativo*. En el capítulo presente nos ocuparemos del estudio sistemático de las propiedades de las matrices polinomiales, es decir, de las matrices, cuyos elementos son polinomios en λ con coeficientes de un cuerpo conmutativo principal. Aplicaremos después estos resultados al problema consistente en hallar la forma de Jordan de la matriz de una aplicación lineal.

§ 13. Factores invariantes

13.1. Equivalencia. Consideremos una matriz cuadrada de orden n

$$\begin{bmatrix} f_{11}(\lambda) & \dots & f_{1n}(\lambda) \\ \vdots & & \vdots \\ f_{n1}(\lambda) & \dots & f_{nn}(\lambda) \end{bmatrix},$$

cuyos elementos son polinomios en la letra λ con coeficientes de un campo principal K. Llamaremos las matrices de este tipo *polinomiales* o *λ-matrices*[1]. Frecuentemente resulta necesario realizar con λ-matrices las transformaciones siguientes:

I. *La multiplicación de una de las filas por un número de K diferente de cero.*

II. *La adición a una de las filas de la matriz de otra fila multiplicada por un polinomio arbitrario $f(\lambda)$.*

[1] Suponemos en todo momento que las matrices consideradas son cuadradas, aunque muchos de los resultados pueden ser extendidos directamente al caso de λ-matrices rectangulares.

III. *La multiplicación de una de las columnas por un número de K diferente de cero.*

IV. *La adición a una de las columnas de los elementos de otra columna multiplicados por un polinomio arbitrario $f(\lambda)$.*

Estas transformaciones se llaman *transformaciones elementales de λ-matrices.* Si con una λ-matriz se realiza una transformación elemental se obtiene de nuevo una λ-matriz; con esta matriz se puede realizar otra transformación elemental, etc. Se dice que una λ-matriz F es *equivalente* a una λ-matriz G, si F se puede obtener de G mediante una cadena de transformaciones elementales. En muchas ocasiones resulta útil el lema siguiente:

Mediante las transformaciones elementales I *y* II *se pueden cambiar entre sí dos filas cualesquiera de una λ-matriz y mediante las transformaciones elementales* III *y* IV *se pueden cambiar entre sí dos columnas cualesquiera de la misma.*

En efecto, supongamos que es necesario cambiar entre sí la i-ésima y la j-ésima columnas de una matriz. Es fácil ver que esto se consigue realizando las siguientes transformaciones elementales: 1) agregamos a la i-ésima fila la j-ésima; 2) a la j-ésima fila de la matriz nueva agregamos su i-ésima fila multiplicada por -1; 3) multiplicamos la j-ésima fila de la matriz obtenida por -1 y 4) agregamos a la i-ésima fila de la última matriz su j-ésima fila multiplicada por -1. Si realizamos transformaciones análogas con las columnas, lograremos cambiar de posición la i-ésima y la j-ésima columnas. Hemos demostrado el lema.

De este lema se deduce que *si la matriz F difiere de G en el orden de las columnas o de las filas, la matriz F es equivalente a G.*

De la definición de equivalencia de λ-matrices se desprenden directamente las propiedades siguientes:

1. *La relación de equivalencia es transitiva:* si F es equivalente a G y G es equivalente a H, resulta que F es equivalente a H.

En efecto, G se puede obtener, por hipótesis, de H y F de G mediante una cadena de transformaciones elementales; por consiguiente, F se puede obtener mediante una cadena de transformaciones elementales de H.

2. *La relación de equivalencia es simétrica:* si F es equivalente a G, G es equivalente a F. En otras palabras, si F se puede obtener de G mediante una cadena de transformaciones elementales, también G se puede obtener de F mediante una cadena de transformaciones elementales.

Demostremos primero que si F se puede obtener de G mediante una transformación elemental, también G se puede obtener de F mediante una transformación elemental. Consideremos para ello, uno a uno, los cuatro tipos de las transformaciones elementales. Supongamos que F se obtiene de G mediante la transformación de tipo I, es decir, multiplicando una fila i-ésima de G por un número $\alpha \neq 0$.

Entonces, multiplicando la i-ésima fila de F por α^{-1}, obtendremos, evidentemente, G. Supongamos ahora que F se obtiene de G mediante una transformación de tipo II, por ejemplo, agregando a la i-ésima fila de la matriz G su j-ésima fila multiplicada por $f(\lambda)$. En este caso, agregando a la i-ésima fila de la matriz F su j-ésima fila multiplicada por $-f(\lambda)$, obtendremos de nuevo G. Lo mismo se puede decir acerca de las transformaciones de tipo III y IV. Vemos, por consiguiente, que para toda transformación elemental existe la transformación elemental inversa que anula el resultado de la primera. Por esto, si la matriz F se obtiene de G mediante una cadena de transformaciones elementales, resulta que realizando las transformaciones inversas en el orden contrario podremos obtener de la matriz F la matriz G que es lo que se quería demostrar.

3. *La relación de equivalencia es reflexiva:* toda matriz es equivalente a sí misma.

Por ejemplo, realizando con F dos transformaciones recíprocamente inversas, obtendremos de nuevo F.

13.2. Forma diagonal. Acabamos de demostrar que la relación de equivalencia es transitiva, simétrica y reflexiva. De aquí se deduce que las λ-matrices se descomponen en clases de matrices equivalentes. Surge el problema: ¿puede indicarse una forma para las λ-matrices tal que en cada una de estas clases haya una matriz de la forma dada, y sólo una? Las formas con esta propiedad se denominan *canónicas*. Demostraremos que para las λ-matrices la forma diagonal, con algunas condiciones complementarias de divisibilidad, es canónica en este sentido.

DEFINICIÓN. *Una λ-matriz de tipo*

$$\begin{bmatrix} f_1(\lambda) & & & \\ & f_2(\lambda) & & \\ & & \ddots & \\ & & & f_n(\lambda) \end{bmatrix}$$

se llama canónica diagonal, si todo elemento diagonal $f_i(\lambda)$ divide al elemento siguiente $f_{i+1}(\lambda)$ y si el coeficiente principal de todos los polinomios $f_1(\lambda), \ldots, f_n(\lambda)$ diferentes de cero es 1.

De aquí se deduce que si entre los elementos diagonales de una matriz canónica diagonal aparecen ceros, estos elementos deben ocupar las posiciones últimas, ya que el cero no puede dividir a ningún polinomio no nulo. Por otro lado, si entre los elementos diagonales figuran *números* diferentes de cero, éstos deben ser iguales a 1 y deben ocupar las posiciones primeras, ya que 1 no es divisible por ningún polinomio con coeficiente principal 1, a excepción del polinomio 1. Por consiguiente, en el caso más general la matriz canónica diagonal es de la forma

$$\begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & f_1(\lambda) & & & & & \\ & & & & \ddots & & & & \\ & & & & & f_k(\lambda) & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix}$$

donde $f_1(\lambda), \ldots, f_k(\lambda)$ no son constantes, sus coeficientes principales son iguales a 1 y cada uno de ellos es divisor del siguiente.

TEOREMA 1. *Toda λ-matriz puede ser reducida mediante un número finito de tranformaciones elementales a la forma canónica diagonal.*

DEMOSTRACIÓN. Sea G una λ-matriz dada. Si $G=O$, nada hay que demostrar, puesto que O es ya de la forma normal. Supongamos por lo tanto que $G \neq O$. Escojamos entre todas las λ-matrices equivalentes a G aquella cuyo elemento del ángulo superior izquierdo sea diferente de cero y sea del grado mínimo. Sea

$$F = \begin{bmatrix} f_{11}(\lambda) & \ldots & f_{1n}(\lambda) \\ \vdots & & \vdots \\ f_{n1}(\lambda) & \ldots & f_{nn}(\lambda) \end{bmatrix}$$

esta matriz. Demostremos que $f_{11}(\lambda)$ divide a todos los elementos de la primera fila y a todos los elementos de la primera columna de la matriz F. En efecto, sea

$$f_{1i}(\lambda) = f_{11}(\lambda)\, q_i(\lambda) + r_i(\lambda) \quad (i = 1,\ 2,\ \ldots,\ n), \tag{1}$$

donde $q_i(\lambda)$ y $r_i(\lambda)$ son el cociente y el resto de la división de $f_{1i}(\lambda)$ por $f_{11}(\lambda)$. Realicemos con F la siguiente transformación elemental: restemos de los elementos de la i-ésima columna los elementos de la primera multiplicados por $q_i(\lambda)$. La igualdad (1) muestra que el elemento de la primera fila y de la i-ésima columna de la matriz nueva resultará igual a $r_i(\lambda)$. Si $r_i(\lambda) \neq 0$, el grado de $r_i(\lambda)$ es inferior al grado del divisor $f_{11}(\lambda)$. Cambiando entre sí la primera y la i-ésima columnas, obtendremos una matriz equivalente a F en cuyo ángulo superior izquierdo figura un polinomio $r_i(\lambda)$ de grado menor que el de $f_{11}(\lambda)$; pero esto contradice a la definición de la matriz F. Por consiguiente, $r_i(\lambda) = 0$.

Indicando ahora por $q_i(\lambda)$ el cociente de la división de $f_{1i}(\lambda)$ por $f_{11}(\lambda)$, realicemos con la matriz F las siguientes transformaciones elementales: de la segunda columna restemos la primera multiplicada por $q_2(\lambda)$, restemos después de la tercera columna la primera multiplicada por $q_3(\lambda)$, etc. Realicemos a continuación transformaciones análogas con las filas. La matriz F resultará sustituida por la matriz equivalente

$$H=\begin{bmatrix} f_{11}(\lambda) & 0 & \dots & 0 \\ 0 & h_{22}(\lambda) & \dots & h_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ 0 & h_{n2}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}, \tag{2}$$

donde $h_{ij}(\lambda)$ son unos polinomios determinados.

Todos los polinomios $h_{ij}(\lambda)$ son divisibles por $f_{11}(\lambda)$. Efectivamente, si $f_{11}(\lambda)$ no divide a uno de ellos, digamos a $h_{ij}(\lambda)$, entonces sumando a la primera fila de la matriz H su i-ésima fila obtendremos una matriz Q con las siguientes propiedades:

1) Q es equivalente a G,

2) el elemento superior de la izquierda de la matriz Q es diferente de cero y es del menor grado,

3) en la primera fila de la matriz Q figura el elemento $h_{ij}(\lambda)$ que no es divisible por el primer elemento de esta fila.

Sin embargo, hemos visto que de las dos propiedades primeras se desprende que todos los elementos de la primera fila son divisibles por su primer elemento. Por consiguiente, la tercera propiedad contradice a las dos primeras y nuestra proposición queda demostrada. Hemos probado, pues, que para toda λ-matriz G existe una matriz equivalente H de tipo (2), donde todos los $h_{ij}(\lambda)$ son divisibles por $f_{11}(\lambda)$. Realicemos ahora transformaciones elementales con la matriz

$$H_1=\begin{bmatrix} h_{22}(\lambda) & \dots & h_{2n}(\lambda) \\ \vdots & & \vdots \\ h_{n2}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}.$$

Toda transformación elemental de H_1 puede ser considerada también como una transformación elemental de la matriz H. Es fácil ver que la primera fila y la primera columna de la matriz H no varían en estas transformaciones. Además, puesto que todos los elementos de la matriz H_1 son divisibles por $f_{11}(\lambda)$, todos los elementos de las matrices nuevas, que surgen de H_1 como resultado de transformaciones elementales, también serán divisibles por $f_{11}(\lambda)$.

Aplicando a la matriz H_1 el resultado demostrado anteriormente, veremos que H_1 puede ser reducida mediante transformaciones elementales a la forma

$$\begin{bmatrix} h_{22}(\lambda) & 0 & \dots & 0 \\ 0 & h_{33}(\lambda) & \dots & h_{3n}(\lambda) \\ \vdots & \vdots & & \vdots \\ 0 & h_{n3}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix}$$

y, por consiguiente, la matriz H a la forma

$$\begin{bmatrix} f_{11}(\lambda) & & & & \\ & h_{22}(\lambda) & & & \\ & & h_{33}(\lambda) & \dots & h_{3n}(\lambda) \\ & & \vdots & & \vdots \\ & & h_{n3}(\lambda) & \dots & h_{nn}(\lambda) \end{bmatrix},$$

donde todos los $h_{ij}(\lambda)$ $(i, j=3, \dots, n)$ son divisibles por $h_{22}(\lambda)$ y $h_{22}(\lambda)$ es divisible por $f_{11}(\lambda)$. Continuando este proceso obtendremos al cabo de un número finito de pasos la forma canónica diagonal requerida.

De nuestra demostración se puede extraer fácilmente un método práctico de reducción de λ-matrices a la forma canónica diagonal. Su idea consiste en disminuir primero, empleando las transformaciones elementales, el grado del elemento que figura en la primera fila y en la primera columna y en convertir en cero los demás elementos de las mismas. Después de haber logrado este primer objetivo, aplicamos el mismo método al ángulo H_1 que queda, etc.

El teorema 1 afirma que *toda clase de matrices equivalentes contiene al menos una matriz de forma canónica diagonal*. En el punto siguiente quedará demostrado que esta matriz es única en cada clase.

13.3. Máximos comunes divisores de menores. Sea F una λ-matriz de orden n. Consideremos todos sus menores de orden k $(k=1, 2, \dots, n)$. Estos menores son unos polinomios en λ. Indiquemos por $D_k(\lambda)$ su máximo común divisor[1]. Si resulta que todos los menores de un orden k son iguales a cero, aceptaremos por definición que $D_k(\lambda)=0$. En particular, $D_1(\lambda)$ es el máximo común divisor de los elementos de la matriz F, mientras que $D_n(\lambda)$ es igual al determinante de F, dividido por su coeficiente principal.

TEOREMA 2 *Las λ-matrices equivalentes tienen un mismo máximo común divisor de los menores de orden k $(k=1, 2, \dots, n)$.*

Sean F_1 y F_2 dos λ-matrices equivalentes. Indiquemos los máximos comunes divisores de sus menores de orden k por $D_{k1}(\lambda)$ y $D_{k2}(\lambda)$, respectivamente. Debemos probar que $D_{k1}(\lambda)=D_{k2}(\lambda)$. Sabemos que F_2 puede ser obtenida de F_1 mediante una cadena de transformaciones elementales. Supongamos primero que esta cadena consta sólo de una transformación elemental. Por ejemplo, supongamos que F_2 se obtiene de F_1 multiplicando todos los elementos de la i-ésima fila de la matriz F_1 por un número $\alpha \neq 0$. En este caso los menores respectivos de F_1 y de F_2 o bien son idénticos o bien difieren en el

[1] Convendremos en que el máximo común divisor es siempre el común divisor del grado mayor de *coeficiente principal* igual a 1. Por ello, todos los polinomios $D_k(\lambda)$ diferentes de cero son de coeficiente principal igual a 1.

factor constante α. Pero un factor constante no influye en el cálculo del máximo común divisor, de modo que $D_{k1}=D_{k2}$. Lo mismo sucederá en el caso en que F_2 se obtenga de F_1 multiplicando por α los elementos de una de las columnas de la matriz F_1. Supongamos ahora que F_2 se obtiene de F_1 mediante las transformaciones de tipo II o IV; por ejemplo, aceptemos que F_2 se obtiene al agregar a la i-ésima fila de la matriz F_1 los elementos de la j-ésima fila multiplicados por $f(\lambda)$. Demostremos que D_{k2} es divisible por D_{k1}.

En efecto, existen tres clases de menores de orden k de las matrices F_1 y F_2. A la primera pertenecen aquellos que no contienen elementos de la i-ésima fila. En este caso los menores respectivos de las matrices F_1 y F_2 son, obviamente, iguales. A la segunda clase pertenecen aquellos menores que contienen elementos de la i-ésima y de la j-ésima filas. Estos menores de las matrices F_1 y F_2 también serán iguales, ya que el valor de un determinante no se altera si a los elementos de una de sus filas se suman cantidades proporcionales a los elementos de otra de sus filas. Finalmente, a la tercera clase pertenecen los menores que contienen elementos de la i-ésima fila y no contienen elementos de la j-ésima fila. Los menores respectivos de esta clase son de forma

$$M_1=\begin{vmatrix} \cdot & \cdot & \cdot \\ f_{i\nu_1} & \dots & f_{i\nu_k} \\ \cdot & \cdot & \cdot \end{vmatrix}, \qquad M_2=\begin{vmatrix} \cdot & \cdot & \cdot \\ f_{i\nu_1}+f_{j\nu_1}\cdot f(\lambda) & \dots & f_{i\nu_k}+f_{j\nu_k}\cdot f(\lambda) \\ \cdot & \cdot & \cdot \end{vmatrix},$$

donde las filas de ambos menores que no han sido escritas coinciden. En virtud del teorema de adición de determinantes, tenemos

$$M_2=\begin{vmatrix} \cdot & \cdot & \cdot \\ f_{i\nu_1} & \dots & f_{i\nu_k} \\ \cdot & \cdot & \cdot \end{vmatrix}+f(\lambda)\begin{vmatrix} \cdot & \cdot & \cdot \\ f_{j\nu_1} & \dots & f_{j\nu_k} \\ \cdot & \cdot & \cdot \end{vmatrix}=M_1+f(\lambda)N_1,$$

donde N_1 es un menor de la matriz F_1. Todos los menores de orden k de la matriz F_1 son divisibles por D_{k1}. De nuestros razonamientos se ve que D_{k1} divide a todos los menores de orden k de la matriz F_2, ya que éstos o bien coinciden con los menores respectivos de la matriz F_1 o bien se expresan linealmente en términos de los mismos. Pero en tal caso, D_{k1} aparecerá como factor en el máximo común divisor de los menores de orden k de la matriz F_2, es decir, será un factor en D_{k2}. Por lo tanto, si F_2 se obtiene de F_1 mediante una sola transformación elemental, resulta que D_{k2} es divisible por D_{k1}. Realizando con F_2 la transformación elemental inversa, obtendremos F_1. Por ello, D_{k1} debe ser también divisible por D_{k2}. Pero, si los coeficientes principales de dos polinomios coinciden y si estos polinomios son divisibles uno por otro, ellos deben coincidir. Es decir, $D_{k1}=D_{k2}$. Por ahora hemos demostrado la igualdad de los máximos comunes divisores suponiendo que F_2 se obtiene de F_1

mediante una sola transformación elemental. Sin embargo, si $D_k(\lambda)$ no varía en cualquier transformación elemental concreta, es obvio que $D_k(\lambda)$ tampoco varía en el caso de varias transformaciones sucesivas. Por lo tanto, podemos considerar que el teorema 2 ha quedado demostrado en su forma general.

Calculemos los polinomios $D_1(\lambda), \ldots, D_n(\lambda)$ de una matriz de forma canónica diagonal

$$D = \begin{bmatrix} d_1(\lambda) & & & & \\ & d_2(\lambda) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & d_n(\lambda) \end{bmatrix}.$$

Para obtener un menor de orden k debemos suprimir $n-k$ filas y $n-k$ columnas de D. Si suprimimos en D la i-ésima fila, en su i-ésima columna quedarán sólo ceros. Por ello, para obtener un menor diferente de cero, debemos suprimir en D todas las columnas cuyos números coinciden con los de las filas suprimidas. Es decir, los menores de orden k diferentes de cero deben ser de la forma

$$\begin{vmatrix} d_{\nu_1}(\lambda) & & & & \\ & d_{\nu_2}(\lambda) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & d_{\nu_k}(\lambda) \end{vmatrix} = d_{\nu_1}(\lambda)\, d_{\nu_2}(\lambda) \ldots d_{\nu_k}(\lambda). \tag{3}$$

El máximo común divisor de estos menores es $D_k(\lambda)$. De las desigualdades $1 \leqslant \nu_1 \leqslant \nu_2 \leqslant \ldots \leqslant \nu_k \leqslant n$ se deduce que $1 \leqslant \nu_1$, $2 \leqslant \nu_2$, $\ldots$, $k \leqslant \nu_k$. Por lo tanto $d_{\nu_i}(\lambda)$ es divisible por $d_i(\lambda)$, de modo que $d_{\nu_1}(\lambda) \ldots d_{\nu_k}(\lambda)$ es divisible por $d_1(\lambda) \ldots d_k(\lambda)$. Vemos, por consiguiente, que todos los menores de orden k de la matriz D son divisibles por el menor

$$\begin{vmatrix} d_1(\lambda) & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & d_k(\lambda) \end{vmatrix} = d_1(\lambda) \ldots d_k(\lambda). \tag{4}$$

Si este menor es igual a cero, todos los menores de orden k de la matriz D también son iguales a cero. Por definición, tenemos en este caso $D_k(\lambda) = 0$. Si el menor (4) es diferente de cero, los polinomios $d_1(\lambda), \ldots, d_k(\lambda)$ son diferentes de cero y sus coeficientes principales son iguales a 1. Pero entonces también el coeficiente principal del menor (4) es igual a 1. Puesto que (4) divide a todos los menores

(3), resulta que $D_k(\lambda)$ coincide con (4). En ambos casos tenemos, por consiguiente,

$$D_k(\lambda) = d_1(\lambda) d_2(\lambda) \ldots d_k(\lambda) \quad (k = 1, 2, \ldots, n). \tag{5}$$

Así se calculan los polinomios $D_k(\lambda)$ de una matriz canónica diagonal de elementos diagonales $d_1(\lambda), \ldots, d_n(\lambda)$.

Consideremos ahora una λ-matriz arbitraria F. Indiquemos por $D_k(\lambda)$ el máximo común divisor de los menores de orden k de esta matriz. Según el teorema 1, la matriz F puede ser reducida mediante transformaciones elementales a la forma canónica diagonal

$$D = \begin{bmatrix} d_1(\lambda) & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & d_n(\lambda) \end{bmatrix}.$$

De acuerdo con el teorema 2 los polinomios $D_k(\lambda)$ calculados para la matriz D coinciden con los respectivos polinomios $D_k(\lambda)$ calculados para F. Por consiguiente, *los polinomios $D_k(\lambda)$ de la matriz F y los elementos diagonales de la matriz canónica diagonal D, a la que se puede reducir F, están ligados por las relaciones* (5).

Supongamos que $D_1(\lambda), \ldots, D_r(\lambda)$ son diferentes de cero y que los demás polinomios $D_{r+1}(\lambda), \ldots, D_n(\lambda)$, si es que existen, son iguales a cero. Entonces de (5) obtenemos

$$\begin{aligned} D_1(\lambda) &= d_1(\lambda), & d_1(\lambda) &= D_1(\lambda), \\ D_2(\lambda) &= d_1(\lambda) d_2(\lambda), & d_2(\lambda) &= D_2(\lambda) : D_1(\lambda), \\ &\cdots\cdots\cdots & &\cdots\cdots\cdots \\ D_r(\lambda) &= d_1(\lambda) d_2(\lambda) \ldots d_r(\lambda), & d_r(\lambda) &= D_r(\lambda) : D_{r-1}(\lambda), \\ D_{r+1}(\lambda) &= d_1(\lambda) d_2(\lambda) \ldots d_r(\lambda) d_{r+1}(\lambda), & d_{r+1}(\lambda) &= D_{r+1}(\lambda) : D_r(\lambda). \end{aligned}$$

Puesto que $d_{r+1}(\lambda) = 0$, resulta que $d_{r+2}(\lambda), \ldots, d_n(\lambda)$ también deben ser iguales a cero y obtenemos definitivamente

$$d_1(\lambda) = D_1(\lambda),\ d_2(\lambda) = D_2(\lambda) : D_1(\lambda),\ \ldots,\ d_r(\lambda) = D_r(\lambda) : D_{r-1}(\lambda),$$
$$d_{r+1}(\lambda) = \ldots = d_n(\lambda) = 0. \tag{6}$$

De esta forma hemos obtenido el teorema siguiente.

TEOREMA 3. *Si los máximos comunes divisores $D_k(\lambda)$ de los menores de orden k de una λ-matriz F son diferentes de cero para $k = 1$, 2, ..., r y si $D_{r+1}(\lambda) = 0$, los elementos diagonales $d_k(\lambda)$ de la matriz canónica diagonal, a la que se reduce F mediante transformaciones elementales, se expresan en términos de $D_k(\lambda)$ por las fórmulas* (6) *y, por consiguiente, la matriz F los define unívocamente.*

Los polinomios $d_1(\lambda), \ldots, d_n(\lambda)$ se llaman *factores invariantes de la matriz F*.

El número r que figura en las igualdades (6) tiene un sentido muy simple: es el rango de la matriz F. Efectivamente, el rango de la matriz F es el orden máximo de los menores de F diferentes de cero. Si este orden es igual a r, tenemos en consecuencia que $D_r(\lambda)\neq 0$ y que $D_{r+1}(\lambda)=0$. Recíprocamente, las condiciones $D_r(\lambda)\neq 0$ y $D_{r+1}(\lambda)=0$ significan que uno de los menores de orden r es diferente de cero y que todos los menores de orden $r+1$ son iguales a cero. Por consiguiente, el rango de F es igual a r.

13.4. Condiciones de equivalencia. Empleando los resultados del punto anterior es fácil encontrar las condiciones que garanticen que dos λ-matrices dadas sean equivalentes. Representaremos estas condiciones en dos formas.

PRIMERA CONDICIÓN DE EQUIVALENCIA. *Para que unas matrices polinomiales de orden n sean equivalentes es necesario y suficiente que los máximos comunes divisores de sus menores de orden k coincidan para $k=1, 2, \ldots, n$.*

Puesto que la coincidencia de los máximos comunes divisores de los menores equivale a la coincidencia de los factores invariantes respectivos, la primera condición de equivalencia puede ser enunciada del modo siguiente: *para la equivalencia de λ-matrices es necesario y suficiente que sus factores invariantes correspondientes sean iguales.*

La demostración es evidente. En efecto, si dos λ-matrices F y G son equivalentes, sus máximos comunes divisores $D_k(\lambda)$ son idénticos (teorema 2). Viceversa, si los polinomios $D_k(\lambda)$ de F y de G son iguales, las matrices F y G se reducen mediante transformaciones elementales a una misma matriz canónica diagonal (teorema 3). Pero dos matrices equivalentes a una tercera son equivalentes; por consiguiente F es equivalente a G que es lo que se quería demostrar.

SEGUNDA CONDICIÓN DE EQUIVALENCIA. *Para que unas matrices polinomiales F y G de orden n sean equivalentes es necesario y suficiente que satisfagan la relación*

$$G=PFQ,$$

donde P y Q son unas matrices polinomiales de determinantes constantes diferentes de cero[1].

---

[1] Hemos definido la equivalencia de λ-matrices mediante las transformaciones elementales. Con frecuencia el concepto de equivalencia se define también de otra manera. Se dice que dos λ-matrices G y F son equivalentes, si existen unas matrices cuadradas regulares P y Q de determinantes constantes que satisfacen la relación $G=PFQ$. Entonces, la segunda condición de equivalencia puede ser interpretada como el teorema de equivalencia de estos conceptos.

Antes de pasar a la demostración de esta proposición hagamos algunas observaciones. Sea

$$A = \begin{array}{cl} \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \alpha & & & & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \\ i\text{-ésima fila,} \\ \\ \\ \\ \end{array} \end{array}$$

donde α es un número diferente de cero. Multiplicando a la izquierda por A una matriz cualquiera F, veremos que todos los elementos de la matriz F permanecerán intactos, menos los elementos de la i-ésima fila que resultarán multiplicados por α. Por consiguiente, *toda transformación elemental de tipo* I *realizada con la matriz F equivale a la multiplicación de F a la izquierda por una matriz A convenientemente escogida.* Análogamente, si multiplicamos la matriz F a la izquierda por la matriz

$$B = \begin{array}{cl} \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & f(\lambda) & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} & \begin{array}{l} \\ \\ i\text{-ésima fila} \\ \\ j\text{-ésima fila,} \\ \\ \\ \end{array} \end{array}$$

donde todos los elementos diagonales son iguales a la unidad, el elemento de la i-ésima fila y de la j-ésima columna es igual a $f(\lambda)$ y los restantes elementos son iguales a cero, resultará que a los elementos de la i-ésima fila de F se sumarán los elementos de su j-ésima fila multiplicados por $f(\lambda)$. Por consiguiente, *toda transformación elemental de tipo* II *equivale a la multiplicación a la izquierda de la matriz F por una matriz B convenientemente escogida.*

Finalmente, es fácil comprobar de esta misma forma que *las transformaciones elementales de la matriz F de tipo* III *y* IV *equivalen a la multiplicación de F a la derecha por las matrices respectivas A y B.*

Pasemos ahora a demostrar la segunda condición de equivalencia.

NECESIDAD. Sea G una matriz equivalente a la matriz F. Esto significa que G se puede obtener de F mediante una cadena de transformaciones elementales sucesivas. Podemos sustituir cada una de las transformaciones elementales por la multiplicación por una matriz de tipo A o B a la izquierda o a la derecha, respectivamente. Como resultado obtendremos la igualadad

$$G = P_1 P_2 \ldots P_p F Q_1 Q_2 \ldots Q_q, \tag{7}$$

donde cada una de las matrices P_i y Q_j es de tipo A o B (i, $j = 1$, 2, ...). Pongamos

$$P = P_1 P_2 \ldots P_p \text{ y } Q = Q_1 Q_2 \ldots Q_q.$$

Puesto que los determinantes de las matrices B son iguales a la unidad y los determinantes de las matrices A son números constantes diferentes de cero, resulta que los determinantes de las matrices P y Q también serán unos números constantes diferentes de cero. La relación (7) muestra que

$$G = PFQ$$

y la necesidad queda demostrada.

SUFICIENCIA. Supongamos al contrario que

$$G = PFQ, \tag{8}$$

donde P y Q son matrices polinomiales de determinantes constantes diferentes de cero. El máximo común divisor $D_n(\lambda)$ de todos los menores de orden n de la matriz P es igual al determinante de P, dividido por su coeficiente principal. Puesto que este determinante es un número constante, resulta que $D_n(\lambda) = 1$. Para $k = n$ tenemos de las relaciones (5)

$$D_n(\lambda) = d_1(\lambda) d_2(\lambda) \ldots d_n(\lambda) = 1,$$

y de aquí resulta

$$d_1(\lambda) = d_2(\lambda) = \ldots = d_n(\lambda) = 1,$$

donde $d_1(\lambda)$, ..., $d_n(\lambda)$ son los factores invariantes de la matriz P. Pero los factores invariantes de la matriz unidad E también son todos iguales a la unidad, ya que E es de forma canónica diagonal. Según el primer criterio de equivalencia, de aquí se deduce que la matriz P es equivalente a E y, por consiguiente, puede ser obtenida de E mediante una cadena de transformaciones elementales. Toda transformación elemental puede ser sustituida por la multiplicación por una matriz de tipo A o B. Así P quedará representada en la forma siguiente

$$P = P_1 \ldots P_k E Q_1 \ldots Q_l = P_1 \ldots P_k Q_1 \ldots Q_l,$$

donde P_i y Q_j son matrices de tipo A y B.

Aplicando estos mismos razonamientos a la matriz Q, obtenemos un desarrollo análogo

$$Q = M_1 \ \dots \ M_s N_1 \ \dots \ N_t.$$

Introduciendo en (8) estas descomposiciones, llegamos a la igualdad

$$G = P_1 \ \dots \ P_k Q_1 \ \dots \ Q_l F M_1 \ \dots \ M_s N_1 \ \dots \ N_t \tag{9}$$

de la cual se ve que G se obtiene multiplicando sucesivamente la matriz F por las matrices P_i, Q_i, M_i y N_i de tipo A o B. Pero cada multiplicación de esta índole equivale a una transformación elemental. Por consiguiente, G es equivalente a F y hemos concluido la demostración.

Ejemplos y problemas

1. Redúzcanse a la forma canónica diagonal mediante transformaciones elementales las matrices

$$\begin{bmatrix} \lambda-2 & -1 & 0 \\ 0 & \lambda-2 & -1 \\ 0 & 0 & \lambda-2 \end{bmatrix}, \quad \begin{bmatrix} \lambda(\lambda+1) & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & (\lambda+1)^2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}.$$

2. Empleando los máximos comunes divisores de los menores, hállese la forma canónica diagonal de las λ-matrices

$$\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 5 & 4 & 3 & \lambda+2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \alpha+\lambda & \beta & 1 & 0 \\ -\beta & \alpha+\lambda & 0 & 1 \\ 0 & 0 & \alpha+\lambda & \beta \\ 0 & 0 & -\beta & \alpha+\lambda \end{bmatrix}.$$

3. Demuéstrese que toda λ-matriz rectangular de m filas y de n columnas puede ser reducida mediante transformaciones elementales a la forma

$$\begin{bmatrix} d_1\lambda & 0 & \dots & 0 & \dots & 0 \\ 0 & d_2(\lambda) & \dots & 0 & \dots & 0 \\ . & . & . & . & . & . \\ 0 & 0 & \dots & d_m(\lambda) & \dots & 0 \end{bmatrix} \quad \text{o a la forma} \quad \begin{bmatrix} d_1(\lambda) & \dots & 0 \\ . & . & . \\ 0 & \dots & d_n(\lambda) \\ 0 & \dots & 0 \\ . & . & . \\ 0 & \dots & 0 \end{bmatrix}.$$

¿Cómo debe enunciarse la segunda condición de equivalencia para estas matrices?

4. Demuéstrese que mediante las transformaciones elementales de tipo I y II toda λ-matriz cuadrada puede ser reducida a la forma

$$\begin{bmatrix} f_{11}(\lambda) & f_{12}(\lambda) & \dots & f_{1n}(\lambda) \\ & f_{22}(\lambda) & \dots & f_{2n}(\lambda) \\ & . & . & . \\ & & & f_{nn}(\lambda) \end{bmatrix},$$

donde los elementos diagonales o bien son iguales a cero o bien son de coeficiente principal igual a 1.

5. Consideremos en lugar de λ-matrices las matrices formadas por números enteros. Las transformaciones elementales de estas matrices se definen del modo siguiente: I, multiplicación de una fila por ± 1; II, adición a los elementos de una fila de los elementos de otra cualquiera multiplicados por un número entero; III y IV, transformaciones similares de las columnas. Una matriz diagonal

formada por números enteros se llama canónica, si sus elementos diagonales son no negativos y si todo elemento diagonal anterior divide al que le sigue. Demuéstrense los teoremas: a) toda matriz formada por números enteros puede ser reducida, mediante un número finito de transformaciones elementales, a la forma canónica diagonal; b) esta forma canónica diagonal es única y c) para la equivalencia de dos matrices F y G formadas por números enteros es necesario y suficiente que para ellas sea válida la relación $G=PFQ$, donde P y Q son unas matrices formadas por números enteros, cuyos determinantes son iguales a ± 1.

§ 14. Divisores elementales

Los factores invariantes determinan, salvo una equivalencia, la matriz polinomial F. Sin embargo, si F se descompone en células diagonales, la relación entre los factores invariantes de la matriz F y los factores invariantes de las células resulta compleja. Por ello, conviene considerar en algunas cuestiones, en lugar de los factores invariantes, los así llamados divisores elementales de la matriz F, ya que el comportamiento de estos últimos en el caso de descomposición de F es más sencillo.

14.1. Relación con los factores invariantes. Consideremos una λ-matriz arbitraria F, cuyos elementos son polinomios en λ con coeficientes del campo principal K. No someteremos el cuerpo conmutativo K a ninguna restricción. Sean $d_1(\lambda)$, $d_2(\lambda)$, ..., $d_n(\lambda)$ los factores invariantes de la matriz F. Parte de estos factores puede ser igual a cero; por ello, supondremos, para concretar, que $d_1(\lambda)$, ..., $d_r(\lambda)$ son diferentes de cero y que $d_{r+1}(\lambda)=\ldots=d_n(\lambda)=0$. El número r, como hemos visto anteriormente, es el rango de la matriz F. Descompongamos cada uno de los polinomios $d_1(\lambda)$, ..., $d_r(\lambda)$ en factores irreducibles sobre el cuerpo conmutativo K. Sea, por ejemplo,

$$d_i(\lambda)=[\varepsilon_1(\lambda)]^{n_1}[\varepsilon_2(\lambda)]^{n_2}\ \ldots\ [\varepsilon_k(\lambda)]^{n_k},$$

donde $\varepsilon_1(\lambda)$, ..., $\varepsilon_k(\lambda)$ son distintos polinomios irreducibles de coeficiente principal igual a 1. Las expresiones $[\varepsilon_1(\lambda)]^{n_1}$, ..., $[\varepsilon_k(\lambda)]^{n_k}$ se llaman *divisores elementales* del factor invariante $d_i(\lambda)$. Los divisores elementales de todos los factores invariantes de la matriz F que no sean constantes se llaman *divisores elementales* de esta matriz. Por ejemplo, si los factores invariantes de la matriz F son iguales, respectivamente a 1, λ, $\lambda^2(\lambda+1)$ y $\lambda^2(\lambda+1)^2$, sus divisores elementales serán λ, λ^2, λ^2, $\lambda+1$, $(\lambda+1)^2$. Un divisor elemental de tipo $[\varepsilon(\lambda)]^k$, donde $\varepsilon(\lambda)$ es un polinomio irreducible, se llama *perteneciente* al polinomio $\varepsilon(\lambda)$. En el ejemplo considerado los divisores elementales λ, λ^2 y λ^2 pertenecen a λ, mientras que $\lambda+1$ y $(\lambda+1)^2$ pertenecen a $\lambda+1$.

Tomemos ahora una λ-matriz F y escribamos todos sus divisores elementales. Si uno de los divisores elementales de F figura en varios factores invariantes de F, lo escribiremos tantas veces cuantas aparece en los factores invariantes. Demostremos que el sistema de divisores elementales obtenido de esta forma determina plenamente todos los factores invariantes de la matriz F diferentes de una constante; si tomamos en consideración, además, el orden y el rango de F, quedarán determinados todos los factores invariantes de la matriz F.

TEOREMA 1. *El orden, el rango y el sistema de divisores elementales de una λ-matriz F determinan plenamente sus factores invariantes y, por consiguiente, determinan F salvo una equivalencia.*

La demostración se aclara fácilmente con el ejemplo siguiente. Supongamos que F es de orden 6 y de rango 4 y que sus divisores elementales son λ, λ^2, λ^2, $\lambda+1$, $(\lambda+1)^3$, $\lambda-1$ y $\lambda-1$. Puesto que el orden de F es 6, la matriz F tiene seis factores invariantes $d_1(\lambda), \ldots, d_6(\lambda)$. Además, $d_5(\lambda)=d_6(\lambda)=0$, ya que el rango de F debe ser 4. Si descomponemos $d_1(\lambda), \ldots, d_4(\lambda)$ en factores debemos obtener los siete divisores elementales indicados. Puesto que, sin embargo, $d_4(\lambda)$ es divisible por $d_3(\lambda)$, $d_2(\lambda)$ y $d_1(\lambda)$, resulta que en $d_4(\lambda)$ figuran los divisores elementales de potencia superior pertenecientes a todos los polinomios irreducibles. Luego, $d_4(\lambda)=\lambda^2(\lambda+1)^3(\lambda-1)$. Entre los divisores elementales restantes λ, λ^2, $\lambda+1$ y $\lambda-1$ los de potencia superior deben componer $d_3(\lambda)$; por consiguiente, $d_3(\lambda)=\lambda^2(\lambda+1)(\lambda-1)$. A su vez, los de potencia superior de entre los que quedan deben formar $d_2(\lambda)$, es decir, $d_2(\lambda)=\lambda$. Puesto que todos los divisores elementales han sido ya distribuidos, resulta que $d_1(\lambda)=1$. Es obvio que este método se puede aplicar en cualquier caso, lo que demuestra el teorema.

Los divisores elementales dependen del campo principal K. Por ejemplo, supongamos que los factores invariantes de una λ-matriz F son iguales a λ^2+1 y $(\lambda^2+1)^2$. Si el campo principal es cuerpo de los números reales, el polinomio λ^2+1 es irreducible y los divisores elementales de la matriz F son λ^2+1 y $(\lambda^2+1)^2$. Sin embargo, si el campo principal es cuerpo de los números complejos, se tiene $\lambda^2+1=(\lambda-i)(\lambda+i)$ y los divisores elementales de F serán $\lambda+i$, $(\lambda+i)^2$, $\lambda-i$ y $(\lambda-i)^2$.

14.2. Divisores elementales de una matriz descompuesta. Para obtener los divisores elementales de una λ-matriz que tiene la forma canónica diagonal es suficiente tomar, de acuerdo con la definición, todos los divisores elementales de sus elementos diagonales. Mostremos que esta misma regla tiene lugar también para una λ-matriz diagonal cualquiera.

LEMA. *El sistema de divisores elementales de una λ-matriz diagonal cualquiera F es la unión de los divisores elementales de sus elementos diagonales.*

Supongamos que los elementos diagonales de F son los polinomios $f_1(\lambda)$, $f_2(\lambda)$, $\ldots$, $f_n(\lambda)$. Podemos aceptar, sin perder generalidad, que todos estos polinomios son diferentes de cero y que sus coeficientes principales son iguales a 1. Indiquemos por $D_k(\lambda)$ el máximo común divisor de los menores de orden k de la matriz F. Puesto que los coeficientes principales de los polinomios $f_1(\lambda), \ldots$ $\ldots, f_n(\lambda)$ son iguales a la unidad, resulta que $D_n(\lambda)$ es el determinante de la matriz F, es decir,

$$D_n(\lambda) = f_1(\lambda) f_2(\lambda) \ldots f_n(\lambda).$$

Pero

$$D_n(\lambda) = d_1(\lambda) d_2(\lambda) \ldots d_n(\lambda),$$

donde $d_1(\lambda), \ldots, d_n(\lambda)$ son los factores invariantes de la matriz F. Por esto, indicando por $\varepsilon_1(\lambda), \varepsilon_2(\lambda), \ldots, \varepsilon_s(\lambda)$ los distintos factores irreducibles de los polinomios $f_1(\lambda), \ldots, f_n(\lambda)$, podemos ver que todo divisor elemental de la matriz F es una potencia de uno de los polinomios $\varepsilon_1(\lambda), \ldots, \varepsilon_s(\lambda)$. Despejemos ahora en $f_1(\lambda), \ldots, f_n(\lambda)$ las potencias máximas de $\varepsilon_1(\lambda)$ por las que son divisibles estos polinomios. Sea

$$f_i(\lambda) = [\varepsilon_1(\lambda)]^{s_i} g_i(\lambda) \quad (i = 1, 2, \ldots, n),$$

donde $g_i(\lambda)$ no es divisible por $\varepsilon_1(\lambda)$. Queremos probar que $[\varepsilon_1(\lambda)]^{s_1}, \ldots, [\varepsilon_1(\lambda)]^{s_n}$ es precisamente el sistema de divisores elementales de la matriz F pertenecientes al polinomio irreducible $\varepsilon_1(\lambda)$. Puesto que los divisores elementales de la matriz F no dependen del orden de sus filas y de sus columnas, podemos disponer estas filas y columnas de modo que

$$s_1 \leqslant s_2 \leqslant \ldots \leqslant s_n. \tag{1}$$

Hallemos la potencia superior con la que $\varepsilon_1(\lambda)$ figura en $D_k(\lambda)$. Por definición, $D_k(\lambda)$ es el máximo común divisor de los menores de orden k de la matriz F, entre los cuales, como hemos visto anteriormente (p. 13.3), serán diferentes de cero sólo los menores iguales a

$$f_{v_1}(\lambda) \ldots f_{v_k}(\lambda) = [\varepsilon_1(\lambda)]^{s_{v_1} + \ldots + s_{v_k}} g_{v_1}(\lambda) \ldots g_{v_k}(\lambda).$$

En vista de las desigualdades (1), la menor potencia de $\varepsilon_1(\lambda)$ la tiene el menor

$$f_1(\lambda) \ldots f_k(\lambda) = [\varepsilon_1(\lambda)]^{s_1 + \ldots + s_k} g_1(\lambda) \ldots g_k(\lambda).$$

Por consiguiente, $D_k(\lambda)$ contiene $\varepsilon_1(\lambda)$ en la potencia $s_1 + \ldots + s_k$. Sustituyendo aquí k por $k-1$, obtenemos que $D_{k-1}(\lambda)$ contiene $\varepsilon_1(\lambda)$ en la potencia $s_1 + \ldots + s_{k-1}$. Pero el factor invariante $d_k(\lambda)$ es igual al cociente de $D_k(\lambda)$ por $D_{k-1}(\lambda)$. Por esto $d_k(\lambda)$ contiene $\varepsilon_1(\lambda)$ exactamente en la potencia s_k. Luego, los divisores elementales de la matriz F pertenecientes al polinomio irreducible $\varepsilon_1(\lambda)$ coinciden con los divisores elementales pertenecientes a $\varepsilon_1(\lambda)$ de los elementos diagonales de la matriz F. Puesto que nuestro razonamiento es válido también para los demás polinomios $\varepsilon_2(\lambda), \ldots, \varepsilon_s(\lambda)$. el lema queda demostrado en el caso más general.

TEOREMA 2. *El sistema de divisores elementales de una matriz descompuesta es igual a la unión de los sistemas de divisores elementales de sus células.*

Sea F una λ-matriz de forma celular diagonal

$$F=\begin{bmatrix} F_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & F_m \end{bmatrix}.$$

Las transformaciones elementales de cada una de las células $F_1, \ldots, F_m$ pueden ser consideradas como transformaciones de toda la matriz F. Estas transformaciones no alteran la estructura celular diagonal de la matriz F y las transformaciones, realizadas con una de las células, no alteran la forma de las restantes. Por esto, mediante transformaciones elementales de la matriz F se pueden reducir todas las células a la forma diagonal. Aplicando el lema vemos que los divisores elementales de las matrices $F, F_1, \ldots, F_m$ serán las uniones de los divisores elementales de aquellos elementos diagonales que figuran en estas matrices. En particular, los divisores elementales de la matriz F se obtienen uniendo los divisores elementales de todos los elementos diagonales, es decir, uniendo los divisores elementales de todas las células $F_1, \ldots, F_m$, que es lo que se quería demostrar.

Ejemplos y problemas

1. Hállense los divisores elementales de las matrices siguientes:

$$\begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 & \lambda+2 \\ 0 & 1 & \lambda+2 & \\ 1 & \lambda+2 & & \\ \lambda+2 & & & \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 0 & 0 & 0 & \lambda^2 \\ 0 & 0 & \lambda(\lambda-1) & \\ 0 & (\lambda-1)^2 & & \\ \lambda(\lambda-1) & & & \end{bmatrix}.$$

2. Hállense los factores invariantes de las matrices:

$$\begin{bmatrix} \lambda(\lambda+1) & & & \\ & \lambda^2 & & \\ & & (\lambda+1)^2 & \\ & & & \lambda(\lambda-1) \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 & 2 & 3 \\ & \lambda & 1 & 2 \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 1 & \lambda & 0 \\ 0 & 0 & 1 & \lambda \end{bmatrix}.$$

3. Hállense los divisores elementales de la matriz:

$$\begin{bmatrix} \lambda^2+2 & \lambda^2+1 & \lambda^2+1 \\ 3 & \lambda^2+1 & 3 \\ \lambda^2+1 & \lambda^2+1 & \lambda^2+1 \end{bmatrix}$$

en el cuerpo de los números racionales, en el cuerpo de los números reales y en el cuerpo de los números complejos.

§ 15. Formas normales de la matriz de una aplicación lineal

15.1. División de λ-matrices. Se llama *polinomio matricial* en la variable λ la expresión de tipo

$$F(\lambda)=A_0\lambda^m+A_1\lambda^{m-1}+A_2\lambda^{m-2}+\ldots+A_m, \tag{1}$$

donde $A_0, \ldots, A_m$ son matrices cuadradas de un mismo orden y con elementos de un campo principal K. Dos polinomios se llaman iguales, si son iguales las matrices que figuran en estos polinomios en los términos con la misma potencia de la variable λ. Los λ-polinomios matriciales se suman y se multiplican siguiendo las reglas habituales. Está claro que todo λ-polinomio puede ser representado mediante una matriz, cuyos términos son polinomios corrientes en λ, y viceversa. Por ejemplo

$$\begin{bmatrix}1 & 2\\ 0 & 3\end{bmatrix}+\begin{bmatrix}5 & 6\\ 7 & -2\end{bmatrix}\lambda+\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\lambda^2=\begin{bmatrix}\lambda^2+5\lambda+1 & 6\lambda+2\\ 7\lambda & \lambda^2-2\lambda+3\end{bmatrix}.$$

Luego, los λ-polinomios matriciales son simplemente una forma especial de representación de λ-matrices.

Si la matriz A_0 de un λ-polinomio (1) es diferente de la nula, se dice que m es el grado del polinomio matricial. El polinomio $F(\lambda)$ se llama *regular*, si la matriz A_0 es invertible.

Es evidente que *el grado de una suma de λ-polinomios matriciales no sobrepasa el máximo de los grados de los sumandos*. Mediante ejemplos es fácil comprobar que el grado del producto de dos λ-polinomios matriciales puede resultar menor que la suma de los grados de los factores. Sin embargo, *si al menos uno de los dos polinomios que se multiplican es regular, el grado del producto es igual exactamente a la suma de los grados de los factores.*

En efecto, si

$$\begin{aligned}A(\lambda)&=A_0\lambda^m+A_1\lambda^{m-1}+\ldots+A_m,\\ B(\lambda)&=B_0\lambda^p+B_1\lambda^{p-1}+\ldots+B_p\end{aligned} \tag{2}$$

y si B_0 es invertible, el término principal del producto $A(\lambda)\cdot B(\lambda)$ es igual a $A_0B_0\lambda^{m+p}$ y de $A_0B_0=O$ resulta que $O=A_0B_0B_0^{-1}=A_0$ lo que contradice a lo supuesto.

TEOREMA 1. *Cualesquiera que sean un λ-polinomio matricial $A(\lambda)$ y un λ-polinomio regular $B(\lambda)$ existen unos λ-polinomios $P(\lambda)$, $S(\lambda)$, $Q(\lambda)$ y $R(\lambda)$ que satisfacen las condiciones*

a) $A(\lambda)=B(\lambda)P(\lambda)+S(\lambda)$ y $A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda)$;

b) *o bien $S(\lambda)=O$ o bien el grado de $S(\lambda)$ es menor que el grado de $B(\lambda)$; o bien $R(\lambda)=O$ o bien el grado de $R(\lambda)$ es menor que el grado de $B(\lambda)$.*

Los polinomios indicados se determinan unívocamente por las condiciones a) *y* b) *y se dice que* $P(\lambda)$ *y* $S(\lambda)$ *son el cociente y el resto a la izquierda y que* $Q(\lambda)$ y $R(\lambda)$ *son el cociente y el resto a la derecha de la división de* $A(\lambda)$ *por* $B(\lambda)$.

La demostración consistirá en que indicaremos un método que permite encontrar los cocientes y los restos y que, en el caso considerado, coincide con el algoritmo de división de los polinomios corrientes. Supongamos, por ejemplo, que $A(\lambda)$ y $B(\lambda)$ son de forma (2). Si es $m < p$, podemos tomar $P(\lambda) = O$ y $S(\lambda) = A(\lambda)$. Sea, pues, $m \geqslant p$. Formamos las siguientes diferencias sucesivas

$$\begin{aligned} S_1(\lambda) &= A(\lambda) - A_0 B_0^{-1} \lambda^{m-p} B(\lambda) = C_0^{(1)} \lambda^{m_1} + \ldots, \\ S_2(\lambda) &= S_1(\lambda) - C_0^{(1)} B_0^{-1} \lambda^{m_1 - p} B(\lambda) = C_0^{(2)} \lambda^{m_2} + \ldots, \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S_{t+1}(\lambda) &= S_t(\lambda) - C_0^{(t)} B_0^{-1} \lambda^{m_t - p} B(\lambda) = C_0^{(t+1)} \lambda^{m_{t+1}} + \ldots, \end{aligned}$$

donde los puntos suspensivos representan términos de grados inferiores y donde $m_{t+1} < p$. Sumando estas igualdades y tomando

$$\begin{aligned} S(\lambda) &= S_{t+1}(\lambda), \\ P(\lambda) &= A_0 B_0^{-1} \lambda^{m-p} + C_0^{(1)} B_0^{-1} \lambda^{m_1 - p} + \ldots + C_0^{(t)} B_0^{-1} \lambda^{m_t - p}, \end{aligned}$$

obtenemos la primera de las relaciones a). Análogamente se determinan $Q(\lambda)$ y $R(\lambda)$.

Resta demostrar la unicidad. Pero, si

$$A(\lambda) = B(\lambda) P(\lambda) + S(\lambda) = B(\lambda) P_*(\lambda) + S_*(\lambda),$$

se tiene

$$B(\lambda)[P(\lambda) - P_*(\lambda)] = S_*(\lambda) - S(\lambda).$$

Puesto que el polinomio $B(\lambda)$ es regular, resulta que si la diferencia $P(\lambda) - P_*(\lambda)$ es no nula, el grado del primer miembro es no menor que el grado de $B(\lambda)$, mientras que el grado del segundo miembro es, indudablemente, menor que el grado de $B(\lambda)$. Por esto debe ser $P(\lambda) = P_*(\lambda)$ y $S(\lambda) = S_*(\lambda)$.

15.2. Equivalencia escalar. Según el p. 13.4, dos λ-matrices $A(\lambda)$ y $B(\lambda)$ son equivalentes si, y sólo si, existen unas λ-matrices $U(\lambda)$ y $V(\lambda)$, de determinantes no nulos que no dependen de λ, que satisfacen la relación

$$A(\lambda) = U(\lambda)\, B(\lambda)\, V(\lambda). \tag{3}$$

Diremos que la matriz $A(\lambda)$ es *escalarmente equivalente* a la matriz $B(\lambda)$, si existen unas matrices regulares U y V, formadas *por elementos que no dependen de* λ, que satisfacen la relación (3). Las matrices formadas por elementos que no dependen de λ se llamarán *escalares*.

TEOREMA 2. *Si dos λ-polinomios de primer grado $A\lambda+B$ y $C\lambda+D$ son regulares y equivalentes, también son escalarmente equivalentes.*

Por hipótesis tenemos

$$A\lambda+B=U(\lambda)(C\lambda+D)V(\lambda), \tag{4}$$

donde $U(\lambda)$ y $V(\lambda)$ son matrices de determinantes constantes diferentes de cero. Indiquemos por P y S el cociente y el resto a la izquierda de la división de $U(\lambda)$ por $A\lambda+B$ e indiquemos por Q y R el cociente y el resto a la derecha de la división de $V(\lambda)$ por $A\lambda+B$. Es decir,

$$U=(A\lambda+B)P+S \quad \text{y} \quad V=Q(A\lambda+B)+R. \tag{5}$$

Las matrices S y R son escalares, ya que son de grado menor que la unidad. Probemos que

$$A\lambda+B=S(C\lambda+D)R. \tag{6}$$

Efectivamente, multiplicando por U^{-1} ambos miembros de la igualdad (4), sustituyendo V por su expresión (5) y agrupando los términos, obtenemos

$$[U^{-1}-(C\lambda+D)Q](A\lambda+B)=(C\lambda+D)R.$$

Comparando los grados del primero y del segundo miembros veremos que la expresión que figura entre los corchetes debe ser igual a una matriz escalar T; tenemos, pues, que

$$T=U^{-1}-(C\lambda+D)Q \quad \text{y} \quad T(A\lambda+B)=(C\lambda+D)R. \tag{7}$$

De aquí resulta que

$$E=U(C\lambda+D)Q+UT=(A\lambda+B)V^{-1}Q+UT=$$
$$=(A\lambda+B)V^{-1}Q+[(A\lambda+B)P+S]T,$$

es decir,

$$E=(A\lambda+B)[V^{-1}Q+PT]+ST.$$

Pero el segundo miembro puede ser de grado nulo sólo en el caso en que sea cero la expresión que figura entre los corchetes, de donde

$$E=ST \quad \text{y} \quad T=S^{-1}.$$

Comparando con (7), obtenemos (6), donde S y, por consiguiente, R son matrices escalares invertibles.

15.3. Criterio de semejanza de matrices. Según el p. 3.1, dos matrices A y B sobre un cuerpo conmutativo K se llaman *semejantes* si existe una matriz invertible T, formada por elementos de K, tal que

$$A=T^{-1}BT. \tag{8}$$

En el capítulo III se ha explicado la importancia que tiene el encontrar las condiciones de semejanza de unas matrices dadas. Empleando los resultados obtenidos es fácil indicar las condiciones necesarias y suficientes de semejanza de matrices y resolver, con ello, uno de los problemas principales de la teoría de matrices.

TEOREMA 3. *Para que las matrices A y B, definidas sobre un cuerpo conmutativo arbitrario K, sean semejantes es necesario y suficiente que sus matrices características $\lambda E - A$ y $\lambda E - B$ sean equivalentes.*

La necesidad es evidente, ya que de (8) se deduce que

$$\lambda E - A = T^{-1} (\lambda E - B) T,$$

es decir, que $\lambda E - B$ y $\lambda E - A$ son equivalentes y, es más, son escalarmente equivalentes. Viceversa, supongamos que las matrices $\lambda E - B$ y $\lambda E - A$ son equivalentes. Puesto que representan unos λ-polinomios matriciales regulares y de primer grado, resulta que estas matrices son, en virtud del teorema 2, escalarmente equivalentes, es decir, existen unas matrices S y R escalares y regulares tales que

$$\lambda E - A = S (\lambda E - B) R.$$

Comparando en esta igualdad los coeficientes de λ y los términos independientes, obtenemos

$$E = SR \quad \text{y} \quad A = SBR,$$

de donde resulta

$$A = R^{-1} B R \tag{9}$$

que es lo que se quería demostrar.

De los teoremas demostrados se puede extraer el siguiente algoritmo para determinar la semejanza de las matrices A y B: formamos las matrices características $\lambda E - A$ y $\lambda E - B$ y, empleando el proceso descrito en el p. 13.2, las reducimos mediante transformaciones elementales a la forma canónica diagonal. Si estas formas coinciden, las matrices A y B son semejantes; si las formas son distintas, las matrices A y B no son semejantes.

A veces, además de establecer el hecho mismo de semejanza, es necesario hallar también la matriz transformadora T tal que $B = T^{-1} A T$. Con este fin, para pequeños valores del orden n de las matrices consideradas, se toma $T = \| t_{ij} \|$, se escribe la igualdad matricial $TB = AT$ en forma de n^2 igualdades entre los elementos de TB y AT y se consideran estas igualdades como ecuaciones lineales homogéneas respecto a las n^2 incógnitas $t_{ij} (i, j = 1, 2, \ldots, n)$. Resolviendo este sistema, se determina T.

El método expuesto resulta muy voluminoso para grandes valores de n y en este caso es preferible seguir el camino indicado en la propia demostración del teorema 3. Ante todo, conociendo las transformaciones elementales que reducen las matrices $\lambda E - A$ y $\lambda E - B$ a la forma canónica diagonal y conociendo, por consiguiente, las transformaciones elementales que reducen $\lambda E - A$ a $\lambda E - B$, podemos encontrar, de acuerdo con el p. 13.4, unas λ-matrices $U(\lambda)$ y $V(\lambda)$ tales que

$$\lambda E - A = U(\lambda)(\lambda E - B)V(\lambda).$$

Calculando el resto a la derecha R de la división de $V(\lambda)$ por $\lambda E - A$, tendremos, según (9), $A = R^{-1}BR$, es decir, R será una de las matrices transformadoras que buscamos.

Notemos que para determinar R no es necesario realizar de hecho la división de $V(\lambda)$ por $\lambda E - A$. En efecto, representando $V(\lambda)$ en la forma

$$V(\lambda) = V_0\lambda^k + V_1\lambda^{k-1} + \ldots + V_k$$

y aplicando el esquema de división del p. 15.1, obtenemos

$$R = V_0A^k + V_1A^{k-1} + \ldots + V_k. \tag{10}$$

Análogamente, representando el polinomio $U(\lambda)$ en la forma «izquierda»

$$U(\lambda) = \lambda^l U_0 + \lambda^{l-1}U_1 + \ldots + U_l$$

y realizando la división a la izquierda por $\lambda E - A$, obtenemos para el resto a la izquierda S la expresión

$$S = A^lU_0 + A^{l-1}U_1 + \ldots + U_l. \tag{11}$$

Estas fórmulas para los restos de la división de un λ-polinomio matricial por el binomio $\lambda E - A$ son totalmente análogas a la fórmula de Bezout $r = f(a)$ para el resto de la división de un polinomio corriente $f(\lambda)$ por el binomio $\lambda - a$. Por esto las fórmulas (10) y (11) a veces se denominan fórmulas matriciales de Bezout para los restos.

15.4. Forma normal de Jordan. En el p. 12.2 hemos introducido unas matrices de forma especial que hemos llamado *matrices de Jordan* y hemos estudiado algunas propiedades de las aplicaciones lineales, cuyas matrices en un sistema de coordenadas adecuado tienen la forma de Jordan. Sin embargo, hemos dejado sin resolver el problema principal acerca de las condiciones en las que la matriz de una aplicación puede ser reducida a la forma de Jordan. Pero ahora tenemos ya todos los medios necesarios para resolver este problema.

TEOREMA 4. *Toda matriz cuadrada sobre el cuerpo de los números complejos, así como sobre cualquier otro cuerpo conmutativo algebraicamente cerrado, es semejante a una matriz de la forma de Jordan. Dos matrices de Jordan son semejantes si, y sólo si, están compuestas por las mismas células de Jordan y difieren una de la otra a lo sumo en la disposición de las células a lo largo de la d·ıgonal principal.*

Anteponemos a la demostración del teorema dos ˡemas que tienen también interés por sí solos.

LEMA 1. *La matriz característica de una célula de Jordan tiene sólo un divisor elemental $(\lambda-\rho)^n$, donde n es el orden de la célula y ρ es su valor propio.*

La matriz característica de una célula de Jordan dada (véase el p. 12.2) es de la forma

$$\lambda E - A = \begin{bmatrix} \lambda-\rho & -1 & 0 & \dots & 0 & 0 \\ & \lambda-\rho & -1 & \dots & 0 & 0 \\ & & & \dots & \dots & \dots \\ & & & & \lambda-\rho & -1 \\ & & & & & \lambda-\rho \end{bmatrix}.$$

Calculemos el máximo común divisor $D_k(\lambda)$ de los menores de orden k de la matriz $\lambda E - A$. Ante todo, tenemos

$$D_n(\lambda) = |\lambda E - A| = (\lambda-\rho)^n.$$

Después, $D_{n-1}(\lambda)$ es el máximo común divisor de todos los menores de orden $n-1$. Pero entre los últimos figura el menor

$$\begin{vmatrix} -1 & 0 & \dots & 0 & 0 \\ \lambda-\rho & -1 & \dots & 0 & 0 \\ & \lambda-\rho & \dots & 0 & 0 \\ & & \dots & \dots & \dots \\ & & & \lambda-\rho & -1 \end{vmatrix} = (-1)^{n-1}$$

que se obtiene al suprimir la primera columna y la última fila en la matriz $\lambda E - A$. Puesto que este menor es igual a ± 1, resulta que $D_{n-1}(\lambda) = 1$. Indiquemos por $d_1(\lambda), \dots, d_n(\lambda)$ los factores invariantes de la matriz $\lambda E - A$. De las relaciones

$$D_{n-1}(\lambda) = d_1(\lambda) \ \dots \ d_{n-1}(\lambda) = 1,$$
$$D_n(\lambda) = d_1(\lambda) \ \dots \ d_{n-1}(\lambda)\, d_n(\lambda) = (\lambda-\rho)^n$$

se desprende que $d_1(\lambda) = \dots = d_{n-1}(\lambda) = 1$ y que $d_n(\lambda) = (\lambda-\rho)^n$. Por consiguiente, $\lambda E - A$ tiene sólo un divisor elemental y este divisor es igual a $(\lambda-\rho)^n$.

LEMA 2. *El sistema de divisores elementales de la matriz característica de una matriz de Jordan se compone de los divisores elementales de sus células de Jordan y determina unívocamente, salvo el orden de secuencia de las células a lo largo de la diagonal principal, la forma de la matriz de Jordan.*

Una matriz de Jordan es, por definición, una matriz celular diagonal con células de Jordan a lo largo de la diagonal principal. Por ello, la matriz característica de una matriz de Jordan se descompone en las matrices características de las células de Jordan aisladas. De aquí se deduce, en virtud del p. 14.2, que el sistema de divisores elementales de la matriz característica de una matriz de Jordan consta de los divisores elementales de las matrices características de cada una de las células de Jordan, habiendo un divisor por cada una de estas células. Luego, el sistema de divisores elementales de la matriz característica de una matriz de Jordan determina la forma de esta matriz unívocamente, salvo el orden de secuencia de las células a lo largo de la diagonal principal.

Las matrices características de matrices semejantes son equivalentes y, por lo tanto, tienen los mismos sistemas de divisores elementales. De aquí se deduce que las matrices semejantes de Jordan deben estar formadas por células iguales de Jordan y, para concluir la demostración del teorema 4, resta sólo saber cómo obtener a partir de toda matriz dada A la matriz de Jordan semejante a ella. Sea $(\lambda-\rho_1)^{n_1}, \ldots, (\lambda-\rho_s)^{n_s}$ el conjunto completo de los divisores elementales de la matriz característica $\lambda E-A$. Indiquemos por B la matriz celular diagonal, cuyas células diagonales son las células de Jordan con los divisores elementales señalados. Luego, la matriz $\lambda E-B$ tendrá los mismos divisores elementales que tiene $\lambda E-A$. Pero entonces, según el p. 14.1, las matrices $\lambda E-A$ y $\lambda E-B$ son equivalentes y de aquí se desprende, en virtud del p. 15.3, que la matriz A es semejante a la matriz de Jordan B. Hemos demostrado el teorema.

Los razonamientos expuestos ofrecen también una respuesta a la pregunta de cómo hallar a partir de una matriz dada A su matriz semejante de Jordan. Para ello es suficiente formar la matriz característica $\lambda E-A$, reducirla mediante transformaciones elementales a la forma canónica diagonal, descomponer en factores los polinomios diagonales, hallar los divisores elementales y construir a partir de los mismos la matriz de Jordan. Sea, por ejemplo,

$$A=\begin{bmatrix} 3 & 1 & -3 \\ -7 & -2 & 9 \\ -2 & -1 & 4 \end{bmatrix}.$$

Formamos la matriz característica

$$\lambda E-A=\begin{bmatrix} \lambda-3 & -1 & 3 \\ 7 & \lambda+2 & -9 \\ 2 & 1 & \lambda-4 \end{bmatrix}$$

y determinamos sus factores invariantes. Es fácil ver que estos factores son $1, 1, (\lambda-1)(\lambda-2)^2$. Por consiguiente, los divisores elementales serán $\lambda-1$ y $(\lambda-2)^2$ y la matriz de Jordan es

$$B=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Para terminar hagamos una observación más. Si los divisores elementales de la matriz $\lambda E-A$ resultan de primer grado, las células de Jordan de la matriz de Jordan correspondiente B serán de primer orden, es decir, la matriz B será diagonal. Recíprocamente, si la correspondiente matriz de Jordan es diagonal, los divisores elementales serán de primer grado. Por consiguiente, *para que una matriz dada sea semejante a una diagonal es necesario y suficiente que todos los divisores elementales de su matriz característica sean de primer grado.*

15.5. Forma normal natural. En el cuerpo de los números complejos toda matriz es equivalente a una matriz de Jordan determinada. Si el campo principal K es un cuerpo conmutativo arbitrario, la reducción a la forma de Jordan puede resultar imposible. En todo caso, para que esta reducción sea posible es necesario, y como puede verse fácilmente también es suficiente, que el polinomio característico de la matriz se descomponga sobre K en factores lineales. Es por esto que surge el problema de indicar una forma normal a la que puede ser reducida la matriz sobre el cuerpo conmutativo del que se toman sus elementos. Existe una cantidad infinita de formas normales de esta índole. Entre éstas se obtiene con mayor facilidad la forma normal llamada *natural*.

Sea

$$f(\lambda)=\alpha_0+\alpha_1\lambda+\ldots+\alpha_{n-1}\lambda^{n-1}+\lambda^n$$

un polinomio cualquiera de grado no nulo y de coeficiente principal igual a la unidad. Aceptaremos que los coeficientes del polinomio $f(\lambda)$ pertenecen a un cuerpo conmutativo K. La matriz

$$A=\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & \ldots & 0 & 1 \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \ldots & -\alpha_{n-2} & -\alpha_{n-1} \end{bmatrix}$$

se llama matriz *asociada* del polinomio $f(\lambda)$.

LEMA 3. *Si $f(\lambda)$ es un polinomio de grado no nulo y de coeficiente principal 1 y si A es su matriz asociada, los factores invariantes de la matriz característica $\lambda E-A$ son iguales a $1, 1, \ldots, 1, f(\lambda)$.*

En efecto, si $f(\lambda)=\alpha_0+\alpha_1\lambda+\ldots+\alpha_{n-1}\lambda^{n-1}+\lambda^n$, se tiene

$$\lambda E-A=\begin{bmatrix} \lambda & -1 & 0 & \ldots & 0 & 0 \\ 0 & \lambda & -1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \lambda & -1 \\ \alpha_0 & \alpha_1 & \alpha_2 & \ldots & \alpha_{n-2} & \lambda+\alpha_{n-1} \end{bmatrix}.$$

Suprimiendo la primera columna y la última fila obtenemos un menor igual a $(-1)^{n-1}$. Por consiguiente, el máximo común divisor $D_{n-1}(\lambda)$ de los menores de orden $n-1$ es igual a la unidad. En cuanto a $D_n(\lambda)$, este polinomio es igual al determinante de la matriz $\lambda E-A$. Desarrollando este determinante según los elementos de la última fila, encontraremos directamente que

$$D_n(\lambda)=\alpha_0+\alpha_1\lambda+\ldots+\alpha_{n-1}\lambda^{n-1}+\lambda^n.$$

De $D_{n-1}(\lambda)=1$ se deduce que $d_1(\lambda)=\ldots=d_{n-1}(\lambda)=1$ y de $D_n(\lambda)=f(\lambda)$ se desprende que $d_n(\lambda)=f(\lambda)$ que es lo que se quería demostrar.

Se dice que la matriz A es *de forma normal natural*, si A se descompone en células $A_1, A_2, \ldots, A_s$ que son las matrices asociadas de unos polinomios

$f_1(\lambda), f_2(\lambda), \ldots, f_s(\lambda)$ cada uno de los cuales es divisible por el anterior. Probemos que los factores invariantes de la matriz A son iguales a $1, \ldots, 1, f_1(\lambda), f_2(\lambda), \ldots, f_s(\lambda)$, donde el número de unidades es igual a la suma de los grados de los polinomios $f_1(\lambda), \ldots, f_s(\lambda)$ disminuida en s. Efectivamente, la matriz característica es de la forma

$$\lambda E - A = \begin{bmatrix} \lambda E_1 - A_1 & & & \\ & \lambda E_2 - A_2 & & \\ & & \ddots & \\ & & & \lambda E_s - A_s \end{bmatrix}.$$

Según el lema, toda célula $\lambda E_i - A_i$ puede ser reducida mediante transformaciones elementales a la forma diagonal en la que los elementos diagonales serán $1, \ldots, 1, f_i(\lambda)$. Cambiando después en la matriz $\lambda E - A$ el orden de las filas y de las columnas, la reduciremos a la forma diagonal con los elementos diagonales $1, \ldots, 1$ $f_1(\lambda), \ldots, f_s(\lambda)$. Puesto que todo elemento aquí es divisible por el anterior, ésta será la forma canónica diagonal y los elementos $1, \ldots, 1, f_1(\lambda), \ldots, f_s(\lambda)$ serán los factores invariantes de la matriz $\lambda E - A$. Como quiera que el orden de la matriz A era igual a la suma de los grados de los polinomios $f_1(\lambda), \ldots, f_s(\lambda)$ y el número total de factores invariantes es igual al orden de la matriz, resulta que el número de unidades que figuran entre los factores invariantes coincide con el señalado.

De nuestro razonamiento se desprende, en particular, que la forma normal natural queda unívocamente determinada por los factores invariantes de su matriz característica.

Ahora podemos demostrar en unas palabras el teorema siguiente.

TEOREMA 5. *Toda matriz A con elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a una forma normal natural, y sólo una.*

Sean $1, \ldots, 1, f_1(\lambda), \ldots, f_s(\lambda)$ los factores invariantes de la matriz $\lambda E - A$. Puesto que la suma de los grados de todos los factores invariantes debe ser igual al orden de la matriz A, resulta que el número de unidades que figuran aquí es igual a la suma de los grados de los polinomios $f_1(\lambda), \ldots, f_s(\lambda)$ disminuida en s. Construyamos para cada uno de los polinomios $f_i(\lambda)$ su matriz asociada B_i y consideremos la matriz celular diagonal de células diagonales $B_1, \ldots, B_s$. Como todo polinomio $f_i(\lambda)$ es divisible por el anterior, B es la matriz normal natural. Según hemos demostrado anteriormente, los factores invariantes de la matriz $\lambda E - B$ son iguales a $1, \ldots, 1, f_1(\lambda), \ldots, f_s(\lambda)$ y coinciden, por consiguiente, con los factores invariantes de la matriz $\lambda E - A$. De aquí se deduce, en virtud del teorema 3, que A es semejante a B. Con esto queda demostrada la posibilidad de reducir la matriz A a la forma normal natural. La unicidad resulta de que la matriz natural B se determina unívocamente por los factores invariantes, que, a su vez, se determinan unívocamente por la matriz A.

15.6. Otras formas normales. La ventaja de la forma normal natural estriba en que es absolutamente unívoca: tanto las propias células diagonales como el orden en que están dispuestas a lo largo de la diagonal principal se determinan unívocamente. Entre los defectos figura el que esta forma no ofrece la reducción a las células de menor orden posible así como el que no comprende la forma de Jordan como un caso particular.

El primero de estos defectos puede ser superado del modo siguiente. Convengamos en llamar una matriz B casinatural, si B se descompone en células $B_1, \ldots, B_s$ que representan las matrices asociadas de polinomios de tipo $[\varepsilon_1(\lambda)]^{m_1}, \ldots, [\varepsilon_s(\lambda)]^{m_s}$, donde $\varepsilon_1(\lambda), \ldots, \varepsilon_s(\lambda)$ son polinomios irreducibles sobre el campo principal K de coeficientes principales iguales a 1. Puesto que B_i es la matriz asociada del polinomio $[\varepsilon_i(\lambda)]^{m_i}$, los factores invariantes de la matriz característica $\lambda E_i - B_i$ son iguales a $1, \ldots, 1, [\varepsilon_i(\lambda)]^{m_i}$. Por consiguiente, $\lambda E_i - B_i$ tiene un único divisor elemental $[\varepsilon_i(\lambda)]^{m_i}$ y los divisores elementales de la matriz $\lambda E - B$ serán $[\varepsilon_1(\lambda)]^{m_1}, \ldots, [\varepsilon_s(\lambda)]^{m_s}$. Puesto que el rango de la matriz $\lambda E - B$ es igual a su orden y es igual a la suma de los grados de los polinomios $[\varepsilon_i(\lambda)]^{m_i}$, resulta que la matriz casinatural se determina por sus divisores elementales unívocamente, salvo el orden de disposición de las células a lo largo de la diagonal principal.

Igual que en el punto anterior de aquí se desprende directamente que toda matriz A formada por elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a la forma normal casinatural. Esta forma queda determinada por la matriz A unívocamente, salvo el orden en el que siguen las células a lo largo de la diagonal principal.

Las células de la forma casinatural no pueden ser descompuestas sobre el cuerpo conmutativo K en células de menor orden, ya que si esta descomposición fuese posible, resultaría que la matriz característica de la célula tendría por lo menos dos divisores elementales en lugar de uno.

Para concluir examinemos la forma normal que puede ser considerada como una generalización de la forma de Jordan al caso de subcuerpos conmutativos arbitrarios K del cuerpo de los números complejos [1]). Sea $\varepsilon(\lambda)$ un polinomio irreducible con coeficientes de K. Aceptaremos que el polinomio $\varepsilon(\lambda)$ es de grado no nulo y que su coeficiente principal es igual a la unidad.

LEMA 4. *Si el polinomio característico de una matriz A formada por elementos de un cuerpo conmutativo numérico K es irreducible sobre K y es igual a $\varepsilon(\lambda)$, la matriz celular*

$$B = \begin{bmatrix} A & E & 0 & \ldots & 0 \\ & A & E & \ldots & 0 \\ & & \cdot & \cdot & \cdot \\ & & & A & E \\ & & & & A \end{bmatrix},$$

donde E es la matriz unidad, tiene una matriz característica que posee un único divisor elemental $[\varepsilon(\lambda)]^m$, donde m es el número de células diagonales de la matriz B.

Sea $\lambda E - A = P$; entonces

$$\lambda E - B = \begin{bmatrix} P & -E & 0 & \ldots & 0 & 0 \\ & P & -E & \ldots & 0 & 0 \\ & & & \cdot & \cdot & \cdot \\ & & & & P & -E \\ & & & & & P \end{bmatrix}. \tag{12}$$

Si agregamos ahora a las células de una columna cualquiera de la matriz (12) las células de otra cualquiera de sus columnas multiplicadas por una λ-matriz arbitraria S, esta operación equivaldrá, obviamente, a una serie de transformaciones elementales de tipo II o IV realizadas con la matriz $\lambda E - B$, de modo que el resultado será una matriz equivalente a $\lambda E - B$. Tengamos en cuenta esta observación y realicemos sucesivamente las siguientes transformaciones con la matriz $\lambda E - B$: agreguemos a su primera columna la segunda columna multiplicada por P; agreguemos a la primera columna de la matriz nueva su tercera columna multiplicada por P^2, a su segunda columna la tercera multiplicada por P, etc. Después de estas transformaciones obtendremos la

1) Hablando más generalmente, al caso de cualesquiera cuerpos conmutativos perfectos. Pero entonces, en lugar del cuerpo de los números complejos habrá que tomar la adherencia algebraica del cuerpo conmutativo correspondiente.

matriz

$$\begin{bmatrix} 0 & -E & 0 & \dots & 0 \\ 0 & 0 & -E & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -E \\ P^m & P^{m-1} & P^{m-2} & \dots & P \end{bmatrix}.$$

Agregando ahora a la última fila la primera, la segunda, ..., filas multiplicadas, respectivamente, por P^{m-1}, P^{m-2}, ... y cambiando el orden de las columnas, obtendremos la matriz

$$\begin{bmatrix} -E & & & \\ & \ddots & & \\ & & -E & \\ & & & P^m \end{bmatrix}. \tag{13}$$

Puesto que la matriz (13) tiene la forma celular diagonal y sus células iniciales coinciden, salvo el signo, con matrices unidades, resulta que los divisores elementales de la matriz (13) coinciden con los divisores elementales de la matriz $P^m=(\lambda E-A)^m$ que pasamos ahora a examinar.

Por hipótesis, el polinomio característico de la matriz A es igual a $\varepsilon(\lambda)$ y es irreducible sobre el cuerpo conmutativo K. De aquí se desprende que todas sus raíces en el cuerpo de los números complejos son diferentes y que, por consiguiente, existe una matriz compleja T tal que TAT^{-1} es de forma diagonal y que

$$\lambda E-TAT^{-1}=\begin{bmatrix} \lambda-\alpha_1 & & & \\ & \lambda-\alpha_2 & & \\ & & \ddots & \\ & & & \lambda-\alpha_n \end{bmatrix}.$$

La matriz P^m es equivalente a la matriz $TP^mT^{-1}=(TPT^{-1})^m$ y la última es de la forma

$$[T(\lambda E-A)T^{-1}]^m=(\lambda E-TAT^{-1})^m=\begin{bmatrix} (\lambda-\alpha_1)^m & & \\ & \ddots & \\ & & (\lambda-\alpha_n)^m \end{bmatrix}.$$

Por esto, los divisores elementales de la matriz P^m en el cuerpo de los números complejos son iguales a $(\lambda-\alpha_1)^m, \dots, (\lambda-\alpha_n)^m$. Todos ellos pertenecen a diferentes polinomios irreducibles. Los factores invariantes de la matriz P^m son $1, \dots, 1, (\lambda-\alpha_1)^m \dots (\lambda-\alpha_n)^m=[\varepsilon(\lambda)]^m$. Pero el polinomio $\varepsilon(\lambda)$ es irreducible sobre el cuerpo conmutativo K y, por consiguiente, la matriz $\lambda E-P^m$ tiene en este cuerpo conmutativo sólo un divisor elemental, a saber, $[\varepsilon(\lambda)]^m$. Es decir, la matriz (13) y con ella también la matriz (12) tienen sólo un divisor elemental $[\varepsilon(\lambda)]^m$. Hemos demostrado el lema.

Si A es la matriz asociada de $\varepsilon(\lambda)$, diremos que la matriz B es *la célula generalizada de Jordan* correspondiente al divisor elemental $[\varepsilon(\lambda)]^m$. Asimismo diremos que una matriz tiene *la forma generalizada de Jordan*, si se descompone en células generalizadas de Jordan. Está claro que una matriz generalizada de Jordan se determina plenamente por sus divisores elementales.

TEOREMA 6. *Toda matriz A formada por elementos de un cuerpo conmutativo K se reduce sobre este cuerpo a la forma generalizada de Jordan. Está forma* *queda determinada por la matriz A unívocamente, salvo el orden de disposición de las células a lo largo de la diagonal principal.*

En efecto, sean $[\varepsilon_1(\lambda)]^{m_1}, \ldots, [\varepsilon_s(\lambda)]^{m_s}$ los divisores elementales de la matriz $\lambda E - A$ en el cuerpo conmutativo K. Construyamos para todo polinomio $[\varepsilon_i(\lambda)]^{m_i}$ la correspondiente célula generalizada de Jordan B_i y consideremos la matriz celular diagonal B con las células B_i a lo largo de la diagonal principal.

En virtud del lema 4, los divisores elementales de la matriz $\lambda E - B$ son iguales a los respectivos divisores elementales de la matriz $\lambda E - A$. Puesto que $\lambda E - B$ y $\lambda E - A$ son, además, del mismo rango y del mismo orden, resulta que A es semejante a B. La unicidad se deduce de que los divisores elementales de la matriz $\lambda E - A$ determinan unívocamente la matriz B.

Si el campo principal K es cuerpo de los números complejos, todos los polinomios irreducibles son del primer grado. Por consiguiente, las matrices A y B de las células generalizadas de Jordan serán del primer orden, de modo que las células generalizadas se convierten en las células corrientes de Jordan.

Consideremos también el caso en que el campo principal K es el cuerpo de todos los números reales. Los polinomios irreducibles sobre K serán de dos tipos: 1) polinomios de primer grado $\lambda - \rho$; las respectivas células generalizadas de Jordan serán células corrientes de Jordan y 2) $\varepsilon(\lambda) = \lambda^2 + p\lambda + q$, donde $p^2 - 4q < 0$; la célula asociada A es de la forma

$$\begin{bmatrix} 0 & 1 \\ -q & -p \end{bmatrix} \tag{14}$$

y la célula generalizada de Jordan correspondiente a $(\lambda^2 + p\lambda + q)^m$ tiene la forma de la matriz B del lema 4, donde en lugar de A se puede tomar una matriz de otro tipo

$$A = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix};$$

α y β son los coeficientes de la parte real y de la parte imaginaria, respectivamente, de la raíz compleja del polinomio $\lambda^2 + p\lambda + q$. Esta sustitución es posible, ya que el polinomio característico de la matriz

$$\lambda E - A = \begin{bmatrix} \lambda - \alpha & -\beta \\ \beta & \lambda - \alpha \end{bmatrix}$$

es igual a $\lambda^2 + p\lambda + q$.

Ejemplos y problemas

1. Determínese cuáles de las matrices

$$\begin{bmatrix} 3 & 1 & -3 \\ -4 & -2 & 6 \\ -1 & -1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 42 & 130 & 25 \\ -8 & -24 & -5 \\ -23 & -73 & -13 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & -1 \\ 7 & 4 & -22 \\ 2 & 1 & -5 \end{bmatrix},$$

$$\begin{bmatrix} 8 & -13 & 16 \\ -8 & 18 & -22 \\ -11 & 22 & -27 \end{bmatrix}, \quad \begin{bmatrix} 20 & -89 & -32 \\ 11 & -51 & 20 \\ 20 & -95 & -36 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{bmatrix}$$

son semejantes. Hállense sus matrices de Jordan semejantes.

2. Demuéstrese que el polinomio característico de una matriz A es igual al producto de todos los factores invariantes y que el polinomio mínimo es igual al último de los factores invariantes de la matriz característica $\lambda E - A$.

3. Hállense los polinomios mínimos de las matrices indicadas en el problema 1.

4. Demuéstrese que sobre el cuerpo de los números complejos toda matriz cuadrada es semejante a una matriz diagonal si, y sólo si, su polinomio mínimo no tiene raíces múltiples.

5. Demuéstrese que las matrices recíprocamente transpuestas A y A' son siempre semejantes.

6. Si A es una matriz regular y B es una matriz cualquiera, la matriz AB es semejante a la matriz BA. ¿Serán semejantes AB y BA siendo arbitrarias las matrices A y B?

7. Hállense las formas normales de las matrices

$$\begin{bmatrix} -3 & -1 & -3 \\ 22 & 9 & -27 \\ 5 & 2 & -6 \end{bmatrix} \text{ y } \begin{bmatrix} 1 & -1 & 1 & 3 \\ -1 & 1 & 3 & -1 \\ -1 & -3 & -1 & -1 \\ 3 & 1 & -1 & -1 \end{bmatrix}$$

sobre el cuerpo de los números racionales, sobre el cuerpo de los números reales y sobre el cuerpo de los números complejos.

8. Muéstrese que para obtener la forma normal sobre el cuerpo de los números complejos se puede tomar en lugar de las células de Jordan células de tipo

$$\begin{bmatrix} \rho & \alpha & 0 & \dots & 0 & 0 \\ & \rho & \alpha & \dots & 0 & 0 \\ & & \rho & \dots & 0 & 0 \\ & & & \dots & \dots & \\ & & & & \rho & \alpha \\ & & & & & \rho \end{bmatrix},$$

donde α es un número fijo cualquiera diferente de cero.

9. CARACTERISTICA DE SEGRE Sea $\mathcal{A}$ una aplicación lineal de un espacio vectorial complejo y sean $\rho_1, \dots, \rho_s$ distintos valores propios de esta aplicación. Indiquemos por $\sigma_{1k}, \sigma_{2k}, \dots, \sigma_{jk}$ los órdenes de las células de Jordan en la forma normal de Jordan de $\mathcal{A}$, correspondientes al valor propio ρ_k. El símbolo

$$[(\sigma_{11}, \sigma_{21}, \dots)(\sigma_{12}, \sigma_{22}, \dots) \dots (\sigma_{1s}, \sigma_{2s}, \dots)]$$

se denomina *característica de Segre de la aplicación* $\mathcal{A}$. Por ejemplo, las matrices

$$\begin{bmatrix} 1 & 1 & & \\ 0 & 1 & & \\ & & 1 & \\ & & & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & & \\ 0 & 2 & & \\ & & 2 & 1 \\ & & 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 0 & \\ & 2 & 1 & \\ & & 2 & \\ & & & 3 \end{bmatrix} \text{ y } \begin{bmatrix} 2 & 1 & 0 & 0 \\ & 2 & 1 & 0 \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}$$

tienen, respectivamente, las siguientes características de Segre: [(2,1) (1)], [(2,2)], [(3) (1)] y [(4)]. Calcúlense las características de Segre de las matrices indicadas en el problema 1.

10. CARACTERÍSTICA DE WEYR. Sea $\mathcal{A}$ la misma aplicación del problema anterior. Indiquemos por

$$\alpha_{i1}, \alpha_{i1}+\alpha_{i2}, \dots, \alpha_{i1}+\alpha_{i2}+\dots+\alpha_{ip}$$

los defectos de las aplicaciones $\mathcal{A}-\rho_i\mathcal{E}$, $(\mathcal{A}-\rho_i\mathcal{E})^2, \dots, (\mathcal{A}-\rho_i\mathcal{E})^p$, donde p es la multiplicidad del valor propio ρ_i. La fila $\{\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ip}\}$ se denomina *característica de Weyr de la aplicación* $\mathcal{A}$ correspondiente al valor propio ρ_i. Demuéstrese que las características de Weyr de las aplicaciones indicadas en el problema 9 son iguales, respectivamente, a

$$[\{2, 1\}, \{1\}], [\{2, 2\}], [\{1, 1, 1\}, \{1\}] \text{ y } [\{1, 1, 1, 1\}].$$

11. Demuéstrese que la característica de Segre y la característica de Weyr correspondientes a un mismo valor propio están relacionadas del modo siguiente.

Supongamos que la característica de Segre es (6, 4, 3, 3). Consideremos el diagrama de puntos

$$\begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & & & \end{array}$$

El número de puntos en las sucesivas columnas de este diagrama será precisamente la característica de Weyr. Para el caso considerado es igual a $\{4, 4, 4, 2, 1, 1\}$.

12. Toda matriz con divisores elementales irreducibles se denomina *semisimple*. Demuéstrese que cualquier matriz A puede ser descompuesta en la suma de una matriz semisimple y otra nilpotente permutables entre sí y que esta descomposición es única.

§ 16. Funciones de matrices

En este parágrafo serán considerados aquellos problemas del cálculo de matrices en la solución de los cuales se emplea la posibilidad de reducir las matrices a la forma normal de Jordan. De acuerdo con esto se aceptará que el campo principal es el cuerpo de todos los números complejos.

16.1. Polinomio en una matriz de Jordan. Las funciones de matrices más sencillas son los polinomios. Más tarde daremos la definición general de función de una matriz, mientras que ahora daremos la expresión explícita de un polinomio en una matriz que tiene la forma normal de Jordan. Consideremos primero una célula aislada de Jordan de orden n

$$A=\begin{bmatrix} \rho & 1 & 0 & \dots & 0 & 0 \\ & \rho & 1 & \dots & 0 & 0 \\ & & & \dots & \dots & \dots \\ & & & & \rho & 1 \\ & & & & & \rho \end{bmatrix}. \tag{1}$$

Demostremos que para cualesquiera valores naturales de m es válida la fórmula

$$A^m=\begin{bmatrix} \rho^m & \binom{m}{1}\rho^{m-1} & \dots & \binom{m}{n-1}\rho^{m-n+1} \\ & \rho^m & \dots & \binom{m}{n-2}\rho^{m-n+2} \\ & & \dots & \dots \\ & & & \rho^m \end{bmatrix}, \tag{2}$$

donde se ha tomado

$$\binom{m}{k}=\frac{m(m-1)\dots(m-k+1)}{1\cdot 2\dots k}.$$

El método de demostración más sencillo es por inducción según m. Para $m=1$ la fórmula (2) coincide con (1) y por lo tanto es verídica. Por otro lado, si la igualdad (2) es válida para un valor de m, multiplicándola por A, obtendremos, mediante el cálculo directo, que la fórmula (2) es también válida para A^{m+1}.

Sea ahora $f(\lambda)$ un polinomio en λ:

$$f(\lambda)=\alpha_0+\alpha_1\lambda+\alpha_2\lambda^2+\ldots+\alpha_k\lambda^k.$$

Tenemos, por definición,

$$f(A)=\alpha_0 E+\alpha_1 A+\alpha_2 A^2+\ldots+\alpha_k A^k.$$

Introduciendo aquí en lugar de las matrices A^m sus valores de (2), veremos que en la i-ésima fila y en la $(i+s)$-ésima columna de la matriz $f(A)$ aparece la expresión

$$\sum_{m=0}^{k}\alpha_m\frac{m(m-1)\ldots(m-s)}{1\cdot 2\ldots s}\rho^{m-s}=\frac{1}{1\cdot 2\ldots s}f^{(s)}(\rho).$$

Por consiguiente, obtenemos definitivamente que

$$f(A)=\begin{bmatrix} f(\rho) & \frac{1}{1!}f'(\rho) & \frac{1}{2!}f''(\rho) & \ldots & \frac{1}{(n-1)!}f^{(n-1)}(\rho)\\ & f(\rho) & \frac{1}{1!}f'(\rho) & \ldots & \frac{1}{(n-2)!}f^{(n-2)}(\rho)\\ & & & \ldots\ldots & \ldots\ldots\ldots\\ & & & & f(\rho)\end{bmatrix}. \qquad (3)$$

Hemos calculado el valor de un polinomio en una célula de Jordan. Pero una matriz general de Jordan A es una suma directa de células aisladas de Jordan:

$$A=A_1\dot{+}A_2\dot{+}\ \ldots\ \dot{+}A_s$$

y según el p. 1.4 tenemos

$$f(A)=f(A_1)\dot{+}f(A_2)\dot{+}\ldots\dot{+}f(A_s). \qquad (4)$$

Aquí $f(A_1)$, $\ldots$ $f(A_s)$ son polinomios en células aisladas de Jordan y sus expresiones vienen dadas por la fórmula (3). Este resultado se puede aplicar también para el cálculo de polinomios en matrices A que no tienen la forma de Jordan. En efecto, determinamos primero una matriz T tal que la matriz $T^{-1}AT=B$ tenga la forma normal de Jordan, calculamos después $f(B)$ empleando las fórmulas (3) y (4) y, finalmente, teniendo en cuenta la relación

$$f(A)=f(TBT^{-1})=Tf(B)T^{-1}$$

(p. 3.1) obtenemos el valor de $f(A)$.

16.2. Funciones escalares. El concepto general de funciones matriciales se define de la misma forma absolutamente que el concepto de funciones numéricas corrientes. A saber, consideremos un conjunto de matrices $\mathfrak{M}$. Si a toda matriz A de $\mathfrak{M}$ se pone en correspondencia una matriz B, se dice que B es una función de A definida sobre $\mathfrak{M}$. Queremos ahora poner en correspondencia a toda función numérica corriente $\rho = f(\lambda)$ — que está definida sobre un conjunto de números complejos y que satisface las condiciones indicadas más abajo — una determinada función matricial $f(A)$. Esta correspondencia se obtiene del modo siguiente. Sean dadas una función numérica $\rho = f(\lambda)$ y una matriz arbitraria A. Indiquemos por $\rho_1, \rho_2, \ldots, \rho_s$ los diferentes valores propios de la matriz A. Reduzcamos A a la forma normal de Jordan

$$T^{-1}AT = B = B_1 \dotplus B_2 \dotplus \ldots \dotplus B_t,$$

donde $B_1, \ldots, B_t$ son células de Jordan, y consideremos una de estas células, por ejemplo, la célula

$$B_i = \begin{bmatrix} \rho_i & 1 & 0 \ldots 0 \\ & \rho_i & 1 \ldots 0 \\ & & \ldots \\ & & \rho_i \end{bmatrix} \tag{5}$$

que corresponde al divisor elemental $(\lambda - \rho_i)^{n_i}$. Si la función $f(\lambda)$ está definida en una vecindad del punto ρ_i y tiene derivadas finitas $f'(\rho_i), \ldots, f^{(n_i-1)}(\rho_i)$, tomamos por definición

$$f(B_i) = \begin{bmatrix} f(\rho_i) & \frac{1}{1!}f'(\rho_i) & \ldots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\rho_i) \\ & f(\rho_i) & \ldots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\rho_i) \\ & & \ldots\ldots & \ldots\ldots \\ & & & f(\rho_i) \end{bmatrix}. \tag{6}$$

Además, si $f(\lambda)$ está definida en una vecindad de cada uno de los puntos $\rho_1, \ldots, \rho_s$ y tiene en estas vecindades derivadas de orden adecuado, tomamos también

$$f(B) = f(B_1) \dotplus f(B_2) \dotplus \ldots \dotplus f(B_t) \tag{7}$$

y

$$f(A) = Tf(B)T^{-1} = T(f(B_1) \dotplus \ldots \dotplus f(B_t))T^{-1}. \tag{8}$$

La matriz $f(A)$ se llama *valor* de la función $f(\lambda)$ para $\lambda = A$. Más abajo se demuestra que $f(A)$ no depende de cómo se reduce la mátriz A a la forma normal[1] y es, por consiguiente, una función matricial de A. Esta función se denomina *correspondiente* a la función numérica $f(\lambda)$. Está claro que no todas, ni mucho menos, funciones matriciales poseen las correspondientes funciones numéricas.

[1] Es decir, de cómo se escoge la matriz T.

Aquellas para las cuales existen funciones numéricas correspondientes se denominan *funciones escalares*.

Indiquemos algunas propiedades elementales de las funciones escalares:

A) Si $f(\lambda)$ es un polinomio en λ, el valor de la función escalar $f(A)$ coincide con el valor del polinomio $f(\lambda)$ para $\lambda = A$ en el sentido del p.1.2.

Efectivamente, la propia definición de las funciones escalares se ha realizado de modo que, para el caso de polinomios, coincida con la antigua.

B) Sea A una matriz y sean $f_1(\lambda)$ y $f_2(\lambda)$ unas funciones numéricas para las cuales tienen sentido las expresiones $f_1(A)$ y $f_2(A)$. Si $f(\lambda) = f_1(\lambda) + f_2(\lambda)$, también $f(A)$ tiene sentido y $f(A) = f_1(A) + f(A_2)$.

C) Si A es una matriz, $f_1(\lambda)$ y $f_2(\lambda)$ son unas funciones numéricas tales que $f_1(A)$ y $f_2(A)$ tienen sentido y si $f(\lambda) = f_1(\lambda) f_2(\lambda)$, también $f(A)$ tiene sentido y $f(A) = f_1(A) f_2(A)$.

Las demostraciones de las propiedades B) y C) son semejantes y por ello nos limitaremos al caso de la propiedad C). Para calcular los valores $f_1(A)$, $f_2(A)$ y $f(A)$ debemos, por definición, reducir A a la forma normal de Jordan B y emplear las fórmulas (7) y (8). Si se logra demostrar que $f(B) = f_1(B) f_2(B)$, de la fórmula (8) resultará directamente que $f(A) = f_1(A) f_2(A)$. Por otro lado, se tiene

$$f(B) = f(B_1) \dotplus f(B_2) \dotplus \ldots \dotplus f(B_t),$$

$$f_1(B) f_2(B) = f_1(B_1) f_2(B_1) \dotplus \ldots \dotplus f_1(B_t) f_2(B_t)$$

y, por consiguiente, todo se reduce a la demostración de las igualdades

$$f(B_i) = f_1(B_i) f_2(B_i) \quad (i = 1, 2, \ldots, t),$$

donde B_i son células de Jordan. Tomando los valores de $f_1(B_i)$ y $f_2(B_i)$ según las fórmulas (6) y multiplicándolos, veremos que en la k-ésima fila y en la $(k+j)$-ésima columna de la matriz $f_1(B_i) f_2(B_i)$ aparece el elemento igual a

$$f_1(\rho) \cdot \frac{1}{j!} f_2^{(j)}(\rho) + \frac{1}{1!} f_1'(\rho) \cdot \frac{1}{(j-1)!} f_2^{(j-1)}(\rho) + \ldots + \frac{1}{j!} f_1^{(j)}(\rho) \cdot f_2(\rho).$$

Esta expresión puede ser representada en la forma

$$\frac{1}{j!}\left[f_1(\rho) f_2^{(j)}(\rho) + \frac{j}{1!} f_1'(\rho) f_2^{(j-1)} \rho + \ldots + f_1^{(j)}(\rho) f_2(\rho)\right]$$

que, de acuerdo con la regla de la derivada de un producto de funciones, coincide con $\frac{1}{j!} f^{(j)}(\rho)$. Por consiguiente

$$f_1(B_i) f_2(B_i) = f(B_i)$$

y la proposición C) queda demostrada.

Análogamente, empleando esta vez la regla de la derivación de una función de función, se podría demostrar que siendo $\varphi(\lambda)$ y $f(\varphi(\lambda))$ funciones numéricas que satisfacen las condiciones en las que la expresión $f(\varphi(A))$ tenga sentido y siendo $\psi(\lambda)=f(\varphi(\lambda))$ se tiene $\psi(A)=f(\varphi(A))$.

D) Sea A una matriz cuyos valores propios son $\rho_1, \rho_2, \ldots, \rho_n$ con la particularidad de que todo valor propio aparece aquí tantas veces como su multiplicidad lo indique. Si $f(\lambda)$ es una función numérica y si $f(A)$ tiene sentido, los valores propios de la matriz $f(A)$ son iguales a $f(\rho_1), f(\rho_2), \ldots, f(\rho_n)$.

En efecto, los valores propios respectivos de las matrices $f(A)$ y $T^{-1}f(A)T=f(T^{-1}AT)$ coinciden y, por lo tanto, podemos aceptar que A tiene la forma normal de Jordan. Las fórmulas (5) y (6) muestran que en este caso $f(A)$ tiene la forma triangular con la particularidad de que a lo largo de la diagonal principal de $f(A)$ figuran los números $f(\rho_1), f(\rho_2), \ldots, f(\rho_n)$. Puesto que los elementos diagonales de una matriz triangular son sus valores propios, la proposición D) queda demostrada.

Consideremos dos ejemplos. 1) Sea $f(\lambda)=\lambda^{-1}$. Esta función está definida en todo punto, salvo $\lambda=0$, y para todos los valores de λ diferentes de cero tiene derivadas de cualquier orden. Por consiguiente, si la matriz A no posee valores propios nulos, es decir, si A es regular, resulta que $f(A)$ tiene sentido. Pero $\lambda\cdot f(\lambda)=1$ y por lo tanto $A\cdot f(A)=E$, de donde tenemos $f(A)=A^{-1}$. Es decir, a la función λ^{-1} le corresponde la matriz inversa.

2) Sea $f(\lambda)=\sqrt{\lambda}$. Para $\lambda\neq 0$ esta función tiene derivadas finitas de cualquier orden. Por consiguiente, la expresión $\sqrt{A}$ tiene sentido para todas las matrices regulares A[1]. Tomando $\lambda=A$ en la relación

$$f(\lambda)f(\lambda)=\lambda$$

obtenemos

$$f(A)f(A)=A.$$

Hemos demostrado, por consiguiente, que de toda matriz regular se puede extraer la raíz cuadrada.

16.3. Representación de los valores de funciones por polinomios. En todos los cursos del Algebra superior se considera el problema de cómo a partir de un sistema dado de números diferentes $\rho_1, \rho_2, \ldots, \rho_s$ y de otro sistema cualquiera de números $\alpha_1, \alpha_2, \ldots, \alpha_s$ obtener un polinomio $f(\lambda)$ que en los puntos $\rho_1, \rho_2, \ldots, \rho_s$ tome

[1] Para evitar la multiformidad de $\sqrt{\lambda}$ y hacer más rigurosos los razonamientos, es suficiente efectuar en el plano complejo de la variable λ un corte desde el origen de coordenadas a lo largo de un rayo que no contenga ninguno de los valores propios de la matriz A y considerar sólo una de las ramas del radical.

los valores $\alpha_1, \alpha_2, \ldots, \alpha_s$, respectivamente. La solución se ofrece en forma del conocido polinomio de interpolación de Lagrange.

En lo sucesivo será importante saber constriur los polinomios cuando tanto ellos como sus derivadas hasta un orden determinado toman los valores dados en los puntos $\rho_1, \rho_2, \ldots, \rho_s$. Este problema es, por consiguiente, una generalización directa del problema anterior. Enunciemos la proposición referente a su solución en forma de un lema especial.

LEMA. *Sean $\rho_1, \rho_2, \ldots, \rho_s$ unos números diferentes dados y sea dada una tabla de $(k+1)s$ números cualesquiera α_{ij}. Existe un polinomio $p(\lambda)$ que en todo punto ρ_i toma el valor α_{i0} mientras que su j-ésima derivada toma el valor α_{ij} $(i=1, 2, \ldots, s;\ j=1, \ldots, k)$.*

Primero conviene construir un polinomio auxiliar $p_i(\lambda)$ que tanto él como sus derivadas hasta el orden k tomen los valores requeridos solamente en el punto ρ_i y se anulen en los demás puntos. Tomemos

$$\varphi_i(\lambda)=\beta_{i0}+\beta_{i1}(\lambda-\rho_i)+\ldots+\beta_{ik}(\lambda-\rho_i)^k,$$
$$\Phi_i(\lambda)=(\lambda-\rho_1)^{k+1}\ldots(\lambda-\rho_{i-1})^{k+1}(\lambda-\rho_{i+1})^{k+1}\ldots(\lambda-\rho_s)^{k+1},$$
$$p_i(\lambda)=\varphi_i(\lambda)\Phi_i(\lambda),$$

donde $\beta_{i0}, \beta_{i1}, \ldots, \beta_{ik}$ son unos números por ahora indeterminados. Es evidente que para cualesquiera valores de $\beta_{i0}, \ldots, \beta_{ik}$ se tiene

$$p_i(\rho_j)=p_i'(\rho_j)=\ldots=p_i^{(k)}(\rho_j)=0 \quad (j\neq i).$$

De acuerdo a la regla de la derivación de un producto tenemos

$$p_i^{(j)}(\rho_i)=\varphi_i^{(j)}(\rho_i)\Phi_i(\rho_i)+j\varphi_i^{(j-1)}(\rho_i)\Phi_i'(\rho_i)+\ldots+\varphi_i(\rho_i)\Phi_i^{(j)}(\rho_i),$$

es decir,

$$\alpha_{ij}=j!\,\beta_{ij}\Phi_i(\rho_i)+j!\,\beta_{i,\,j-1}\Phi_i'(\rho_i)+\ldots+\beta_{i0}\Phi_i^{(j)}(\rho_i). \tag{9}$$

Como $\Phi_i(\rho_i)\neq 0$, tomando $j=0, 1, \ldots, k$ podemos determinar de las relaciones (9) sucesivamente los números $\beta_{i0}, \beta_{i1}, \ldots, \beta_{ik}$ y, con ello, calcular $p_i(\lambda)$. El polinomio

$$p(\lambda)=p_1(\lambda)+p_2(\lambda)+\ldots+p_s(\lambda)$$

satisfará, obviamente, las condiciones del lema.

Consideremos una función numérica $f(\lambda)$ y una matriz A tal que el valor $f(A)$ esté definido. Probemos que existe entonces un polinomio $p(\lambda)$ tal que $p(A)$ es igual a $f(A)$. Indiquemos por $\rho_1, \rho_2, \ldots, \rho_s$ los distintos valores propios de la matriz A. Sea n su orden. De acuerdo con el lema que acabamos de demostrar podemos construir un polinomio $p(\lambda)$ que satisfaga las condiciones siguientes[1]

$$p(\rho_i)=f(\rho_i),\quad p'(\rho_i)=f'(\rho_i),\quad \ldots,\quad p^{(n-1)}(\rho_i)=f^{(n-1)}(\rho_i) \tag{10}$$
$$(i=1, \ldots, s).$$

[1] Si algunas de las derivadas $f^{(j)}(\rho_i)$ sobran para la determinación de $f(A)$, los números correspondientes de (10) pueden ser sustituidos por ceros.

Para determinar el sentido de la expresión $f(A)$ sólo necesitamos conocer los valores que toman en los puntos $\rho_1, \rho_2, \ldots, \rho_s$ la función $f(\lambda)$ y sus derivadas hasta el orden $n-1$ a lo sumo. Puesto que estos valores de $f(\lambda)$ y de $p(\lambda)$ coinciden, resulta que $f(A)=p(A)$. Es decir, hemos obtenido el siguiente resultado:

TEOREMA 1. *Los valores de todas las funciones escalares de una matriz A pueden ser representados mediante polinomios en A* [1].

En particular, considerando la función $f(\lambda)=\sqrt{\lambda}$ vemos que *para toda matriz regular A existe un polinomio $p(\lambda)$ tal que*

$$p(A)\,p(A)=A.$$

Empleando el teorema 1 es fácil resolver el problema que hemos dejado pendiente en el punto anterior acerca de la unicidad de la determinación del valor de $f(A)$. En efecto, conociendo los valores de la función $f(\lambda)$ y de sus derivadas en los puntos $\rho_1, \ldots, \rho_s$ podemos construir el polinomio $p(\lambda)$ cuyo valor $p(A)$ no depende de cómo se reduce la matriz A a la forma normal de Jordan y coincide, al mismo tiempo, con el valor $f(A)$. Es decir, el valor $f(A)$, definido en el punto anterior mediante la reducción de la matriz A a la forma normal, no depende de cómo se realiza esta reducción.

Hagamos una observación más. Sea $f(\lambda)$ una función numérica y sea A una matriz tal que $f(A)$ tiene sentido. En virtud del teorema 1 podemos hallar un polinomio $p(\lambda)$ tal que $p(A)=f(A)$. Dada la función $f(\lambda)$, el polinomio $p(\lambda)$ depende sólo de los divisores elementales de la matriz A. Pero los divisores elementales de la matriz A y de la matriz transpuesta A' coinciden y, por ello, se tiene $p(A')=f(A')$. Es fácil deducir del punto 1.3 que $p(A')=p(A)'$. Es decir, *para toda función escalar $f(A)$ tenemos $f(A')=f(A)'$*

16.4. Divisores elementales de funciones. Estudiemos el problema de cómo determinar a partir de los divisores elementales de una matriz A los divisores elementales de una de sus funciones escalares $f(A)$. Reduzcamos A a la forma normal

$$T^{-1}AT=B=B_1\dotplus B_2\dotplus\ldots\dotplus B_t, \tag{11}$$

donde $B_1, \ldots, B_t$ son células de Jordan. Por definición,

$$f(A)=Tf(B)\,T^{-1}$$

y, por consiguiente, los divisores elementales de las matrices $f(A)$ y $f(B)$ coinciden. De (11) se desprende que

$$f(B)=f(B_1)\dotplus f(B_2)\dotplus\ldots\dotplus f(B_t);$$

[1] Notemos una vez más que, según los razonamientos expuestos en el texto, todo valor $f(A)$ de una función escalar f dada se puede representar mediante un polinomio $p(A)$. Sin embargo, para una misma función f este polinomio será diferente, según sean diferentes las matrices A.

luego, el sistema de los divisores elementales de la matriz $f(B)$ es la unión de los sistemas de los divisores elementales de las células $f(B_1), \ldots, f(B_t)$. Es decir, nuestro problema inicial se reduce al siguiente: dada una célula de Jordan B_i de divisor elemental $(\lambda-\rho_i)^{n_i}$ hallar los divisores elementales de $f(B_i)$.

En virtud de las fórmulas (5) y (6) tenemos

$$\lambda E_i-f(B_i)=\begin{bmatrix} \lambda-f(\rho_i) & -f'(\rho_i) & \ldots & -\frac{1}{(n_i-1)!}f^{(n_i-1)}(\rho_i) \\ & \lambda-f(\rho_i) & \ldots & -\frac{1}{(n_i-2)!}f^{(n_i-2)}(\rho_i) \\ & & \ldots & \ldots \\ & & & \lambda-f(\rho_i) \end{bmatrix}. \quad (12)$$

Determinemos los máximos comunes divisores $D_1(\lambda), D_2(\lambda), \ldots$ $\ldots, D_{n_i}(\lambda)$ de los menores de primero, segundo, $\ldots$, n_i-ésimo orden de esta matriz. El mayor de ellos $D_{n_i}(\lambda)$ es igual al determinante de la matriz, es decir,

$$D_{n_i}(\lambda)=(\lambda-f(\rho_i))^{n_i}.$$

Todos los demás son divisores de $D_{n_i}(\lambda)$ y, por consiguiente, son de la forma $(\lambda-f(\rho_i))^{\alpha}$. Consideremos $D_{n_i-1}(\lambda)$. Este polinomio debe ser un divisor de todos los menores de orden n_i-1 de la matriz (12) y, en particular, del menor $\Delta(\lambda)$ que se obtiene suprimiendo la primera columna y la última fila. Sin embargo, si en este menor se introduce en lugar de λ el número $f(\rho_i)$, se obtiene una matriz de forma triangular y con los elementos $-f'(\rho_i)$ a lo largo de la diagonal principal y, por consiguiente,

$$\Delta(\rho_i)=(-f'(\rho_i))^{n_i-1}. \quad (13)$$

Supongamos ahora que $f'(\rho_i)\neq 0$. La igualdad (13) muestra entonces que $\Delta(\lambda)$ no es divisible por $\lambda-f(\rho_i)$. Pero el polinomio $D_{n_i-1}(\lambda)$ debe ser un divisor común de los polinomios $\Delta(\lambda)$ y $D_{n_i}(\lambda)$, es decir, $D_{n_i-1}(\lambda)=1$. Los demás polinomios $D_{n_i-2}(\lambda), \ldots D_2(\lambda)$, $D_1(\lambda)$ son divisores de $D_{n_i-1}(\lambda)$ y, por ello, también son iguales a la unidad. Calculando los cocientes $D_{k+1}:D_k$ vemos que los factores invariantes de la matriz (12) serán $1, \ldots, 1, (\lambda-f(\rho_i))^{n_i}$, debido a lo cual la matriz (12) tendrá sólo un divisor elemental $(\lambda-f(\rho_i))^{n_i}$. De aquí se desprende el teorema siguiente:

TEOREMA 2. *Sea A una matriz de valores propios $\rho_1, \ldots, \rho_s$ y sea $f(\lambda)$ una función tal que $f'(\rho_i)\neq 0$ $(i=1, \ldots, s)$. Entonces, si la matriz $f(A)$ existe, sus divisores elementales se pueden obtener sustituyendo cada uno de los divisores elementales $(\lambda-\rho_i)^{n_i}$ de la matriz A por la expresión $(\lambda-f(\rho_i))^{n_i}$.*

Por ejemplo, si A es una matriz regular y $f(\lambda)=\lambda^{-1}$, se tiene $f(A)=A^{-1}$ y $f'(\rho_i)=-\rho_i^{-2}\neq 0$. Luego, si sustituimos todo divisor elemental $(\lambda-\rho_i)^{n_i}$ de la matriz A por la expresión $(\lambda-\rho_i^{-1})^{n_i}$, obtendremos el sistema de divisores elementales de la matriz inversa.

El teorema 2 permite determinar los divisores elementales de la matriz $f(A)$ pertenecientes a aquellos valores propios $f(\rho_i)$ para los cuales $f'(\rho_i)\neq 0$. No es difícil obtener la regla correspondiente también para el caso en que $f'(\rho_i)=0$. Supongamos que para un valor propio de una matriz A se tiene

$$f'(\rho_i)=f''(\rho_i)=\ldots=f^{(k-1)}(\rho_i)=0 \text{ y } f^{(k)}(\rho_i)\neq 0. \tag{14}$$

Queremos determinar en qué divisores elementales de la matriz $f(A)$ se transforma cada uno de los divisores elementales $(\lambda-\rho_i)^{n_i}$ de la matriz A. Es obvio que este problema se reduce al siguiente: determinar los divisores elementales de la matriz (6) en la condición (14). Para $k\geqslant n_i$ la matriz (6) resulta diagonal y sus divisores elementales serán $\lambda-f(\rho_i), \ldots, \lambda-f(\rho_i)$. El caso $k=1$ ha sido examinado anteriormente: en este caso la matriz (12) tiene un único divisor elemental $(\lambda-f(\rho_i))^{n_i}$. Por esto sólo nos interesarán los valores de k comprendidos entre 1 y n_i.

Consideremos un espacio lineal auxiliar $\mathfrak{L}$ de dimensión n_i y de base $a_1, a_2, \ldots, a_{n_i}$. Pongamos para abreviar $\frac{1}{j!}f^{(j)}(\rho_i)=\alpha_{j+1}$ y $n_i=n$ e indiquemos por $\mathscr{C}$ la aplicación lineal del espacio $\mathfrak{L}$ que tiene la matrtz

$$C=f(B_i)-f(\rho_i)E_i.$$

Tenemos, por consiguiente,

$$\left.\begin{array}{ll}
a_1\mathscr{C} & =\alpha_{k+1}a_{k+1}+\alpha_{k+2}a_{k+2}+\ldots+\alpha_n a_n, \\
a_2\mathscr{C} & = \qquad\qquad \alpha_{k+1}a_{k+2}+\ldots+\alpha_{n-1}a_n, \\
\cdots & \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{n-k}\mathscr{C} & = \qquad\qquad\qquad\qquad \alpha_{k+1}a_n, \\
a_j\mathscr{C} & =o \qquad\qquad\qquad\qquad (j>n-k).
\end{array}\right\} \tag{15}$$

Puesto que nos interesan los divisores elemantales de la aplicación $\mathscr{C}$, tomaremos en $\mathfrak{L}$ otra base en la que la matriz de la aplicación $\mathscr{C}$ tenga una forma más sencilla. Sea

$$e_i=\beta_{ii}a_i+\beta_{i,\,i+1}a_{i+1}+\ldots+\beta_{in}a_n \quad (i=1, \ldots, n);$$

entonces

$$e_i\mathscr{C}=\beta_{ii}\alpha_{k+1}a_{i+k}+(\beta_{ii}\alpha_{k+2}+\beta_{i,\,i+1}\alpha_{k+1})\,a_{i+k+1}+\ldots$$

Escojamos los números β_{ij} de modo que se cumplan las relaciones

$$e_i\mathscr{C}=e_{i+k},\ e_j\mathscr{C}=o \quad (i=1, \ldots, n-k;\ j>n-k). \tag{16}$$

Esto ofrece el siguiente sistema de ecuaciones respecto de β_{ij}:

$$\alpha_{k+1}\beta_{ii}=\beta_{i+k,\,i+k},\ \alpha_{k+1}\beta_{i,\,i+1}+\alpha_{k+2}\beta_{ii}=\beta_{i+k,\,i+k+1}, \ldots$$

Teniendo en cuenta la condición $\alpha_{k+1}\neq 0$, de aquí se pueden determinar sucesivamente los valores $\beta_{ii}, \beta_{i,\,i+1}, \ldots$, expresándolos en términos de los valores de β con mayor primer índice. Para los valores β_{ij}, donde $i\geqslant n-k$, no se obtiene ninguna ecuación, de modo que estos valores se pueden escoger arbitrariamente. En particular, tomando $\beta_{n-k,\,n-k}=\ldots=\beta_{nn}=1$, obtendremos para los demás coeficientes iniciales $\beta_{11}, \ldots, \beta_{n-k+1,\,n-k+1}$ valores nulos. Las ecuaciones (15) pueden ser resueltas entonces respecto de $a_1, a_2, \ldots, a_n$; por esto el sistema $e_1, e_2, \ldots, e_n$ será una base nueva de $\mathfrak{L}$ que satisface las condiciones (16). Descompongamos los vectores $e_1, \ldots, e_n$ en los sistemas

$$\begin{array}{llll} e_1, & e_{1+k}, & e_{1+2k}, & \ldots \\ e_2, & e_{2+k}, & e_{2+2k}, & \ldots \\ \cdot & \cdot & \cdot & \cdot \\ e_k, & e_{2k}, & e_{3k}, & \ldots \end{array}$$

y consideremos los subespacios $\mathfrak{L}_1, \ldots, \mathfrak{L}_k$, tendidos sobre estos sistemas. Las relaciones (16) muestran que $\mathfrak{L}_i$ es un subespacio invariante de base $e_i, e_{i+k}, e_{i+2k}, \ldots$ y que

$$\mathfrak{L}=\mathfrak{L}_1\dot{+}\mathfrak{L}_2\dot{+}\ldots\dot{+}\mathfrak{L}_k. \tag{17}$$

En virtud de los resultados del p. 12.3, de aquí se desprende que la matriz de la aplicación $\mathscr{C}$ se descompone en k células de Jordan de divisores elementales $\lambda^{m_1}, \lambda^{m_2}, \ldots, \lambda^{m_k}$, donde $m_1, m_2, \ldots, m_k$ son las dimensiones de los subespacios $\mathfrak{L}_1, \mathfrak{L}_2, \ldots, \mathfrak{L}_k$. La matriz de la aplicación inicial $f(B_i)$ está ligada a la matriz C mediante la fórmula $f(B_i)=f(\rho_i)E_i+C$ y, por lo tanto, los divisores elementales de la matriz $f(B_i)$ son $(\lambda-f(\rho_i))^{m_1}, \ldots, (\lambda-f(\rho_i))^{m_k}$. Si empleamos el símbolo $[\alpha]$ para indicar el mayor número entero que no sobrepasa a α, obtendremos para $m_1, \ldots, m_k$ las expresiones siguientes:

$$m_1=\left[\frac{n-1}{k}\right]+1,\ m_2=\left[\frac{n-2}{k}\right]+1,\ \ldots,\ m_k=\left[\frac{n-k}{k}\right]+1=\left[\frac{n}{k}\right].$$

Por consiguiente, *si la matriz A tiene un divisor elemental $(\lambda-\rho)^m$ y si $f'(\rho)=\ldots=f^{(k-1)}(\rho)=0$ y $f^{(k)}(\rho)\neq 0$, entonces al pasar de A a $f(A)$ este divisor elemental se descompone en los divisores elementales $(\lambda-f(\rho))^{m_1}, \ldots, (\lambda-f(\rho))^{m_k}$, donde*

$$m_1=\left[\frac{m-1}{k}\right]+1,\ m_2=\left[\frac{m-2}{k}\right]+1,\ \ldots,\ m_k=\left[\frac{m-k}{k}\right]+1.$$

Consideremos un ejemplo. Supongamos que A tiene los divisores elementales $(\lambda-1)^8$ y $(\lambda+2)^2$; es preciso determinar los divisores elementales de la matriz $A^6-3A^4+3A^2-E$. En estas condiciones tenemos $f(\lambda)=\lambda^6-3\lambda^4+3\lambda^2-1=(\lambda-1)^3(\lambda+1)^3$, $f'(-2)\neq 0$, $f'(1)=f''(1)=0$ y $f'''(1)\neq 0$. Al pasar a $f(A)$ el divisor elemental $(\lambda+2)^2$ se convierte en $(\lambda-f(-2))^2=(\lambda-27)^2$, mientras que el divisor elemental $(\lambda-1)^8$ se descompone en λ^3, λ^3 y λ^2. Por consiguiente, los divisores elementales de la matriz $f(A)$ serán λ^3, λ^3, λ^2 y $(\lambda-27)^2$.

16.5. Series de potencias. Una sucesión de matrices cuadradas

$$A_1,\ A_2,\ \ldots,\ A_m,\ A_{m+1},\ \ldots \tag{18}$$

de un mismo orden se llama *convergente* hacia la matriz A, si los elementos de las matrices (18) que aparecen en la intersección de una columna y de una fila dadas convergen hacia el elemento correspondiente de la matriz A. De esta definición se desprende directamente que si las matrices A_m y B_m convergen con el crecimiento de m hacia A y B, respectivamente, las matrices A_m+B_m y A_mB_m convergen hacia $A+B$ y AB. En particular, si T es una matriz constante y la matriz A_m converge hacia A, resulta que $T^{-1}A_mT$ tendrá como límite a la matriz $T^{-1}AT$. Además, si

$$A_m=A_m^{(1)}\dot{+}A_m^{(2)}\dot{+}\ldots\dot{+}A_m^{(s)}\quad (m=1,\ 2,\ \ldots),$$

donde los ordenes de las células no dependen de m, la matriz A_m converge con el crecimiento de m hacia un límite determinado si, y sólo si, cada una de las células $A_m^{(i)}$ converge hacia un límite.

La última observación permite resolver fácilmente el problema de convergencia de las así llamadas series de potencias de matrices. Sea

$$\alpha_0+\alpha_1\lambda+\alpha_2\lambda^2+\ldots+\alpha_m\lambda^m+\ldots \tag{19}$$

una serie formal respecto de la variable λ. La expresión

$$\alpha_0E+\alpha_1A+\alpha_2A^2+\ldots+\alpha_mA^m+\ldots \tag{20}$$

se llama *serie de potencias* correspondiente de la matriz A y el polinomio

$$f_n(A)=\alpha_0E+\alpha_1A+\ldots+\alpha_nA^n$$

se llama *n-ésima suma inicial* de esta serie. Se dice que la serie (20) *converge*, si la sucesión de las sumas iniciales $f_1(A), \ldots, f_m(A), \ldots$ tiene límite; en el caso de su existencia este límite se denomina *suma* de la serie (20).

Reduzcamos la matriz A a la forma normal

$$T^{-1}AT=B=B_1\dotplus B_2\dotplus\ldots\dotplus B_t,$$

donde $B_1, \ldots, B_t$ son células de Jordan. Hemos visto más arriba que la convergencia de la sucesión $f_m(A)$ equivale a la convergencia de la sucesión $T^{-1}f_m(A)T$ $(m=1, 2, \ldots)$. Pero

$$T^{-1}f_m(A)T=f_m(T^{-1}AT)=f_m(B)=f_m(B_1)\dotplus\ldots\dotplus f_m(B_t),$$

es decir, el problema acerca de la convergencia de la serie (20) equivale al siguiente: ¿bajo qué condiciones converge esta serie para las células de Jordan $B_1, \ldots, B_t$? Consideremos una de estas células, por ejemplo, la célula B_i. Sea $(\lambda-\rho_i)^{n_i}$ el divisor elemental que le corresponde. Según la fórmula (3) tenemos

$$f_m(B_i)=\begin{bmatrix} f_m(\rho_i) & \frac{1}{1!}f_m'(\rho_i) & \ldots & \frac{1}{(n_i-1)!}f_m^{(n_i-1)}(\rho_i) \\ & f_m(\rho_i) & \ldots & \frac{1}{(n_i-2)!}f_m^{(n_i-2)}(\rho_i), \\ \ldots & \ldots & \ldots & \ldots \\ & & & f_m(\rho_i) \end{bmatrix},$$

por consiguiente, $f_m(B_i)$ converge hacia un límite con el crecimiento de m si, y sólo si, $f_m(\rho_i), f_m'(\rho_i), \ldots, f_m^{(n_i-1)}(\rho_i)$ convergen hacia unos límites, es decir, si en el punto ρ_i convergen tanto la serie (19) como las series que se obtienen de ella derivándola término por término n_i-1 veces sucesivas. De la teoría de las funciones analíticas se sabe que todas estas series convergen indudablemente, si ρ_i pertenece al interior del círculo de convergencia de la serie (19) o si ρ_i pertenece a la circunferencia del círculo de convergencia y la (n_i-1)-ésima derivada de la serie (19) converge en el punto ρ_i. Es decir, hemos demostrado el teorema siguiente:

TEOREMA 3. *Para que una serie de potencias de una matriz A converja es necesario y suficiente que todo valor propio ρ_i de la matriz A se halle en el interior del círculo de convergencia de la correspondiente serie de potencias $f(\lambda)$ o que se halle sobre la circunferencia del círculo de convergencia, pero con la particularidad de que la serie, que se obtiene derivando n_i-1 veces la serie de $f(\lambda)$, converja en el punto ρ_i, donde n_i es el grado del mayor divisor elemental correspondiente a ρ_i.*

16.6. Matrices conmutables con una matriz dada. Dos matrices A y B se llaman *conmutables*, si $AB=BA$. Toda matriz es conmutable con sí misma y con la matriz unidad. Además, si A es conmutable con las matrices B y C, las igualdades

$$A \cdot BC = BAC = BC \cdot A,$$
$$A(\alpha B + \beta C) = \alpha AB + \beta AC = \alpha BA + \beta CA = (\alpha B + \beta C) A$$

muestran que A es conmutable con el producto y las combinaciones lineales de las mismas. Luego, si B es conmutable con A, resulta que B es conmutable con cualquier polinomio en A. En particular, A es conmutable con los polinomios en A y cualesquiera dos polinomios en A son conmutables.

Según el p. 16.3, los valores de las funciones escalares de una matriz A pueden ser representados mediante polinomios en A. Los polinomios en A son conmutables con toda matriz B que sea conmutable con A. Por consiguiente, *el valor de toda función escalar de una matriz A es conmutable con todas las matrices que son conmutables con A.*

La relación $\alpha E \cdot P = P \cdot \alpha E$ muestra que *las matrices αE son conmutables con todas las matrices del mismo orden.* La recíproca también es válida:

Si una matriz cuadrada A de orden n es conmutable con todas las matrices de orden n, la matriz A es de la forma αE.

Indiquemos por α_{ij} los elementos de la matriz A. Sea P la matriz que contiene la unidad en la p-ésima fila y en la q-ésima columna, mientras que todas sus demás posiciones están ocupadas por ceros. Realizando la multiplicación directa obtenemos

$$AP = \begin{bmatrix} 0 & \dots & \alpha_{1p} & \dots & 0 \\ 0 & \dots & \alpha_{2p} & \dots & 0 \\ . & . & . & . & . \\ 0 & \dots & \alpha_{np} & \dots & 0 \end{bmatrix} \quad \text{y} \quad PA = \begin{bmatrix} 0 & 0 & \dots & 0 \\ . & . & . & . \\ \alpha_{q1} & \alpha_{q2} & \dots & \alpha_{qn} \\ . & . & . & . \\ 0 & 0 & \dots & 0 \end{bmatrix},$$

donde aparecen escritas la q-ésima columna de la primera matriz y la p-ésima fila de la segunda. Como, por hipótesis, $AP = PA$, resulte que $\alpha_{pp} = \alpha_{qq}$ y que $\alpha_{pq} = 0$, si $p \neq q$. Puesto que p y q son arbitrarios, esto significa que A es diagonal y que todos sus elementos diagonales son iguales.

Consideremos ahora un problema más complejo: *hallar todas las matrices conmutables con una matriz dada A.*

Para resolverlo, reduciremos A a la forma normal de Jordan

$$T^{-1}AT = B = B_1 \dot{+} B_2 \dot{+} \dots \dot{+} B_s, \tag{21}$$

donde $B_1, \dots, B_s$ son células de Jordan. Si la matriz X es conmutable con B, es obvio que $Y = TXT^{-1}$ es conmutable con A y, recíprocamente, si Y conmuta con A, también $T^{-1}YT = X$ conmuta con B. Por lo tanto, todo el problema se reduce a determinar las matrices X que conmutan con la matriz B que tiene la forma normal (21). De acuerdo con (21) descompongamos en células también la matriz X:

$$X=\begin{vmatrix} X_{11} & X_{12} & \dots & X_{1s} \\ X_{21} & X_{22} & \dots & X_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ X_{s1} & X_{s2} & \dots & X_{ss} \end{vmatrix}.$$

La condición $BX=XB$ lleva a las igualdades

$$B_pX_{pq}=X_{pq}B_q \quad (p,\ q=1,\ \dots,\ s). \tag{22}$$

Vemos que para cada una de las células X_{pq} se obtiene una sola igualdad (22) de la cual debemos determinar los elementos de X_{pq}.

Supongamos que los órdenes de las matrices B_p y B_q son k y m, respectivamente, y que sus valores propios son ρ y σ. En este caso X_{pq} será una matriz rectangular de k filas y de m columnas. Indiquemos por ξ_{ij} ($i=1,\ \dots,\ k$; $j=1,\ \dots,\ m$) los elementos de la matriz X_{pq} y representemos la relación (22) en forma detallada

$$\begin{bmatrix} \rho & 1 & 0 & \dots & 0 \\ & \rho & 1 & \dots & 0 \\ & & & \dots & \cdot \\ & & & & \rho \end{bmatrix} \begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \xi_{k1} & \xi_{k2} & \dots & \xi_{km} \end{bmatrix} = \begin{bmatrix} \xi_{11} & \xi_{12} & \dots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \dots & \xi_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \xi_{k1} & \xi_{k2} & \dots & \xi_{km} \end{bmatrix} \begin{bmatrix} \sigma & 1 & 0 & \dots & 0 \\ & \sigma & 1 & \dots & 0 \\ & & & \dots & \cdot \\ & & & & \sigma \end{bmatrix}.$$

Realizando aquí la multiplicación y comparando el elemento que se obtiene en la i-ésima fila y j-ésima columna del primer miembro con el elemento correspondiente del segundo miembro, llegamos a las ecuaciones

$$\rho\xi_{ij}+\xi_{i+1,\,j}=\xi_{i,\,j-1}+\sigma\xi_{ij} \quad (i\neq k,\ j\neq 1), \tag{23}$$
$$\rho\xi_{kj}=\xi_{k,\,j-1}+\sigma\xi_{kj} \quad (i=k,\ j\neq 1), \tag{24}$$
$$\rho\xi_{k1}=\sigma\xi_{k1} \quad (i=k,\ j=1). \tag{25}$$

Si $\rho\neq\sigma$, de (25) se deduce que $\xi_{k1}=0$; entonces de (24) obtenemos sucesivamente $\xi_{k2}=\dots=\xi_{km}=0$ y de (23) concluimos que todos los demás ξ_{ij} son también iguales a cero. Por consiguiente, para $\rho\neq\sigma$ tenemos $X_{pq}=O$.

Consideremos el caso $\rho=\sigma$. Las ecuaciones (23), (24) y (25) se convierten entonces en

$$\xi_{i+1,\,j}=\xi_{i,\,j-1} \quad (i=1,\ \dots,\ k-1;\ j=2,\ \dots,\ m), \tag{26}$$
$$\xi_{k,\,j-1}=0 \quad (j=2,\ \dots,\ m). \tag{27}$$

Sea $k\geqslant m$; poniendo $\xi_{11}=\xi_1,\ \xi_{12}=\xi_2,\ \dots,\ \xi_{1m}=\xi_m$, reduciremos las ecuaciones (26) y (27) a las ecuaciones

$$\xi_{ij}=\xi_{j-i+1} \quad (i\leqslant j),$$
$$\xi_{ij}=0 \quad (i>j),$$

de donde resulta que la matriz X_{pq} tiene la forma *triangular lineal*

$$X_{pq}=\begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \dots & \xi_m \\ & \xi_1 & \xi_2 & \dots & \xi_{m-1} \\ & & \dots & \dots & \dots \\ & & & & \xi_1 \end{bmatrix} \tag{28}$$

para $k=m$ y, respectivamente, la forma

$$X_{pq}=\begin{bmatrix} \xi_1 & \xi_2 & \xi_3 & \dots & \xi_m \\ & \xi_1 & \xi_2 & \dots & \xi_{m-1} \\ & & \dots & \dots & \dots \\ & & & & \xi_1 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

para $k>m$. Siendo $k<m$ y tomando $\xi_{1,\,m-k+1}=\xi_1$, $\xi_{1,\,m-k+2}=\xi_2$, ..., $\xi_{1m}=\xi_k$, reduciremos las ecuaciones (23), (24) y (25) a las ecuaciones

$$\begin{aligned} &\xi_{ij}=\xi_{j-i-m+k+1} && (j-i\geqslant m-k), \\ &\xi_{ij}=0 && (j-i<m-k), \end{aligned}$$

que significan que la matriz X_{pq} es de la forma

$$X_{pq}=\begin{bmatrix} 0 & \dots & 0 & \xi_1 & \xi_2 & \dots & \xi_k \\ 0 & \dots & 0 & 0 & \xi_1 & \dots & \xi_{k-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & \xi_1 \end{bmatrix}.$$

Recíprocamente, si las células de la matriz X tienen la forma señalada, las ecuaciones (23), (24) y (25) se satisfacen y, por consiguiente, X conmuta con B.

Por ejemplo, si

$$B=\begin{bmatrix} \rho & 1 & 0 \\ & \rho & 1 \\ & & \rho \end{bmatrix} \dotplus \begin{bmatrix} \rho & 1 \\ & \rho \end{bmatrix} \dotplus \begin{bmatrix} \sigma & 1 & 0 \\ & \sigma & 1 \\ & & \sigma \end{bmatrix} \quad (\rho\neq\sigma), \tag{29}$$

las matrices que conmutan con B son de la forma

$$X=\left[\begin{array}{ccc|cc} \alpha_0 & \alpha_1 & \alpha_2 & \gamma_0 & \gamma_1 \\ & \alpha_0 & \alpha_1 & 0 & \gamma_0 \\ & & \alpha_0 & 0 & 0 \\ \hline 0 & \delta_0 & \delta_1 & \beta_0 & \beta_1 \\ 0 & 0 & \delta_0 & 0 & \beta_0 \end{array}\right] \dotplus \begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 \\ & \lambda_0 & \lambda_1 \\ & & \lambda_0 \end{bmatrix},$$

donde α_i, β_i, γ_i y δ_i son números arbitrarios.

Este resultado se hace sumamente sencillo para matrices de tipo $\rho_1E_1\dotplus\rho_2E_2\dotplus\dots\dotplus\rho_sE_s$, donde $E_1, \dots, E_s$ son matrices unidades y todos los números $\rho_1, \ldots, \rho_s$ son diferentes. En este caso las células X_{pq} son nulas para $p \neq q$ y X adquiere la forma celular diagonal.

En particular, si B es diagonal y sus elementos diagonales son distintos, las matrices conmutables con B son también diagonales.

16.7. Matrices que conmutan con matrices conmutables. Los polinomios en una matriz A poseen la propiedad específica de que, además de conmutar con la propia matriz A, conmutan con cualquier matriz X que es conmutable con A. Resulta que esta propiedad es característica para los polinomios en A.

TEOREMA 4. *Si la matriz C conmuta con todas las matrices que conmutan con B, C es un polinomio en B.*

Es obvio que basta realizar la demostración para el caso en que B tiene la forma normal de Jordan; sea, pues,

$$B = B_1 \dotplus B_2 \dotplus \ldots \dotplus B_s,$$

donde $B_1, \ldots, B_s$ son células de Jordan. Las matrices auxiliares

$$X = \alpha_1 E_1 \dotplus \alpha_2 E_2 \dotplus \ldots \dotplus \alpha_s E_s,$$

donde $\alpha_1, \ldots, \alpha_s$ son números arbitrarios y $E_1, \ldots, E_s$ son matrices unidades, conmutan indudablemente con B y, por ello, las matrices X conmutan también con C. De aquí se deduce, de acuerdo con lo expuesto anteriormente, que C se descompone en células:

$$C = C_1 \dotplus C_2 \dotplus \ldots \dotplus C_s; \tag{30}$$

además de la condición $CB = BC$ resulta que estas células tienen la forma lineal triangular (28). Sea ahora X una matriz arbitraria que conmuta con B. La forma general de la matriz X ha sido determinada en el punto anterior. Según las condiciones del teorema que estamos demostrando, la matriz C debe conmutar con X. Representando X en la forma celular veremos que la igualdad $CX = XC$ equivale a las relaciones

$$C_p X_{pq} = X_{pq} C_q \qquad (p, q = 1, \ldots, s). \tag{31}$$

Si las células correspondientes B_p y B_q tienen diferentes valores propios, nada resulta de (31) ya que en este caso $X_{pq} = O$. Por esto aceptaremos que los valores propios de las células B_p y B_q coinciden. Sea

$$C_p = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_k \\ & \alpha_1 & \ldots & \alpha_{k-1} \\ & & \ldots & \ldots \\ & & & \alpha_1 \end{bmatrix} \quad \text{y} \quad C_q = \begin{bmatrix} \beta_1 & \beta_2 & \ldots & \beta_m \\ & \beta_1 & \ldots & \beta_{m-1} \\ & & \ldots & \ldots \\ & & & \beta_1 \end{bmatrix}.$$

Supongamos, para concretar, que $k < m$. Entonces la relación (31) se convierte en

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_k \\ & \alpha_1 & \dots & \alpha_{k-1} \\ & \dots & \dots & \dots \\ & & & \alpha_1 \end{bmatrix} \begin{bmatrix} 0 & \dots & 0 & \xi_1 & \xi_2 & \dots & \xi_k \\ 0 & \dots & 0 & 0 & \xi_1 & \dots & \xi_{k-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & \xi_1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & \dots & 0 & \xi_1 & \xi_2 & \dots & \xi_k \\ 0 & \dots & 0 & 0 & \xi_1 & \dots & \xi_{k-1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & 0 & \dots & \xi_1 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_m \\ & \beta_1 & \dots & \beta_{m-1} \\ & \dots & \dots & \dots \\ & & & \beta_1 \end{bmatrix},$$

donde $\xi_1, \xi_2, \dots, \xi_m$ son números arbitrarios. Multiplicando en el primer y en el segundo miembros la primera fila por la última columna e igualando los resultados, obtenemos

$$\alpha_1\xi_k + \alpha_2\xi_{k-1} + \dots + \alpha_k\xi_1 = \beta_1\xi_k + \beta_2\xi_{k-1} + \dots + \beta_k\xi_1,$$

de donde, debido a la arbitrariedad de los números $\xi_1, \dots, \xi_k$ resulta que

$$\alpha_1 = \beta_1, \ \alpha_2 = \beta_2, \ \dots, \ \alpha_k = \beta_k. \tag{32}$$

Es fácil comprobar que semejantes igualdades se obtienen también en el caso en que $k \geqslant m$. Las igualdades (30), (31) y (32) constituyen un sistema completo de condiciones a las que debe someterse la matriz incógnita C. Para hacer estas condiciones más claras, procederemos del modo siguiente. Coloquemos las células de Jordan de la matriz B de manera que las células con los valores propios iguales estén al lado una de otra. Sea, por ejemplo,

$$B = (B_1 \dotplus \dots \dotplus B_{m_1}) \dotplus (B_{m_1+1} \dotplus \dots \dotplus B_{m_2}) \dotplus \dots$$
$$\dots \dotplus (B_{m_t+1} \dotplus \dots \dotplus B_s),$$

donde en cada uno de los paréntesis figuran células con valores propios iguales. Indicando las sumas de estos paréntesis por $B^{(1)}, \dots B^{t+1}$, la matriz B quedará dividida en células mayores que denominaremos bloques. De acuerdo con esto también las matrices X y C quedarán divididas en bloques respectivos. Los resultados del punto anterior muestran que todos los bloques no diagonales de la matriz X son iguales a cero; en cuanto a los bloques diagonales de X, éstos tienen la estructura descrita en aquel punto. Las condiciones obtenidas en el punto presente para la matriz C muestran que sus bloques diagonales también se descomponen, igual que en el caso de B, en células, con la particularidad de que las células de la matriz C tienen una forma triangular especial. Las igualdades (32) significan que en las células de la matriz C, pertenecientes a un mismo bloque, los elementos que figuran en una misma línea paralela a la diagonal principal son iguales.

Sea, por ejemplo, B la matriz (29) considerada a título de ejemplo al final del punto anterior. Nuestros resultados muestran que la matriz C que conmuta con toda matriz conmutable con B será de la forma

$$C=\left[\begin{array}{ccccc|ccc} \alpha_0 & \alpha_1 & \alpha_2 & & & & & \\ & \alpha_0 & \alpha_1 & & & & & \\ & & \alpha_0 & & & & & \\ & & & \alpha_0 & \alpha_1 & & & \\ & & & & \alpha_0 & & & \\ \hline & & & & & \lambda_0 & \lambda_1 & \lambda_2 \\ & & & & & & \lambda_0 & \lambda_1 \\ & & & & & & & \lambda_0 \end{array}\right]. \qquad (33)$$

Resta probar que C se puede representar en forma de un polinomio en B. Haremos esto para el caso particular en que B es de la forma (29) y, por consiguiente, C tiene la forma (33); en el caso general los razonamientos son los mismos. Es decir, debemos probar que la matriz (33) es un polinomio en la matriz (29). Busquemos un polinomio $f(\lambda)$ que satisfaga las condiciones

$$f(\rho)=\alpha_0, \quad f'(\rho)=\alpha_1, \quad f''(\rho)=\alpha_2,$$
$$f(\sigma)=\lambda_0, \quad f'(\sigma)=\lambda_1, \quad f''(\sigma)=\lambda_2.$$

Según el p. 16.3 existe un polinomio de esta índole. Aplicando la regla (3) del p. 16.1 veremos que $f(B)=C$.

Ejemplos y problemas

1. Hállese A^n, si

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

2. Hállense *todas* las soluciones de la ecuación $X^2=A$, donde A tiene el mismo valor que en el problema anterior. ¿Cuáles de estas soluciones serán polinomios en A? Calcúlense sen πA, e^A y $\cos \pi A$.

3. Los divisores elementales de una aplicación $\mathcal{A}$ son iguales a λ^5, $(\lambda-\pi)^3$ y $\lambda+\frac{\pi}{2}$. Calcúlense los divisores elementales de las aplicaciones $\cos\mathcal{A}$ y sen $\mathcal{A}$.

4. Si A y B son matrices que conmutan, se tiene $e^Ae^B=e^{A+B}$.

5. ¿Para qué matrices A tiene solución la ecuación $A=e^X$?

6. Si $f(\lambda)$ indica la suma de la serie de potencias (19), la suma de la serie de potencias (20), si es que esta serie converge, será igual a $f(A)$.

7. Hállese la forma general de la matriz que conmuta con una matriz normal de Jordan A, si los divisores elementales de A son $(\lambda-1)^3$, $(\lambda-1)^2$, $\lambda-1$, $(\lambda+2)^2$ y $\lambda+2$.

8. Supongamos que todos los factores invariantes de la matriz $\lambda E-A$, salvo el último, son iguales a 1. Entonces toda matriz conmutable con A es un polinomio en A.

Capítulo V

Espacios unitarios y euclídeos

Los espacios lineales que hemos estudiado en los capítulos anteriores han resultado ser, en determinado sentido, más pobres en conceptos y propiedades que nuestro espacio corriente. En la teoría general de los espacios lineales no han quedado reflejados conceptos como la longitud de un segmento, la magnitud del ángulo y el producto escalar que desempeñan un papel primordial en la geometría. Por esto, si queremos que la teoría general abarque todas las propiedades más esenciales del espacio corriente, debemos introducir, además de las operaciones de adición de vectores y de multiplicación de los mismos por números, la operación de multiplicación escalar. En este capítulo se estudian precisamente las propiedades de los vectores pertenecientes a espacios provistos del producto escalar.

En este capítulo el cuerpo principal es de carácter muy especial: es el cuerpo de los números reales en el caso de espacios euclídeos y es el cuerpo de los números complejos en el caso de espacios unitarios.

§ 17. Espacios unitarios

17.1. Axiomática y ejemplos. Sea $\mathfrak{R}$ nuestro espacio corriente cuyos vectores son los segmentos orientados que parten de un punto inicial O. Se llama producto escalar (a, b) de los vectores a y b el producto de las longitudes de a y de b por el coseno del ángulo que forman estos vectores. De aquí se desprenden directamente las conocidas propiedades del producto escalar:

(a) $(a, b)=(b, a)$;
(b) $(\alpha a, b)=\alpha(a, b)$;
(c) $(a+b, c)=(a, c)+(b, c)$;
(d) si $a \neq o$, se tiene $(a, a)>0$.

Tomemos en el espacio $\Re$ un sistema de coordenadas formado por tres cualesquiera vectores e_1, e_2 y e_3, perpendiculares dos a dos, de longitud 1. Entonces todo vector a admite una representación única de la forma

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3,$$

donde α_1, α_2 y α_3 son las longitudes de las proyecciones del vector a sobre los ejes coordenados, tomadas con signo adecuado. Si

$$b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$$

es otro vector cualquiera, resulta de la definición del producto escalar y de las propiedades (b) y (c) que

$$(a, b) = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3. \tag{1}$$

El espacio $\Re$ es real. Esto se expresa en que las proyecciones, las longitudes y los productos escalares de los vectores son números reales. Sin embargo, en algunos casos surge la necesidad de considerar vectores de proyecciones complejas. A primera vista parece natural tomar de nuevo la expresión (1) para el producto escalar de vectores con coordenadas complejas α_1, α_2, α_3 y β_1, β_2, β_3. En algunos casos se procede precisamente de este modo. El espacio que así resulta se denomina *espacio euclídeo complejo.* Por desgracia, el producto escalar pierde entonces muchas propiedades importantes y entre ellas la propiedad (d) de importancia primordial. En efecto, para el vector

$$a = 3e_1 + 4e_2 + 5ie_3 \quad (i = \sqrt{-1})$$

de la fórmula (1) resulta

$$(a, a) = 9 + 16 + 25i^2 = 0$$

contrariamente a la propiedad (d). Para evitar este inconveniente, en lugar de la expresión (1) se toma como definición del producto escalar de vectores complejos la expresión

$$(a, b) = \alpha_1\overline{\beta}_1 + \alpha_2\overline{\beta}_2 + \alpha_3\overline{\beta}_3, \tag{2}$$

donde la raya superior significa que ha de pasarse a los números complejos conjugados. En el caso en que los vectores a y b son reales, tenemos $\beta_j = \overline{\beta}_j$ y la expresión (2) coincide con (1). Por consiguiente, la nueva definición (2) es una generalización de la anterior. Por otra parte, con la nueva definición la propiedad (d) se cumple sin duda alguna, ya que de (2) resulta:

$$(a, a) = \alpha_1\overline{\alpha}_1 + \alpha_2\overline{\alpha}_2 + \alpha_3\overline{\alpha}_3 = |\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2,$$

donde $|\alpha_j|$ es el módulo del número α_j. Es fácil ver que las propiedades (b) y (c) también se verifican. En cuanto a la propiedad (a), toma una forma distinta en el caso de vectores complejos. Efectivamente, de (2) se tiene

$$(b, a) = \beta_1\bar{\alpha}_1 + \beta_2\bar{\alpha}_2 + \beta_3\bar{\alpha}_3 = \overline{\bar{\beta}_1\alpha_1 + \bar{\beta}_2\alpha_2 + \bar{\beta}_3\alpha_3},$$

es decir,

$$(b, a) = \overline{(a, b)}. \tag{a*}$$

El espacio de vectores complejos en el que el producto escalar se calcula mediante la fórmula (2) se denomina unitario. La fórmula (a*) muestra que las propiedades del espacio unitario difieran, en general, de las propiedades del espacio corriente. No obstante, estas diferencias son de poca importancia. En todo caso, el espacio unitario se aproxima más por sus propiedades al espacio corriente que el espacio euclídeo complejo mencionado anteriormente.

Los razonamientos que hemos expuesto no pueden calificarse de totalmente precisos. Además, hemos considerado el caso de un espacio de tres dimensiones. Por esto debemos dar ahora una definición totalmente rigurosa de los espacios unitarios que sea válida también para espacios de cualquier dimensión.

En la teoría general de espacios lineales hemos realizado casi toda la exposición aceptando que el cuerpo principal K es totalmente arbitrario. En el capítulo presente K será o bien el cuerpo de todos los números complejos o bien el cuerpo de todos los números reales.

Un espacio lineal $\mathfrak{L}$ sobre un campo K se llama *unitario*, si a todo par de vectores a y b de $\mathfrak{L}$ tomados en un orden determinado corresponde un número de K llamado *producto escalar* (a, b) del vector a por el vector b que posee las propiedades siguientes:

1° $(a, b) = \overline{(b, a)}$;
2° $(\alpha a, b) = \alpha (a, b)$;
3° $(a + b, c) = (a, c) + (b, c)$;
4° *si* $a \neq o$, *se tiene* $(a, a) > 0$.

En el caso en que el campo principal K es el cuerpo de los números reales, el espacio unitario $\mathfrak{L}$ se denomina *espacio unitario real* o simplemente *espacio euclídeo real*. En este caso la expresión $\overline{(a, b)}$ coincide, obviamente, con la expresión (a, b) y el axioma 1° adquiere una forma más sencilla: $(a, b) = (b, a)$.

Si el campo K es el cuerpo de los números complejos, el espacio $\mathfrak{L}$ se llama *espacio complejo unitario*. En lo sucesivo las propiedades de los espacios euclídeos reales y las propiedades de los espacios complejos unitarios serán examinadas, en la mayoría de los casos, conjuntamente y por espacio unitario se comprenderá, de acuerdo con la definición, o bien el espacio unitario real o bien el espacio unitario complejo.

Notemos también que en la definición de los espacios unitarios no se exige que el espacio sea de dimensión finita. Por esto cabe hablar también de espacios unitarios de dimensión infinita. Aun cuando algunas propiedades de los espacios unitarios no dependen de la dimensión de los mismos, nos limitaremos a considerar, mientras que no se diga lo contrario, solamente espacios de dimensión finita. La teoría de espacios de dimensión infinita entra de lleno en una disciplina matemática especial que es el Análisis funcional.

De las propiedades 1° y 2° resulta

$$(\alpha a, \beta b) = \alpha (a, \beta b) = \alpha \overline{(\beta b, a)} = \alpha\bar{\beta}\, \overline{(b, a)} = \alpha\bar{\beta}\, (a, b),$$

es decir

$$(\alpha a, \beta b) = \alpha\bar{\beta}\, (a, b). \tag{3}$$

Análogamente de 1°, 2° y 3° se desprende que

$$(a, b+c) = \overline{(b+c, a)} = \overline{(b, a)} + \overline{(c, a)} = (a, b) + (a, c).$$

De aquí obtenemos, mediante el procedimiento corriente, la fórmula general

$$\left(\sum \alpha_j a_j,\ \sum \beta_k b_k\right) = \sum\sum \alpha_j \bar{\beta}_k (a_j, b_k). \tag{4}$$

Para $\alpha = 1$ y $\beta = 0$ de la relación (3) resulta

$$(a, o) = (o, a) = 0.$$

Señalemos dos ejemplos. Consideremos el espacio lineal de filas de longitud n con elementos del campo K. Convengamos en llamar *producto escalar* de la fila $a = [\alpha_1, \ldots, \alpha_n]$ por la fila $b = [\beta_1, \ldots, \beta_n]$ la expresión

$$(a, b) = \alpha_1\bar{\beta}_1 + \alpha_2\bar{\beta}_2 + \ldots + \alpha_n\bar{\beta}_n. \tag{5}$$

De esta expresión se ve que

$$(a, a) = \alpha_1\bar{\alpha}_1 + \ldots + \alpha_n\bar{\alpha}_n = |\alpha_1|^2 + \ldots + |\alpha_n|^2. \tag{6}$$

Puesto que los módulos $|\alpha_j|$ son números reales no negativos, la suma de sus cuadrados será un número real no negativo que será igual a cero sólo en el caso en que sean iguales a cero todos sus sumandos. Por consiguiente, la propiedad 4° aquí se cumple. Es obvio que las propiedades 1°, 2° y 3° también se cumplen, de modo que el espacio de filas con el producto escalar (5) es un espacio unitario de dimensión n sobre el campo K. Este ejemplo es de una importancia primordial ya que más adelante quedará demostrado que todos los espacios unitarios de dimensión n sobre el campo K son isomorfos.

Como ejemplo de espacio unitario de dimensión infinita puede servir el espacio $\mathfrak{L}$ de todas las funciones continuas $f(t)$ de valores complejos definidas sobre el segmento $[0, 1]$. La adición y la multiplicación por número de estas funciones se definen de modo corriente y el producto escalar de una función $f(t)$ por otra función $g(t)$ se define mediante la fórmula

$$(f, g) = \int_0^1 f(t)\overline{g(t)}\,dt. \tag{7}$$

Las propiedades de 1° a 4° se demuestran fácilmente y, por consiguiente, el espacio $\mathfrak{L}$ es unitario. En este ejemplo el campo principal es el cuerpo de todos los números complejos. Si nos limitamos a considerar solamente las funciones continuas de valores reales, se puede tomar como campo principal el cuerpo de los números reales. La fórmula (7) quedará sustituida entonces por la fórmula

$$(f, g) = \int_0^1 f(t)\,g(t)\,dt.$$

17.2. Longitud de un vector. El cuadrado escalar (a, a) de cualquier vector es, según el axioma 4°, un número real no negativo. El valor no negativo de la raíz cuadrada de este número se denomina *longitud* o *norma* del vector a y se designa por $\|a\|$. Es decir, por definición

$$\|a\| = \sqrt{(a, a)}.$$

De esta definición se ve directamente que *el vector nulo es el único vector cuya longitud es igual a cero.* Además, si α es un número, se tiene

$$\|\alpha a\| = \sqrt{(\alpha a, \alpha a)} = \sqrt{\alpha\bar{\alpha}\,(a, a)} = |\alpha|\,\sqrt{(a, a)}, \tag{8}$$

es decir, *al multiplicar un vector por un número su longitud se multiplica por el módulo de este número.* Un vector cuya longitud es igual a la unidad se llama *vector unidad* o *vector normalizado.* La igualdad (8) muestra que al multiplicar un vector no nulo por el número inverso de su longitud se obtiene un vector unidad. Esta operación se llama a veces *normalización* de un vector.

Pasamos a la demostración de una desigualdad importante que relaciona las longitudes de dos vectores con el valor del producto escalar de los mismos.

DESIGUALDAD DE CAUCHY-BUNIAKOVSKI. *Para cualesquiera dos vectores a y b de un espacio unitario es válida la desigualdad*

$$|(a, b)| \leqslant \|a\|\cdot\|b\|,$$

teniendo lugar la igualdad cuando, y sólo cuando, los vectores a y b son linealmente dependientes.

Para cualquier número λ se tiene en virtud del axioma 4ᶜ

$$(a-\lambda b,\ a-\lambda b)\geqslant 0, \tag{9}$$

de donde realizando la multiplicación obtenemos

$$(a,\ a)-\bar{\lambda}(a,\ b)-\lambda\overline{(a,\ b)}+\lambda\bar{\lambda}(b,\ b)\geqslant 0. \tag{10}$$

Si $b=o$, la desigualdad requerida se cumple de un modo trivial ya que ambos miembros suyos resultan ser iguales a cero. Supongamos, por ello, que $b\neq o$. Tomando en la desigualdad (10) en lugar de λ el número $\frac{(a,\ b)}{(b,\ b)}$ y multiplicando todos los miembros de la desigualdad por el número positivo $(b,\ b)$, obtenemos

$$(a,\ a)(b,\ b)-\overline{(a,\ b)}(a,\ b)-(a,\ b)\overline{(a,\ b)}+(a,\ b)\overline{(a,\ b)}\geqslant 0,$$

es decir,

$$(a,\ b)\overline{(a,\ b)}\leqslant(a,\ a)(b,\ b)$$

o

$$|(a,\ b)|\leqslant\|a\|\cdot\|b\|. \tag{11}$$

Luego, hemos demostrado la desigualdad de Cauchy — Buniakovski.

Si a y b son linealmente independientes, se tiene $a-\lambda b\neq o$ y en lugar de (9) podemos tomar la desigualdad estricta

$$(a-\lambda b,\ a-\lambda b)>0.$$

Pero entonces, podemos omitir en todas las desigualdades sucesivas el signo de igualdad y en lugar de (11) obtendremos la desigualdad

$$|(a,\ b)|<\|a\|\cdot\|b\|. \tag{12}$$

En cambio, si a y b son linealmente dependientes, por ejemplo, $a=\alpha b$, tenemos

$$|(a,\ b)|=|(\alpha b,\ b)|=|\alpha|(b,\ b)=\|a\|\cdot\|b\|.$$

Luego, hemos demostrado también la observación complementaria a la desigualdad de Buniakovski.

Veamos qué es lo que significa la desigualdad de Buniakovski en aquellos espacios concretos que hemos considerado anteriormente. Sea $\mathfrak{L}$ el espacio unitario de filas. Tomemos en el mismo unos vectores $a=[\alpha_1,\ \dots,\ \alpha_n]$ y $b=[\beta_1,\ \dots,\ \beta_n]$. Tenemos según la fórmula (6)

$$|(a,\ b)|=|\alpha_1\bar{\beta}_1+\alpha_2\bar{\beta}_2+\dots+\alpha_n\bar{\beta}_n|,$$

$$\|a\|=\sqrt{|\alpha_1|^2+\dots+|\alpha_n|^2}\quad \text{y}\quad \|b\|=\sqrt{|\beta_1|^2+\dots+|\beta_n|^2}.$$

La desigualdad de Cauchy — Buniakovski significa, por consiguiente, que

$$\alpha_1\bar{\beta}_1+\dots+\alpha_n\bar{\beta}_n\leqslant\sqrt{|\alpha_1|^2+\dots+|\alpha_n|^2}\sqrt{|\beta_1|^2+\dots+|\beta_n|^2},$$

donde α_j y β_j son números complejos cualesquiera.

Análogamente, si $\mathfrak{L}$ es el espacio de funciones mencionado anteriormente, la desigualdad de Buniakovski se convierte en la desigualdad

$$\left|\int_0^1 f(t)\overline{g(t)}\,dt\right|^2 \leqslant \int_0^1 |f(t)|^2\,dt \int_0^1 |g(t)|^2\,dt.$$

Volvamos a los espacios unitarios arbitrarios. La desigualdad de Cauchy—Buniakovski permite ahora demostrar fácilmente la siguiente proposición: *la longitud de una suma de vectores no sobrepasa la suma de las longitudes de los sumandos.* Es obvio que basta considerar el caso de dos sumandos solamente, ya que el caso general se deduce de éste por inducción. Tenemos

$$\|a+b\|^2=(a+b,\ a+b)=(a,\ a)+(a,\ b)+$$
$$+\overline{(a,\ b)}+(b,\ b)=(a,\ a)+2\,\mathrm{Re}\,(a,\ b)+(b,\ b),$$

donde $\mathrm{Re}\,(a,\ b)$ es la parte real de $(a,\ b)$. Puesto que

$$\mathrm{Re}\,(a,\ b)\leqslant|(a,\ b)|\leqslant\|a\|\cdot\|b\|,$$

se tiene

$$\|a+b\|^2\leqslant\|a\|^2+2\|a\|\cdot\|b\|+\|b\|^2=(\|a\|+\|b\|)^2,$$

de donde resulta

$$\|a+b\|\leqslant\|a\|+\|b\|$$

que es lo que se quería demostrar.

La expresión $\|a-b\|$ suele llamarse a veces *distancia entre los vectores a* y b. Indicándola por $\rho(a,\ b)$ obtenemos las relaciones siguientes

1) $\rho(a,\ a)=0$; $\rho(a,\ b)>0$, si $a\neq b$;
2) $\rho(a,\ b)=\rho(b,\ a)$;
3) $\rho(a,\ b)+\rho(b,\ c)\geqslant\rho(a,\ c)$.

La demostración de las mismas es evidente; por ejemplo, la última resulta de

$$\rho(a,\ c)=\|a-c\|=\|(a-b)+(b-c)\|\leqslant$$
$$\leqslant\|a-b\|+\|b-c\|=\rho(a,\ b)+\rho(b,\ c).$$

17.3. Sistemas ortonormales. Unos vectores a y b de un espacio unitario $\mathfrak{L}$ se llaman *ortogonales* si el producto escalar de a por b es igual a cero. Si $\mathfrak{L}$ es el espacio corriente, el concepto de ortogonalidad coincide con el concepto de perpendicularidad. Por esto la ortogonalidad puede ser considerada como una generalización del concepto de perpendicularidad.

Del axioma 1° se deduce que la relación de ortogonalidad es simétrica: si el vector a es ortogonal a b, el vector b es ortogonal a a. Es obvio que *el vector nulo es ortogonal a cualquier vector del espacio y es el único vector que posee esta propiedad.*

Un sistema $a_1, a_2, \ldots, a_m$ de vectores de un espacio unitario se llama *ortogonal* si cualesquiera dos vectores a_j y a_k $(j \neq k)$ del mismo son ortogonales. Si el sistema contiene solamente un vector, también lo llamaremos ortogonal.

TEOREMA 1. *Todo sistema ortogonal de vectores no nulos de un espacio unitario $\mathfrak{L}$ es linealmente independiente.*

En efecto, sea $a_1, a_2, \ldots, a_m$ un sistema ortogonal de vectores no núlos y sea

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_m a_m = o.$$

Multiplicando esta igualdad escalarmente por a_j, obtenemos

$$\alpha_1 (a_1, a_j) + \alpha_2 (a_2, a_j) + \ldots + \alpha_m (a_m, a_j) = 0$$

o

$$\alpha_j (a_j, a_j) = 0, \tag{13}$$

ya que debido a la ortogonalidad del sistema todos los demás términos se anulan. Pero $a_j \neq o$; por consiguiente, $(a_j, a_j) \neq 0$ y de (13) resulta $\alpha_j = 0$ que es lo que se quería demostrar.

Sea n la dimensión del espacio $\mathfrak{L}$. El teorema 1 muestra que todo sistema ortogonal de vectores no nulos de $\mathfrak{L}$ no puede contener más de n vectores. Si en $\mathfrak{L}$ existen n vectores no nulos ortogonales, ellos constituyen *una base ortogonal* del espacio $\mathfrak{L}$. Probemos que en $\mathfrak{L}$ siempre existe una base de esta índole. Es más, probemos que *en todo espacio unitario $\mathfrak{L}$ cualquier sistema ortogonal de vectores no nulos puede ser complementado hasta obtener una base ortogonal del espacio $\mathfrak{L}$.* Sea dado en $\mathfrak{L}$ un sistema ortogonal $a_1, \ldots, a_m$. Complementémoslo hasta obtener un sistema ortogonal maximal $a_1, \ldots, a_m, a_{m+1}, \ldots, a_s$ de vectores no nulos. Tal completación es posible, ya que según el teorema 1 ningún sistema ortogonal puede contener más de n vectores diferentes de cero. Probemos que $a_1, \ldots, a_m, \ldots, a_s$ es precisamente la base requerida del espacio $\mathfrak{L}$. Consideremos un vector x cualquiera de $\mathfrak{L}$. Pongamos

$$\xi_k = \frac{(x, a_k)}{(a_k, a_k)} \quad (k = 1, 2, \ldots, s),$$
$$y = \xi_1 a_1 + \xi_2 a_2 + \ldots + \xi_s a_s.$$

Multiplicando escalarmente todos los términos de la última igualdad por a_k, obtenemos

$$(y, a_k) = (x, a_k),$$

de donde resulta

$$(x - y, a_k) = (x, a_k) - (y, a_k) = 0.$$

Por consiguiente, el vector $x - y$ es ortogonal a todos los vectores $a_1, \ldots, a_s$. El sistema $a_1, \ldots, a_s$ es, por hipótesis, un sistema ortogonal maximal de vectores no nulos de $\mathfrak{L}$; luego, $x-y=o$, es decir,

$$x=\xi_1 a_1+\xi_2 a_2+\ldots+\xi_s a_s.$$

Por lo tanto, el sistema linealmente independiente $a_1, \ldots, a_s$ es tal que cualquier vector x se expresa linealmente en términos del mismo. Pero esto significa precisamente que $a_1, \ldots, a_s$ es una base del espacio $\mathfrak{L}$.

Un sistema ortogonal formado por vectores de longitud igual a la unidad se denomina *ortonormal*. Es obvio que normalizando vectores ortogonales no nulos obtenemos de nuevo vectores ortogonales. Por lo tanto, al normalizar los vectores de una base ortogonal del espacio, obtendremos una base ortonormal de este espacio. Hemos visto que todo sistema de vectores ortogonales no nulos puede ser complementado hasta obtener una base ortogonal del espacio. Por esto es válido el teorema siguiente.

TEOREMA 2. *Todo sistema ortonormal de vectores de un espacio $\mathfrak{L}$ puede ser complementado hasta obtener una base ortonormal de este espacio.*

Puesto que cualquier sistema que contiene sólo un vector es ortogonal, de aquí se deduce, en particular, que *en todo espacio unitario existe una base ortonormal.*

Un sistema de vectores $e_1, e_2, \ldots, e_n$ considerados en un orden determinado ha sido llamado sistema de coordenadas de un espacio $\mathfrak{L}$, si $e_1, \ldots, e_n$ forman una base del espacio $\mathfrak{L}$. Si $e_1, \ldots, e_n$ es una base ortonormal, también el sistema de coordenadas se denomina ortonormal. La diferencia que existe entre un sistema de coordenadas arbitrario y un sistema de coordenadas ortonormal es la misma que existe entre un sistema de coordenadas cartesiano oblicuo y un sistema ortogonal del espacio corriente. Efectivamente, sea $\mathfrak{R}$ el espacio corriente de vectores-segmentos provisto del producto escalar habitual. Una base arbitraria del espacio $\mathfrak{R}$ está formada por tres vectores cualesquiera a_1, a_2 y a_3 que parten de un punto O y que no pertenecen a un mismo plano (véase la fig. 1 de la pág. 79). Tomemos un punto A y expresemos el vector $\overrightarrow{OA}$ en términos de los vectores coordenados a_1, a_2 y a_3. Sea

$$\overrightarrow{OA}=\alpha_1 a_1+\alpha_2 a_2+\alpha_3 a_3.$$

Los números α_1, α_2 y α_3 son, según la definición, las coordenadas del vector $\overrightarrow{OA}$. Al mismo tiempo, se ve de la figura que estos números son las coordenadas del punto A calculadas en el sistema de coordenadas cartesiano oblicuo con los ejes a_1, a_2 y a_3, en el que los vectores a_1, a_2 y a_3 se han tomado como segmentos unidades a lo largo de los ejes.

Por consiguiente, un sistema arbitrario de coordenadas en $\mathfrak{R}$ es equivalente, desde el punto de vista de la Geometría analítica, a un sistema de coordenadas cartesiano oblicuo con distintas unidades a lo largo de los ejes coordenados. Por el contrario, un sistema ortonormal de coordenadas e_1, e_2 y e_3 será equivalente, en este sentido, al sistema de coordenadas cartesiano ortogonal habitual con unidades iguales sobre los ejes (fig. 4).

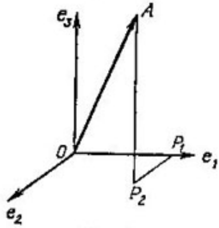

Fig. 4.

Los sistemas ortonormales de coordenadas de los espacios unitarios tienen varias propiedades específicas. Algunas de éstas serán consideradas ahora.

Si $e_1, e_2, \ldots, e_n$ es un sistema ortonormal de coordenadas de un espacio $\mathfrak{L}$, las coordenadas de un vector cualquiera a son iguales respectivamente a los productos escalares (a, e_1), (a, e_2), $\ldots$, (a, e_n).

En efecto, sea

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n.$$

Multiplicando esta igualdad escalarmente por e_k y teniendo en cuenta que $(e_j, e_k) = 0$ para $j \neq k$, obtenemos

$$(a, e_k) = \alpha_k (e_k, e_k) = \alpha_k \quad (k = 1, 2, \ldots, n)$$

que es lo que quería demostrar.

Si en un sistema de coordenadas ortonormal los vectores a y b tienen respectivamente las coordenadas $\alpha_1, \alpha_2, \ldots, \alpha_n$ y $\beta_1, \beta_2, \ldots, \beta_n$, se tiene

$$(a, b) = \alpha_1 \bar{\beta}_1 + \alpha_2 \bar{\beta}_2 + \ldots + \alpha_n \bar{\beta}_n. \tag{14}$$

Efectivamente, sea $e_1, e_2, \ldots, e_n$ un sistema de coordenadas ortonormal dado y sea

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$b = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n.$$

Entonces,

$$(a, b) = (\sum \alpha_j e_j, \quad \sum \beta_k e_k) = \sum\sum \alpha_j \bar{\beta}_k (e_j, e_k) = \sum \alpha_j \bar{\beta}_j.$$

DESIGUALDAD DE BESSEL. *Si $e_1, e_2, \ldots, e_m$ es un sistema ortonormal cualquiera de vectores de un espacio unitario y a es un vector arbitrario del mismo y si tomamos*

$$(a, e_k) = \alpha_k \quad (k = 1, 2, \ldots, m),$$

tiene lugar la desigualdad

$$(a, a) \geqslant |\alpha_1|^2 + |\alpha_2|^2 + \ldots + |\alpha_m|^2.$$

Daremos una demostración independiente de esta igualdad, aunque la misma se puede deducir fácilmente de las proposiciones anteriores. Consideremos un vector auxiliar

$$x=\alpha_1e_1+\alpha_2e_2+\ldots+\alpha_me_m.$$

Tenemos

$$(a-x,\ a-x)=(a,\ a)-(a,\ x)-(x,\ a)+(x,\ x)\geqslant 0, \tag{15}$$

$$(x,\ x)=\left(\sum \alpha_je_j,\ \sum \alpha_je_j\right)=\sum \alpha_j\bar{\alpha}_j, \tag{16}$$

$$(a,\ x)=\left(a,\ \sum \alpha_je_j\right)=\sum \alpha_j\bar{\alpha}_j, \tag{17}$$

$$(x,\ a)=\overline{(a,\ x)}=\sum \alpha_j\bar{\alpha}_j. \tag{18}$$

Introduciendo (16), (17) y (18) en la desigualdad (15) obtenemos

$$(a,\ a)-\sum \alpha_j\bar{\alpha}_j-\sum \alpha_j\bar{\alpha}_j+\sum \alpha_j\bar{\alpha}_j\geqslant 0,$$

es decir,

$$(a,\ a)\geqslant \sum \alpha_j\bar{\alpha}_j$$

que es lo que se quería demostrar.

Si $e_1, e_2, \ldots, e_n$ es una base ortonormal, en lugar de la desigualdad de Bessel tendremos, según (14), la igualdad

$$(a,\ a)=\alpha_1\bar{\alpha}_1+\alpha_2\bar{\alpha}_2+\ldots+\alpha_n\bar{\alpha}_n,$$

que se llama *igualdad de Parseval.* La igualdad de Parseval, además de ser una condición necesaria, es también una condición suficiente para que un sistema $e_1, e_2, \ldots, e_n$ ortonormal sea una base.

Si $e_1, e_2, \ldots, e_n$ es un sistema ortonormal de vectores de un espacio $\mathfrak{L}$ y si para todo vector a de $\mathfrak{L}$ se tiene

$$(a,\ a)=\alpha_1\bar{\alpha}_1+\alpha_2\bar{\alpha}_2+\ldots+\alpha_n\bar{\alpha}_n,$$

donde $\alpha_j=(a,\ e_j)$, $j=1, 2, \ldots, n$, entonces el sistema $e_1, e_2, \ldots, e_n$ es una base del espacio $\mathfrak{L}$.

La demostración se deduce fácilmente de las proposiciones anteriores y queda a cargo del lector. •

Hemos demostrado la existencia de una base ortonormal en todo espacio unitario mediante razonamientos indirectos. No obstante, existe un método directo que permite obtener a partir de cualquier base de un espacio una base ortonormal del mismo. Este método lleva el nombre de proceso de *ortonormalización de Gram—Schmidt* y se emplea con frecuencia al considerar los espacios de funciones. Su esencia consiste en lo siguiente. Sea $a_1, \ldots, a_m$ un sistema linealmente independiente de vectores de un espacio unitario $\mathfrak{L}$. Planteémonos la tarea de construir un sistema ortonormal de vectores $e_1, e_2, \ldots, e_m$ tal que todo vector j-ésimo e_j del mismo se exprese linealmente en términos de los j primeros vectores $a_1, \ldots, a_j$. Realizaremos la construcción por inducción. El vector e_1 debe expresarse, según lo convenido, en términos de a_1 y debe ser de longitud 1. Un vector así se obtiene normalizando a_1:

$$e_1=\frac{1}{\sqrt{(a_1,\ a_1)}}a_1.$$

Supongamos ahora que ya han sido construidos los vectores $e_1, \ldots, e_j$ con las propiedades requeridas correspondientes a un valor de j. Busquemos e_{j+1}. Ante todo, escojamos unos números $\alpha_1, \ldots, \alpha_j$ de modo que el vector auxiliar

$$e'_{j+1} = a_{j+1} + \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_j e_j \tag{19}$$

sea ortogonal a los vectores $e_1, e_2, \ldots, e_j$. Multiplicando (19) escalarmente por e_k $(k=1, 2, \ldots, j)$ veremos que para ello es necesario que

$$\alpha_k = -(a_{j+1}, e_k). \tag{20}$$

Recíprocamente, tomando para α_k los valores de (20) e introduciéndolos en (19), obtenemos un vector e'_{j+1} ortogonal a $e_1, e_2, \ldots, e_j$. Puesto que a_{j+1} no puede expresarse linealmente en términos de $a_1, \ldots, a_j$ y, por consiguiente, no puede ser combinación lineal de los vectores $e_1, \ldots, e_j$, resulta que e'_{j+1} es diferente de cero y, por lo tanto, puede ser normalizado. Tomemos

$$e_{j+1} = \frac{1}{|e'_{j+1}|} e'_{j+1}. \tag{21}$$

Las relaciones (21) y (19) muestran que e_{j+1} se expresa linealmente en términos de $a_1, a_2, \ldots, a_{j+1}$. Además, e_{j+1} está normalizado y es ortogonal a todos los vectores $e_1, \ldots, e_j$. Por consiguiente, se cumplen las suposiciones de inducción y podemos considerar que la sucesión $e_1, e_2, \ldots, e_m$ ha sido construida. Si la sucesión inicial $a_1, a_2, \ldots, a_m$ era una base del espacio, es obvio que la sucesión $e_1, e_2, \ldots, e_m$ obtenida mediante el proceso de ortonormalización también será una base del espacio $\mathfrak{L}$.

17.4. Isomorfismo. En el p. 4.3 hemos quedado en llamar isomorfos dos espacios lineales si entre los elementos de los mismos se puede establecer una correspondencia biyectiva que conserva las operaciones de adición y de multiplicación por número. En los espacios unitarios a estas operaciones se agrega además la de multiplicación escalar. Por esto resulta natural llamar isomorfos los espacios unitarios sólo en el caso en el que ellos se comportan idénticamente respecto a las tres operaciones mencionadas.

DEFINICIÓN. *Dos espacios unitarios $\mathfrak{L}$ y $\mathfrak{L}_1$ sobre un mismo campo de coeficientes se llaman isomorfos, si entre sus elementos se puede establecer una correspondencia biyectiva en la que la suma de dos vectores de $\mathfrak{L}$ se transforma en la suma de los vectores correspondientes de $\mathfrak{L}_1$, el producto de un número por un vector de $\mathfrak{L}$ se transforma en el producto del mismo número por el vector correspondiente de $\mathfrak{L}_1$ y los productos escalares de pares correspondientes de vectores de $\mathfrak{L}$ y de $\mathfrak{L}_1$ coinciden.*

Tendrán interés para nosotros sólo aquellas propiedades de los espacios unitarios que sean propiedades de las tres operaciones principales definidas en estos espacios y que no dependan de la naturaleza de los elementos que constituyen los espacios. Desde este punto de vista los espacios unitarios isomorfos tendrán las mismas propiedades. De aquí se ve la importancia de saber clasificar, salvo un isomorfismo, todos los espacios unitarios. Esta clasificación no difiere en nada de la clasificación de los espacios lineales y queda determinada por el teorema siguiente.

TEOREMA 3. *Para que dos espacios unitarios sobre un mismo campo de coeficientes sean isomorfos es necesario y suficiente que coincidan las dimensiones de estos espacios.*

En efecto, si dos espacios unitarios $\mathfrak{L}$ y $\mathfrak{L}_1$ son isomorfos, también serán isomorfos como espacios lineales, es decir, respecto a las operaciones de adición y de multiplicación por número. Pero los espacios lineales isomorfos son de la misma dimensión y, por consiguiente, las dimensiones de los espacios $\mathfrak{L}$ y $\mathfrak{L}_1$ coinciden. Hemos demostrado la necesidad. Recíprocamente, supongamos que las dimensiones de los espacios $\mathfrak{L}$ y $\mathfrak{L}_1$ coinciden. Tomemos en $\mathfrak{L}$ y $\mathfrak{L}_1$ unas bases ortonormales $e_1, \ldots, e_n$ y $e_1', \ldots, e_n'$. Diremos que los vectores $x \in \mathfrak{L}$ y $x' \in \mathfrak{L}_1$ son *correspondientes* si sus coordenadas en las bases escogidas coinciden. Esta correspondencia es biyectiva y conserva las operaciones de adición y de multiplicación por número (p. 4.3). Por ello sólo debemos probar que los productos escalares de los pares correspondientes de vectores son idénticos. Consideremos dos vectores cualesquiera

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \ldots + \alpha_n e_n,$$
$$b = \beta_1 e_1 + \beta_2 e_2 + \ldots + \beta_n e_n.$$

del espacio $\mathfrak{L}$. Los vectores correspondientes de $\mathfrak{L}_1$ son

$$a' = \alpha_1 e_1' + \alpha_2 e_2' + \ldots + \alpha_n e_n',$$
$$b' = \beta_1 e_1' + \beta_2 e_2' + \ldots + \beta_n e_n'.$$

Puesto que las bases $e_1, \ldots, e_n$ y $e_1', \ldots, e_n'$ son ortonormales, tenemos en virtud de la fórmula (14)

$$(a, b) = \alpha_1 \overline{\beta}_1 + \alpha_2 \overline{\beta}_2 + \ldots + \alpha_n \overline{\beta}_n = (a', b')$$

que es lo que se quería demostrar.

17.5. Sumas ortogonales. Proyecciones. Dos conjuntos de vectores $\mathfrak{M}$ y $\mathfrak{N}$ de un espacio unitario $\mathfrak{L}$ se llaman *ortogonales*, si todo vector del primer conjunto es ortogonal a todo vector del segundo. En particular, se dice que el vector a es ortogonal al conjunto $\mathfrak{M}$, si a es ortogonal a todo vector de $\mathfrak{M}$. A veces, la ortogonalidad de $\mathfrak{M}$ y $\mathfrak{N}$ se indica simbólicamente por $\mathfrak{M} \perp \mathfrak{N}$.

Si dos conjuntos $\mathfrak{M}$ y $\mathfrak{N}$ son ortogonales, la intersección de los mismos o bien es vacía, o bien consta del vector nulo únicamente.

En efecto, si el vector a está contenido en $\mathfrak{M}$ y en $\mathfrak{N}$, se tiene $(a, a) = 0$, de donde $a = o$.

Una suma $\mathfrak{A}_1 + \ldots + \mathfrak{A}_m$ de varios subespacios lineales se llama *ortogonal* si cualesquiera dos subespacios $\mathfrak{A}_j$ y $\mathfrak{A}_k$ $(j \neq k)$ son ortogonales.

Una suma ortogonal de subespacios es siempre una suma directa.

Efectivamente, si la suma es sólo de dos sumandos, la intersección de los mismos, en virtud de la observación anterior, consta solamente del vector nulo y, por consiguiente, la suma es directa. En el caso general, la demostración se realiza por inducción.

Si la suma

$$\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2 + \ldots + \mathfrak{A}_s$$

es ortogonal y si

$$a = a_1 + a_2 + \ldots + a_s,$$
$$b = b_1 + b_2 + \ldots + b_s,$$

donde $a_j \in \mathfrak{A}_j$ y $b_j \in \mathfrak{A}_j$, $j = 1, \ldots, s$, *se tiene*

$$(a, b) = (a_1, b_1) + (a_2, b_2) + \ldots + (a_s, b_s). \tag{22}$$

En efecto, puesto que $\mathfrak{A}_j \perp \mathfrak{A}_k$ para $j \neq k$, resulta que $(a_j, b_k) = 0$. Por consiguiente,

$$(a, b) = \left(\sum a_j, \sum b_j\right) = \sum\sum (a_j, b_k) = \sum (a_j, b_j)$$

que es lo que se quería demostrar.

Consideremos ahora un conjunto no vacío cualquiera $\mathfrak{M}$ de vectores de un espacio unitario $\mathfrak{L}$. El conjunto de todos los vectores del espacio $\mathfrak{L}$ ortogonales a $\mathfrak{M}$ se llama *complemento ortogonal* del conjunto $\mathfrak{M}$ y se indica por $\mathfrak{M}^{\perp}$.

El complemento ortogonal de un conjunto no vacío cualquiera $\mathfrak{M}$ es un subespacio lineal.

En efecto, si a y b pertenecen al complemento ortogonal $\mathfrak{M}^{\perp}$ y c es un vector cualquiera de $\mathfrak{M}$, se tiene

$$(\alpha a + \beta b, c) = \alpha (a, c) + \beta (b, c) = 0.$$

Por consiguiente, cualesquiera que sean α y β el vector $\alpha a + \beta b$ está contenido en $\mathfrak{M}^{\perp}$ y $\mathfrak{M}^{\perp}$ es un subespacio lineal.

TEOREMA 4. *Todo espacio unitario $\mathfrak{L}$ es la suma directa de cualquier subespacio lineal suyo $\mathfrak{A}$ y de su complemento ortogonal $\mathfrak{A}^{\perp}$.*

Sea $e_1, e_2, \ldots, e_m$ una base ortonormal del subespacio $\mathfrak{A}$ y sea $e_{m+1}, \ldots, e_s$ una base ortonormal del subespacio $\mathfrak{A}^{\perp}$. Para demostrar el teorema es suficiente ver que $e_1, \ldots, e_m, \ldots, e_s$ es una base del espacio $\mathfrak{L}$. Supongamos, por el contrario, que el sistema $e_1, \ldots, e_s$ no es una base del espacio $\mathfrak{L}$. Entonces, de acuerdo con el teorema 2, este sistema puede ser complementado hasta obtener una base ortonormal del espacio $\mathfrak{L}$. Sea e uno de los vectores complementarios. Puesto que e es ortogonal a todos los vectores $e_1, \ldots, e_m$, el vector e está contenido en $\mathfrak{A}^{\perp}$. Por consiguiente, $\mathfrak{A}^{\perp}$ contiene un sistema ortonormal y, por ende, linealmente independiente de vectores $e_{m+1}, \ldots, e_s, e$. Pero esto contradice a nuestra hipótesis de que $e_{m+1}, \ldots, e_s$ es una base de $\mathfrak{A}^{\perp}$. Hemos demostrado el teorema.

Del teorema 4 se desprende, en particular, que

$$\mathfrak{A}^{\perp\perp}=\mathfrak{A}. \tag{23}$$

Efectivamente, $\mathfrak{A}^{\perp\perp}$ está contenido indudablemente en $\mathfrak{A}$. Por otra parte, según el teorema 4, para cualquier x de $\mathfrak{A}^{\perp\perp}$ tenemos

$$x=a+b, \quad a\in\mathfrak{A} \quad \text{y} \quad b\in\mathfrak{A}^{\perp}.$$

Multiplicando escalarmente esta igualdad por b, obtenemos $(b, b)=0$, es decir, $b=o$ y $x=a$. Luego, $\mathfrak{A}^{\perp\perp}=\mathfrak{A}$.

Definemos, para concluir, el concepto de proyección ortogonal de un vector sobre un subespacio lineal. Sea $\mathfrak{A}$ un subespacio lineal de un espacio unitario $\mathfrak{L}$. En virtud del teorema 4, $\mathfrak{L}$ es la suma directa del subespacio $\mathfrak{A}$ y de su complemento ortogonal $\mathfrak{A}^{\perp}$. Por consiguiente, todo vector x de $\mathfrak{L}$ puede ser representado unívocamente en forma de una suma

$$x=a+b \quad (a\in\mathfrak{A} \quad \text{y} \quad b\in\mathfrak{A}^{\perp}). \tag{24}$$

El sumando a se llama *proyección del vector x sobre el subespacio $\mathfrak{A}$*. Puesto que de acuerdo con la fórmula (23) $\mathfrak{A}$ es el complemento ortogonal de $\mathfrak{A}^{\perp}$, el sumando b de la igualdad (24) representa la proyección del vector x sobre el subespacio $\mathfrak{A}^{\perp}$.

Multiplicando (24) por un número α, obtenemos

$$\alpha x=\alpha a+\alpha b \quad (\alpha a\in\mathfrak{A} \ \text{y} \ \alpha b\in\mathfrak{A}^{\perp}),$$

es decir, *la proyección del producto de un número por un vector es igual al producto de este número por la proyección del vector*.

Análogamente se demuestra también la proposición de que *la proyección de una suma de vectores sobre un subespacio es igual a la suma de las proyecciones de los sumandos sobre este subespacio.*

Ejemplos y problemas.

1. Sea e_1, e_2 y e_3 un sistema ortonormal de coordenadas de un espacio unitario de tres dimensiones. Pruébese que el sistema de vectores

$$a_1=\frac{1}{3}(2e_1+2e_2-e_3),$$
$$a_2=\frac{1}{3}(2e_1-e_2+2e_3),$$
$$a_3=\frac{1}{3}(e_1-2e_2+2e_3)$$

también constituye un sistema ortonormal de coordenadas de este espacio.

2. En un espacio unitario $\mathfrak{L}$ se ha tomado un sistema ortonormal de coordenadas $e_1, \ldots, e_n$. Pruébese que un sistema de vectores $a_1, \ldots, a_n$ será también una base ortonormal del espacio $\mathfrak{L}$ cuando, y sólo cuando, la matriz formada por las filas coordenadas de estos vectores sea unitaria (véase el p. 1.3).

3. Si $a_1, \ldots, a_m$ es una base ortonormal de un subespacio lineal $\mathfrak{A}$, la proyección de un vector x sobre $\mathfrak{A}$ es igual a

$$(x, a_1)\, a_1 + (x, a_2)\, a_2 + \ldots + (x, a_m)\, a_m.$$

4. En el espacio de todos los polinomios en λ de grado no mayor que n el producto escalar se define mediante la fórmula (7) del p. 17.1. Pruébese que éste será un espacio unitario de dimensión $n+1$. Pruébese que ortonormalizando según Gram—Schmidt en este espacio la sucesión 1, λ y λ^2 obtenemos los polinomios 1, $\sqrt{3}\,(2\lambda-1)$ y $\sqrt{5}\,(6\lambda^2-6\lambda+1)$.

5. Se llama *determinante de Gram* de un sistema de vectores $a_1, a_2, \ldots, a_n$ de un espacio unitario $\mathfrak{L}$ de n dimensiones el determinante

$$\Delta = \begin{vmatrix} (a_1, a_1) & (a_1, a_2) & \ldots & (a_1, a_n) \\ (a_2, a_1) & (a_2, a_2) & \ldots & (a_2, a_n) \\ \cdot & \cdot & \cdot & \cdot \\ (a_n, a_1) & (a_n, a_2) & \ldots & (a_n, a_n) \end{vmatrix}.$$

Supongamos que en $\mathfrak{L}$ se ha escogido un sistema ortonormal de coordenadas. Pruébese que el determinante de Gram Δ es igual al cuadrado del módulo del determinante formado por las filas coordenadas de los vectores $a_1, \ldots, a_n$. Pruébese también que $a_1, a_2, \ldots, a_n$ son linealmente independientes cuando, y sólo cuando, el determinante de Gram de los mismos sea diferente de cero.

6. Dése una interpretación geométrica al proceso de ortonormalización de Gram—Schmidt en el caso del espacio habitual de tres dimensiones de los vectores-segmentos.

§ 18. Aplicaciones conjugadas

18.1. Funciones lineales. Sea $\mathfrak{L}$ un espacio lineal cualquiera de dimensión finita sobre un campo K. Pongamos en correspondencia a todo vector x de $\mathfrak{L}$ un número $f(x)$ de K. Las correspondencias de este tipo han sido llamadas en el p. 4.1 funciones con valores en K definidas sobre $\mathfrak{L}$. Una función $f(x)$ se llama *lineal* si para cualesquiera x e y de $\mathfrak{L}$ y cualesquiera α y β de K se tiene

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y). \tag{1}$$

Tomando en (1) $\alpha = \beta = 0$, obtenemos

$$f(o) = 0.$$

Por esta razón en lugar de «función lineal» se dice a veces «función lineal *homogénea*».

Es fácil ver que *la suma de funciones lineales y el producto de un número por una función lineal son de nuevo funciones lineales.* Probemos, por ejemplo, la primera proposición. Sea $f = g + h$, donde g y h son funciones lineales. De acuerdo con la definición

$$f(\alpha x + \beta y) = g(\alpha x + \beta y) + h(\alpha x + \beta y) =$$
$$= \alpha g(x) + \beta g(y) + \alpha h(x) + \beta h(y) = \alpha f(x) + \beta f(y)$$

que es lo que se quería demostrar.

Conocemos ya que las operaciones de adición de funciones y de multiplicación de las mismas por números satisfacen los axiomas de un espacio lineal. Puesto que la suma de funciones lineales y el producto de una función lineal por un número son de nuevo funciones lineales, resulta que el conjunto de todas las funciones lineales, definidas sobre un espacio vectorial $\mathfrak{L}$, es por sí mismo un espacio lineal. Este espacio se llama espacio *conjugado* (o dual) respecto a $\mathfrak{L}$ y se indica por $\mathfrak{L}'$.

De la igualdad (1) que caracteriza las funciones lineales se desprende inmediatamente una relación más general

$$f(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m) = \alpha_1 f(x_1) + \alpha_2 f(x_2) + \ldots + \alpha_m f(x_m), \quad (2)$$

cuya demostración omitimos ya que es obvia. Los teoremas que siguen esclarecen en gran medida la estructura de las funciones lineales.

TEOREMA 1. *Sea $e_1, e_2, \ldots, e_n$ un sistema de coordenadas cualquiera de un espacio lineal $\mathfrak{L}$. Tomemos una sucesión totalmente arbitraria $\alpha_1, \alpha_2, \ldots, \alpha_n$ de números de K. Entonces existe una función lineal $f(x)$, y sólo una, que está definida sobre $\mathfrak{L}$ y que satisface las condiciones*

$$f(e_j) = \alpha_j \quad (j = 1, 2, \ldots, n). \quad (3)$$

DEMOSTRACIÓN. Sea x un vector de $\mathfrak{L}$:

$$x = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_n e_n. \quad (4)$$

Poniendo en correspondencia a este vector el número $\alpha_1\xi_1 + \alpha_2\xi_2 + \ldots + \alpha_n\xi_n$, obtenemos una función $f(x)$ sobre $\mathfrak{L}$. Es decir, por definición

$$f(x) = \alpha_1\xi_1 + \alpha_2\xi_2 + \ldots + \alpha_n\xi_n.$$

Tomando aquí $x = e_j$ tendremos $f(e_j) = \alpha_j$ $(j = 1, \ldots, n)$. Por otra parte, si

$$y = \eta_1 e_1 + \eta_2 e_2 + \ldots + \eta_n e_n, \quad (5)$$

tenemos

$$f(y) = \alpha_1\eta_1 + \alpha_2\eta_2 + \ldots + \alpha_n\eta_n.$$

De (4) y de (5) se deduce que

$$\alpha x + \beta y = (\alpha\xi_1 + \beta\eta_1) e_1 + \ldots + (\alpha\xi_n + \beta\eta_n) e_n,$$

de donde

$$f(\alpha x + \beta y) = \alpha_1(\alpha\xi_1 + \beta\eta_1) + \ldots + \alpha_n(\alpha\xi_n + \beta\eta_n) = \alpha f(x) + \beta f(y).$$

Por consiguiente, $f(x)$ es una función lineal.

Hemos demostrado que existe una función lineal que satisface las condiciones (3). Probemos que es única. Sea $g(x)$ una función lineal tal que $g(e_j) = \alpha_j$ $(j = 1, \ldots, n)$. Entonces para un vector cualquiera x de coordenadas $\xi_1, \ldots, \xi_n$ tendremos

$$g(x) = g(\xi_1 e_1 + \ldots + \xi_n e_n) = \xi_1 g(e_1) + \ldots + \xi_n g(e_n) = \\ = \xi_1 \alpha_1 + \ldots + \xi_n \alpha_n,$$

es decir, $g(x) = f(x)$.

El teorema 1 ha sido demostrado para espacios lineales $\mathfrak{L}$ cualesquiera. Si el espacio $\mathfrak{L}$ es unitario, las funciones lineales sobre $\mathfrak{L}$ tienen una expresión muy sencilla en términos del producto escalar.

TEOREMA 2. *El producto escalar (x, a) de vectores de un espacio unitario $\mathfrak{L}$ es una función lineal de x siendo a fijo. Con ello a todo vector a de $\mathfrak{L}$ se pone en correspondencia una función lineal definida sobre $\mathfrak{L}$. Además, para distintos vectores se obtienen distintas funciones lineales y todas las funciones lineales sobre $\mathfrak{L}$ pueden ser construidas por este procedimiento.*

La primera afirmación es evidente, ya que de $f(x) = (x, a)$ se deduce que

$$f(\alpha x + \beta y) = (\alpha x + \beta y, a) = \alpha (x, a) + \beta (y, a) = \alpha f(x) + \beta f(y).$$

Demostremos la segunda afirmación. Supongamos que, al contrario, a diferentes vectores a y b corresponde una misma función lineal; entonces esto significa que la igualdad $(x, a) = (x, b)$ tiene lugar para cualesquiera vectores x de $\mathfrak{L}$. Pasando los términos a un mismo miembro, podemos representarla en la forma $(x, a-b) = 0$, de donde tomando $x = a-b$ obtenemos $(a-b, a-b) = 0$. Pero el vector nulo es el único vector cuyo cuadrado escalar es igual a cero, es decir $a-b=o$ o $a=b$. Resta demostrar la tercera afirmación, esto es, que toda función lineal $f(x)$ definida sobre $\mathfrak{L}$ puede ser representada en la forma

$$f(x) = (x, a), \tag{6}$$

donde a depende de f y x es un vector arbitrario de $\mathfrak{L}$. Indiquemos por $\mathfrak{A}$ el conjunto de todos los vectores para los cuales $f(x) = 0$. Puesto que $f(o) = 0$, el vector o indudablemente pertenece a $\mathfrak{A}$. Además, si los vectores a y b pertenecen a $\mathfrak{A}$, se tiene

$$f(\alpha a + \beta b) = \alpha f(a) + \beta f(b) = 0,$$

es decir, también $\alpha a + \beta b$ pertenece a $\mathfrak{A}$. Por consiguiente, $\mathfrak{A}$ es un subespacio lineal del espacio $\mathfrak{L}$. Si $\mathfrak{A} = \mathfrak{L}$, resulta que $f(x) = 0$ para todos los vectores x. Para satisfacer la condición (6) es suficiente tomar en este caso $a = o$. Supongamos por ello que $\mathfrak{A} \neq \mathfrak{L}$. Tomemos un vector no nulo b cualquiera de $\mathfrak{L}$ que sea ortogonal a $\mathfrak{A}$ y tratemos de determinar un número α tal que el vector $a = \alpha b$ satisfaga la relación (6). Sea $f(b) = \beta$ y sea $f(x) = \xi$, donde x es un vector cualquiera de $\mathfrak{L}$. Tenemos

$$f\left(x - \frac{\xi}{\beta} b\right) = f(x) - \frac{\xi}{\beta} f(b) = 0.$$

Indicando por c la diferencia $x-\frac{\xi}{\beta}b$, veremos que c pertenece a $\mathfrak{A}$ y que $x=c+\frac{\xi}{\beta}b$. De aquí resulta

$$(x, a)=\left(\frac{\xi}{\beta}b+c, \alpha b\right)=\frac{\bar{\alpha}\xi}{\beta}(b, b)+\bar{\alpha}(c, b).$$

Puesto que el vector b es ortogonal a $\mathfrak{A}$, tenemos $(c, b)=0$ y, por consiguiente,

$$(x, a)=\frac{\bar{\alpha}\xi}{\beta}(b, b). \tag{7}$$

La igualdad (7) muestra que siendo $\alpha=\bar{\beta}(b, b)^{-1}$ se tiene para cualquier x

$$(x, a)=\xi=f(x)$$

que es lo que se quería demostrar.

18.2. Aplicaciones conjugadas. Emplearemos ahora los resultados del punto anterior con el fin de obtener para toda aplicación lineal de un espacio unitario una aplicación nueva unívocamente determinada y llamada *conjugada* de la dada.

Sea $\mathcal{A}$ una aplicación lineal de un espacio unitario $\mathfrak{L}$. Tomemos en $\mathfrak{L}$ un vector arbitrario y y consideremos la expresión

$$f(x)=(x\mathcal{A}, y), \tag{8}$$

donde x es un vector variable. Puesto que

$$f(\alpha u+\beta v)+((\alpha u+\beta v)\mathcal{A}, y)=(\alpha\cdot u\mathcal{A}+\beta\cdot v\mathcal{A}, y)=\alpha f(u)+\beta f(v),$$

$f(x)$ es una función lineal. Por esto $f(x)$ puede ser representada, de acuerdo con el teorema 2, en la forma

$$f(x)=(x, a), \tag{9}$$

donde el vector a queda determinado unívocamente por la función $f(x)$, es decir, por la aplicación $\mathcal{A}$ y por el vector y. Si consideramos $\mathcal{A}$ como una aplicación dada y hacemos variar el vector y, tendremos entonces para todo vector y un vector determinado a. La aplicación que transforma y en a se indica por $\mathcal{A}^*$ y se llama *conjugada* de $\mathcal{A}$; es decir, $a=y\mathcal{A}^*$. Introduciendo este resultado en (9) y comparando con (8) obtenemos la relación

$$(x\mathcal{A}, y)=(x, y\mathcal{A}^*) \tag{10}$$

que tiene lugar para cualesquiera vectores x e y de $\mathfrak{L}$.

La propiedad (10) caracteriza plenamente la aplicación conjugada $\mathcal{A}^*$. En efecto, sea $\mathcal{B}$ una aplicación con la misma propiedad, es decir, que para cualesquiera x e y

$$(x\mathcal{A}, y)=(x, y\mathcal{B})$$

o

$$(x, y\mathcal{A}^*) - (x, y\mathcal{B}) = (x, y\mathcal{A}^* - y\mathcal{B}) = 0.$$

Esto significa que el vector $y\mathcal{A}^* - y\mathcal{B}$ es ortogonal a todo el espacio, de donde resulta que

$$y\mathcal{A}^* = y\mathcal{B}.$$

La última igualdad tiene lugar para todos los y y, por consiguiente, $\mathcal{A}^* = \mathcal{B}$.

Es fácil demostrar ahora que *la aplicación conjugada $\mathcal{A}^*$ es lineal.* Efectivamente, en virtud de (10) se tiene

$$(x, (\alpha u + \beta v)\,\mathcal{A}^*) = (x\mathcal{A},\ \alpha u + \beta v) = \bar{\alpha}\,(x\mathcal{A}, u) + \bar{\beta}\,(x\mathcal{A}, v) =$$
$$= \bar{\alpha}\,(x, u\mathcal{A}^*) + \bar{\beta}\,(x, v\mathcal{A}^*) = (x, \alpha \cdot u\mathcal{A}^* + \beta \cdot v\mathcal{A}^*)$$

cualquiera que sea el vector x. Debido a la segunda afirmación del teorema 2 esto implica la igualdad

$$(\alpha u + \beta v)\,\mathcal{A}^* = \alpha \cdot u\mathcal{A}^* + \beta \cdot v\mathcal{A}^*$$

que significa precisamente que $\mathcal{A}^*$ es lineal.

La operación de paso a la aplicación conjugada posee las propiedades siguientes:

a) $(\mathcal{A}^*)^* = \mathcal{A}$,

b) $(\alpha\mathcal{A})^* = \bar{\alpha}\mathcal{A}^*$,

c) $(\mathcal{A} + \mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^*$,

d) $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$.

Todas estas propiedades se demuestran de un mismo modo y por ello nos limitaremos a demostrar sólo una de ellas.

Por ejemplo,

$$(x\mathcal{A}\mathcal{B}, y) = (x\mathcal{A}, y\mathcal{B}^*) = (x, y\mathcal{B}^*\mathcal{A}^*),$$

es decir, $(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^*$.

Notemos aquí mismo que las aplicaciones conjugadas de las aplicaciones identidad y nula coinciden con ellas mismas. Efectivamente,

$$(x\mathcal{E}, y) = (x, y) = (x, y\mathcal{E}) \quad \text{y} \quad (x\mathcal{O}, y) = 0 = (x, y\mathcal{O}).$$

Veamos la relación que existe entre las matrices de las aplicaciones conjugadas. Tomemos en el espacio $\mathfrak{L}$ un sistema *ortonormal* de coordenadas $e_1, e_2, \ldots, e_n$ y sea

$$e_i\mathcal{A} = \alpha_{i1}e_1 + \alpha_{i2}e_2 + \ldots + \alpha_{in}e_n,$$
$$e_i\mathcal{A}^* = \beta_{i1}e_1 + \beta_{i2}e_2 + \ldots + \beta_{in}e_n \ (i = 1, \ldots, n).$$

Multiplicando estas igualdades escalarmente por e_j y teniendo en cuenta que el sistema $e_1, \ldots, e_n$ es ortogonal, obtenemos

$$(e_i\mathcal{A}, e_j)=\alpha_{ij} \quad \text{y} \quad (e_i\mathcal{A}^*, e_j)=\beta_{ij} \quad (i, j=1, \ldots, n).$$

De aquí resulta que

$$\alpha_{ij}=(e_i\mathcal{A}, e_j)=(e_i, e_j\mathcal{A}^*)=\overline{(e_j\mathcal{A}^*, e_i)}=\overline{\beta}_{ji}.$$

Por consiguiente, si A es la matriz de la aplicación $\mathcal{A}$, la matriz de la aplicación conjugada $\mathcal{A}^*$ será igual a $\overline{A}'$. Hemos obtenido el siguiente teorema.

TEOREMA 3. *Si una aplicación lineal $\mathcal{A}$ tiene en un sistema ortonormal de coordenadas la matriz A, la aplicación conjugada $\mathcal{A}^*$ tendrá en este mismo sistema la matriz conjugada transpuesta $\overline{A}'$.*

La operación del paso a las aplicaciones conjugadas posee las propiedades a), b), c) y d). En base al teorema 3 deducimos de aquí que estas mismas propiedades posee también la operación del paso a las matrices conjugadas transpuestas. Este resultado se obtiene también mediante cálculo directo (compárese con el p. 1.3).

18.3. Aplicaciones normales. Una serie de propiedades notables pueden ser obtenidas en el caso de las aplicaciones lineales de un espacio unitario que conmutan con sus conjugadas, es decir, que satisfacen la relación

$$\mathcal{A}\mathcal{A}^*=\mathcal{A}^*\mathcal{A}.$$

Estas aplicaciones se llaman *normales*. Recordando que en un sistema ortonormal de coordenadas las aplicaciones conjugadas tienen matrices conjugadas transpuestas, llegamos directamente a la conclusión de que son normales aquellas aplicaciones lineales de un espacio unitario, y sólo aquéllas, cuyas matrices, calculadas en unas bases ortonormales, satisfacen la relación

$$A\overline{A}'=\overline{A}'A.$$

Con la misma facilidad se obtienen también las siguientes propiedades de las aplicaciones normales.

TEOREMA 4. *Todo vector propio a de una aplicación normal $\mathcal{A}$ correspondiente a un valor propio ρ es al mismo tiempo un vector propio de la aplicación conjugada $\mathcal{A}^*$ pero correspondiente al valor propio conjugado de ρ.*

Tenemos

$$\mathcal{A}\mathcal{A}^*=\mathcal{A}^*\mathcal{A} \quad \text{y} \quad a(\mathcal{A}-\rho\mathcal{E})=o.$$

De aquí resulta

$$o=(a(\mathcal{A}-\rho\mathcal{E}), a(\mathcal{A}-\rho\mathcal{E}))=(a(\mathcal{A}-\rho\mathcal{E})(\mathcal{A}^*-\overline{\rho}\mathcal{E}), a)=$$
$$=(a(\mathcal{A}^*-\overline{\rho}\mathcal{E})(\mathcal{A}-\rho\mathcal{E}), a)=(a(\mathcal{A}^*-\overline{\rho}\mathcal{E}), a(\mathcal{A}^*-\overline{\rho}\mathcal{E})),$$

es decir

$$a(\mathcal{A}^* - \bar{\rho}\mathcal{E}) = o$$

que es lo que se quería demostrar.

TEOREMA 5. *Los vectores propios correspondientes a diferentes valores propios de una aplicación normal $\mathcal{A}$ son ortogonales.*

Sea

$$a\mathcal{A} = \rho a \quad \text{y} \quad b\mathcal{A} = \sigma b \quad (\rho \neq \sigma).$$

Entonces

$$\rho(a, b) = (a\mathcal{A}, b) = (a, b\mathcal{A}^*) = \sigma(a, b),$$

es decir,

$$(\rho - \sigma)(a, b) = 0 \quad \text{y} \quad (a, b) = 0.$$

Algo más compleja es la demostración del siguiente teorema principal.

TEOREMA 6. *Para toda aplicación normal $\mathcal{A}$ de un espacio unitario complejo existe una base ortonormal formada por los vectores propios de la aplicación $\mathcal{A}$; la matriz de $\mathcal{A}$ es de forma diagonal en esta base.*

Para la demostración tomamos en el espacio inicial $\mathfrak{L}$ un vector propio cualquiera $a_1 \neq o$ de la aplicación $\mathcal{A}$ e indicamos por $\mathfrak{L}_1$ el subespacio ortogonal a a_1. Si

$$a_1\mathcal{A} = \rho_1 a_1 \quad \text{y} \quad x \in \mathfrak{L}_1,$$

se tiene

$$(a_1, x\mathcal{A}) = (a_1\mathcal{A}^*, x) = \bar{\rho}_1(a_1, x) = 0,$$

es decir, el subespacio $\mathfrak{L}_1$ es invariante respecto de $\mathcal{A}$. De la invariancia de $\mathfrak{L}_1$ se deduce que en él existe un vector propio a_2 de la aplicación $\mathcal{A}$. Indiquemos por $\mathfrak{L}_2$ el subespacio formado por todos los vectores de $\mathfrak{L}$ ortogonales a a_2 y pongamos $\mathfrak{L}'_2 = \mathfrak{L}_1 \cap \mathfrak{L}_2$. Puesto que $\mathfrak{L}_1$ y $\mathfrak{L}_2$ son invariantes respecto de $\mathcal{A}$, también será invariante el espacio $\mathfrak{L}'_2$, en el cual debe existir, por consiguiente, un vector propio no nulo a_3 de la aplicación $\mathcal{A}$. Indicando por $\mathfrak{L}_3$ el conjunto de todos los vectores de $\mathfrak{L}$ ortogonales a a_3 y tomando $\mathfrak{L}'_3 = \mathfrak{L}_1 \cap \mathfrak{L}_2 \cap \mathfrak{L}_3$, obtenemos un subespacio invariante de vectores ortogonales a a_1, a_2 y a_3. Continuando el proceso encontraremos la base ortogonal requerida $a_1, a_2, \ldots, a_n$ del espacio $\mathfrak{L}$ formada por los vectores propios de la aplicación $\mathcal{A}$.

La propiedad de las aplicaciones normales de los espacios complejos establecida en el teorema 6 es característica para estas aplicaciones. Efectivamente, si la matriz A de una aplicación $\mathcal{A}$ es diagonal en una base ortonormal, la matriz $\overline{A}'$ de la aplicación conjugada también será diagonal y, por consiguiente, conmutará con A.

En los espacios unitarios reales la situación es algo diferente. Para analizarla demostremos primero una proposición general relacionada con unas aplicaciones reales lineales cualesquiera.

TEOREMA 7. *Toda aplicación lineal $\mathcal{A}$ de un espacio real no nulo tiene por lo menos un subespacio invariante de dimensión* 1 *o* 2.

Si el polinomio característico $\varphi(\lambda)$ de la aplicación $\mathcal{A}$ tiene una raíz real α, la aplicación $\mathcal{A}$ tiene en $\mathfrak{L}$ un vector propio no nulo. El subespacio tendido sobre este vector será el subespacio invariante requerido de una dimensión.

Supongamos ahora que $\varphi(\lambda)$ no tiene raíces reales. En este caso $\varphi(\lambda)$ tendrá un par de raíces conjugadas $\alpha=\rho+i\sigma$ y $\bar{\alpha}=\rho-i\sigma$, ya que los coeficientes del polinomio $\varphi(\lambda)$ son reales. Tomemos en $\mathfrak{L}$ un sistema de coordenadas $e_1, e_2, \ldots, e_n$ y sea A la matriz de la aplicación $\mathcal{A}$ en este sistema. Consideremos la ecuación

$$[\xi_1, \xi_2, \ldots, \xi_n]\, A=\alpha\, [\xi_1, \xi_2, \ldots, \xi_n], \tag{11}$$

donde $\xi_1, \xi_2, \ldots, \xi_n$ son unas incógnitas cuyos valores determinaremos en el cuerpo de los números complejos. La ecuación (11) puede ser representada en la forma

$$[\xi_1, \xi_2, \ldots, \xi_n]\,(\alpha E-A)=O,$$

donde E es la matriz unidad y O es la fila nula. Esta ecuación equivale a un sistema de ecuaciones lineales homogéneas respecto a las incógnitas $\xi_1, \ldots, \xi_n$ de matriz $(\alpha E-A)'$ (compárese con el p. 11.4). Puesto que el determinante de esta matriz es $\varphi(\alpha)$ y, por consiguiente, es igual a cero, la ecuación (11) tiene en el cuerpo de los números complejos una solución no nula que indicaremos por las mismas letras $\xi_1, \ldots, \xi_n$. Tomemos para abreviar

$$[\xi_1, \xi_2, \ldots, \xi_n]=x,$$

de modo que la relación (11) se convierte en

$$xA=\alpha x. \tag{12}$$

Pasando aquí a los números complejos conjugados, obtenemos $\bar{x}\,\bar{A}=\bar{\alpha}\,\bar{x}$. Pero los elementos de la matriz A son reales y, por lo tanto, $\bar{A}=A$ y

$$\bar{x}A=\bar{\alpha}\,\bar{x}. \tag{13}$$

Puesto que las filas $x+\bar{x}$ e $i(x-\bar{x})$ son reales, en el espacio $\mathfrak{L}$ existen unos vectores a y b, cuyas filas coordenadas serán iguales respectivamente a

$$\left.\begin{aligned} [a]&=\frac{1}{2}(x+\bar{x}),\\ [b]&=\frac{1}{2i}(x-\bar{x}). \end{aligned}\right\} \tag{14}$$

Expresando aquí x y $\bar{x}$ en términos de $[a]$ y $[b]$ y empleando (12) y (13), llegamos a las relaciones

$$\begin{aligned}[a]\, A &= \rho\, [a] - \sigma\, [b], \\ [b]\, A &= \sigma\, [a] + \rho\, [b]\end{aligned} \quad (\alpha = \rho + i\sigma).$$

Estas relaciones equivalen a las igualdades

$$\left.\begin{aligned} a\mathcal{A} &= \rho a - \sigma b, \\ b\mathcal{A} &= \sigma a + \rho b \end{aligned}\right\} \tag{15}$$

que muestran que el subespacio tendido sobre los vectores a y b es invariante respecto a la aplicación $\mathcal{A}$. Hemos demostrado el teorema.

Supongamos ahora que el espacio considerado $\mathfrak{L}$ es un espacio unitario real y que $\mathcal{A}$ es una aplicación normal del mismo, cuyo polinomio, característico $\varphi(\lambda)$ tiene dos raíces conjugadas $\rho + i\sigma$ y $\rho - i\sigma$ $(\sigma \neq 0)$. Tomando para el sistema de coordenadas $e_1, \ldots, e_n$ de $\mathfrak{L}$ una base *ortonormal* cualquiera y repitiendo el razonamiento anterior, obtendremos de nuevo en $\mathfrak{L}$ unos vectores a y b ligados por las relaciones (15). Probemos que los vectores a y b serán ahora *ortogonales*. En efecto, para determinar estos vectores hemos tenido que considerar el espacio de filas $\tilde{\mathfrak{L}}$ sobre el cuerpo de los números complejos. Podemos aceptar que este espacio de filas es unitario, tomando, de acuerdo con el p. 17.1, para el producto escalar de las filas $[\xi_1, \ldots, \xi_n]$ y $[\eta_1, \ldots, \eta_n]$ la expresión $\xi_1\bar{\eta}_1 + \ldots + \xi_n\bar{\eta}_n$. Las filas $[e_1], \ldots, [e_n]$ forman una base ortonormal de $\tilde{\mathfrak{L}}$. La aplicación consistente en la multiplicación de las filas por una matriz A será una aplicación lineal y su matriz coincidirá con A en la base señalada. Puesto que $\bar{A} = A$, la aplicación considerada será normal y las filas x y $\bar{x}$, determinadas durante la demostración del teorema 7, serán unos vectores propios correspondientes a diferentes valores propios $\rho + i\sigma$ y $\rho - i\sigma$. En virtud del teorema 5 tenemos $(x, \bar{x}) = 0$, de donde

$$([a], [b]) = \frac{1}{4i}(x + \bar{x},\ x - \bar{x}) = 0,$$

es decir, $(a, b) = 0$. Tomando para x un vector de longitud $\sqrt{2}$, obtendremos de (14) que a y b serán de longitud 1.

Demostremos además que el subespacio de $\mathfrak{L}$ ortogonal a a y b será invariante respecto de $\mathcal{A}$. En el espacio de filas $\tilde{\mathfrak{L}}$ el subespacio ortogonal a $[a]$ y $[b]$ coincide con el subespacio $\tilde{\mathfrak{L}}_1$ ortogonal a x y $\bar{x}$. Este último subespacio es invariante respecto de $\mathcal{A}$, ya que es la intersección de subespacios ortogonales a los vectores propios x y $\bar{x}$ de una aplicación normal. Sea ahora $y \in \mathfrak{L}$ y $(a, y) = (b, y) = 0$;

entonces $[y] \in \bar{\mathfrak{L}}_1$ e $[y] A \in \bar{\mathfrak{L}}_1$, de donde resulta

$$(y\mathcal{A},\ a) = ([y]\ A,\ [a]) = 0 \quad \text{y} \quad (y\mathcal{A}, b) = 0$$

que es lo que se quería demostrar.

La aplicación $\mathcal{A}_1$ inducida por la aplicación $\mathcal{A}$ en el espacio de dos dimensiones tendido sobre los vectores a y b tiene, en virtud de (15), la matriz

$$A_1 = \begin{bmatrix} \rho & -\sigma \\ \sigma & \rho \end{bmatrix}.$$

Tomando para el número complejo $\rho + i\sigma$ la expresión $r(\cos\varphi - i \operatorname{sen}\varphi)$, podemos representar la matriz A_1 en la forma

$$A_1 = r \begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi \\ -\operatorname{sen}\varphi & \cos\varphi \end{bmatrix}.$$

Por consiguiente, $\mathcal{A}_1$ es el producto de la aplicación de matriz rE y de la aplicación de matriz $\begin{bmatrix} \cos\varphi & \operatorname{sen}\varphi \\ -\operatorname{sen}\varphi & \cos\varphi \end{bmatrix}$. La primera es una aplicación de semejanza de centro en el origen de coordenadas y de coeficiente de dilatación igual a r y la segunda representa, como puede verse fácilmente, un giro de ángulo φ de los vectores alrededor del origen de coordenadas en el plano determinado por a y b.

TEOREMA 8. *Para toda aplicación normal $\mathcal{A}$ de un espacio real unitario $\mathfrak{L}$ existe en $\mathfrak{L}$ una base ortonormal en la que la matriz de la aplicación $\mathcal{A}$ tiene la forma*

$$A = \begin{bmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_k \end{bmatrix} \dot{+} r_1 \begin{bmatrix} \cos\varphi_1 & \operatorname{sen}\varphi_1 \\ -\operatorname{sen}\varphi_1 & \cos\varphi_1 \end{bmatrix} \dot{+} \dots$$

$$\dots \dot{+} r_m \begin{bmatrix} \cos\varphi_m & \operatorname{sen}\varphi_m \\ -\operatorname{sen}\varphi_m & \cos\varphi_m \end{bmatrix}, \qquad (16)$$

donde los números k y m pueden ser iguales a cero.

La demostración es casi idéntica a la demsotración del teorema análogo 6. La diferencia estriba sólo en que no podemos afirmar ahora que todo subespacio $\mathfrak{L}'_i$ invariante respecto de $\mathcal{A}$ contiene un vector propio no nulo a_{i+1} de la aplicación $\mathcal{A}$. Pero si $\mathfrak{L}'_i$ no contiene vectores propios de la aplicación $\mathcal{A}$, en $\mathfrak{L}'_i$ existe, en virtud del teorema 7, un par de vectores recíprocamente ortogonales a_{i+1} y b_{i+1} ligados por relaciones de tipo (15). Para el subespacio $\mathfrak{L}_{i+1}$ se puede tomar entonces el conjunto de vectores de $\mathfrak{L}$ ortogonales a a_{i+1} y b_{i+1}. De las observaciones hechas anteriormente resulta que $\mathfrak{L}_{i+1}$ será invariante respecto de $\mathcal{A}$ y el proceso se puede continuar después según el esquema expuesto en la demostración del teorema 6. Así obtendremos una base ortonormal de $\mathfrak{L}$ compuesta por los vectores $a_1, \ldots, a_k, a_{k+1}, b_{k+1}, \ldots, a_{k+m}, b_{k+m}$ ligados por relaciones de tipo

$$a_p\mathcal{A} = \alpha_p a_p \qquad (p = 1, \ldots, k),$$

$$\left.\begin{array}{l} a_q\mathcal{A} = \rho_q a_q - \sigma_q b_q \\ b_q\mathcal{A} = \sigma_q a_q + \rho_q b_q \end{array}\right\} \qquad (q = 1, \ldots, m, \ \sigma_q \neq 0);$$

estas últimas muestran que la matriz de la aplicación $\mathcal{A}$ tendrá en la base señalada precisamente la forma (16).

Ejemplos y problemas

1. Tomemos en el espacio euclídeo corriente de tres dimensiones $\mathfrak{R}$ una recta orientada que pasa por el origen de coordenadas e indiquemos por $f(x)$ la longitud de la proyección del vector x sobre esta recta tomada con el signo correspondiente. Pruébese que $f(x)$ es una función lineal y que toda función lineal sobre el espacio $\mathfrak{R}$ es de la forma $\alpha f(x)$, donde $\alpha \geqslant 0$ y el eje de proyección se ha escogido convenientemente (compárese con el teorema 2).

2. Una función $f(x)$ definida sobre un espacio lineal complejo se llama *antilineal*, si $f(x+y) = f(x) + f(y)$ y $f(\alpha x) = \bar{\alpha} f(x)$. Pruébese que toda función antilineal sobre un espacio unitario es de la forma (a, x).

3. Demuéstrese que la correspondencia establecida en el problema anterior entre los vectores de un espacio unitario $\mathfrak{L}$ y las funciones antilineales es un isomorfismo entre $\mathfrak{L}$ y el espacio de todas las funciones antilineales sobre $\mathfrak{L}$.

4. En un espacio unitario $\mathfrak{L}$ con una base no ortonormal a_1, a_2 y a_3 están dadas dos aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 1 & 2 & -2 \\ 2 & -1 & 0 \\ -2 & 1 & 0 \end{bmatrix}.$$

Hállese una base *ortonormal* de $\mathfrak{L}$, si se sabe que $\mathcal{A}$ y $\mathcal{B}$ son normales y que el vector a_1 es de longitud 1.

5. Pruébese que sobre un espacio unitario complejo se puede extraer la raíz de cualquier grado natural m de toda aplicación normal, es decir, que para toda aplicación normal $\mathcal{A}$ existe una aplicación $\mathcal{B}$, también normal, tal que $\mathcal{B}^m = \mathcal{A}$. ¿Cuál es el número máximo de estas aplicaciones $\mathcal{B}$?

6. Demuéstrese que siendo $\mathfrak{A}$ un subespacio invariante de una aplicación normal $\mathcal{A}$, el complemento ortogonal $\mathfrak{A}^{\perp}$ también será invariante respecto de $\mathcal{A}$. Si el espacio principal es complejo, la propiedad indicada caracteriza las aplicaciones normales.

§ 19. Aplicaciones unitarias y simétricas

19.1. Aplicaciones unitarias. Una aplicación isomorfa de un espacio unitario sobre sí mismo se llama aplicación *unitaria* de este espacio. Con más detalle: una aplicación lineal regular $\mathcal{U}$ de un espacio unitario $\mathfrak{L}$ se llama unitaria, si no altera el valor del producto escalar, es decir, si para todos los a y b de $\mathfrak{L}$ tiene lugar la relación

$$(a, b) = (a\mathcal{U}, b\mathcal{U}). \tag{1}$$

Las rotaciones del espacio euclídeo corriente de tres dimensiones alrededor del origen de coordenadas O representan el ejemplo más sencillo de aplicaciones unitarias. Las reflexiones especulares de este espacio respecto a un plano cualquiera que pasa por O representan otro ejemplo de aplicaciones unitarias del espacio corriente. Se puede probar fácilmente que con las combinaciones de estos dos tipos de aplicaciones se agotan todas las aplicaciones unitarias del espacio corriente. Por esto las aplicaciones unitarias de espacios se pueden considerar como las aplicaciones análogas a las rotaciones y a las reflexiones especulares del espacio euclídeo corriente.

De la igualdad (1), que caracteriza las aplicaciones unitarias, se deduce que

$$(x,\ y) = (x\mathcal{U},\ y\mathcal{U}) = (x,\ y\mathcal{U}\mathcal{U}^*),$$

de donde resulta que

$$\mathcal{U}\mathcal{U}^* = \mathcal{E},\ \mathcal{U}^* = \mathcal{U}^{-1} \text{ y } \mathcal{U}^*\mathcal{U} = \mathcal{E}. \tag{2}$$

Recíprocamente, de las relaciones (2) se deduce que $\mathcal{U}$ es invertible y que

$$(x,\ y) = (x,\ y\mathcal{U}\mathcal{U}^*) = (x\mathcal{U},\ y\mathcal{U}).$$

Por consiguiente, *una aplicación lineal $\mathcal{U}$ es unitaria cuando, y solo cuando, la aplicación conjugada $\mathcal{U}^*$ coincide con la inversa $\mathcal{U}^{-1}$.*

En particular, las relaciones (2) muestran que *las aplicaciones unitarias son aplicaciones normales en el sentido del punto anterior.*

Tomemos en un espacio $\mathfrak{L}$ un sistema ortonormal de coordenadas y sea $\mathcal{U}$ una aplicación unitaria de este espacio. Si la matriz de la aplicación $\mathcal{U}$ es igual a U, la matriz de la aplicación conjugada es igual a $\overline{U}'$ de acuerdo con el p. 18.2. De la relación (2) se desprende por lo tanto que

$$U\overline{U}' = E. \tag{3}$$

Recíprocamente, si la matriz U de una aplicación lineal $\mathcal{U}$ satisface en un sistema ortonormal de coordenadas la relación (3), la propia aplicación $\mathcal{U}$ satisface las relaciones (2) y es, por consiguiente, unitaria. Las matrices U que satisfacen la relación (3) se llaman *unitarias* (p. 1.3); llegamos, por lo tanto, al resultado siguiente: *toda aplicación unitaria tiene en un sistema ortonormal de coordenadas una matriz unitaria; recíprocamente, si la matriz de una aplicación lineal es unitaria en un sistema ortonormal de coordenadas, la propia aplicación es unitaria.*

Siendo real el campo principal K las matrices de las aplicaciones también son reales y la relación (3) se convierte en

$$UU' = E. \tag{4}$$

Las matrices que satisfacen esta relación han sido llamados en el p. 1.3. ortogonales. Es decir, toda aplicación real unitaria tiene en una base ortonormal la matriz ortogonal. Recíprocamente, si en una base ortonormal la matriz de una aplicación lineal de un espacio real unitario es ortogonal, la aplicación es unitaria.

Las aplicaciones unitarias no alteran, por definición, los valores de los productos escalares. De aquí resulta que las aplicaciones unitarias no alteran las longitudes de los vectores.

La última propiedad es característica para las aplicaciones unitarias: *si una aplicación lineal $\mathfrak{U}$ de un espacio unitario $\mathfrak{L}$ no altera las longitudes de los vectores, la aplicación $\mathfrak{U}$ es unitaria.*

En efecto, sean a y b unos vectores arbitrarios de $\mathfrak{L}$. Tomemos

$$a\mathfrak{U} = a' \text{ y } b\mathfrak{U} = b'.$$

Puesto que la aplicación $\mathfrak{U}$ es lineal, tenemos

$$(a+\alpha b)\,\mathfrak{U} = a' + \alpha b'$$

para cualquier valor α del campo de coeficientes. Por hipótesis, la aplicación $\mathfrak{U}$ no altera las longitudes y por lo tanto

$$(a+\alpha b,\ a+\alpha b) = (a'+\alpha b',\ a'+\alpha b').$$

Realizando aquí la multiplicación y reduciendo los términos semejantes, obtenemos

$$\alpha\,(b,\ a) + \bar{\alpha}\,(a,\ b) = \alpha\,(b',\ a') + \bar{\alpha}\,(a',\ b'). \tag{5}$$

Para $\alpha = 1$ esta igualdad se convierte en

$$(b,\ a) + (a,\ b) = (b',\ a') + (a',\ b'). \tag{6}$$

Si el campo principal es real, tenemos $(b,\ a) = (a,\ b)$ y de (6) resulta $(a,\ b) = (a',\ b')$. En cambio, si K no es real, tomando en (5) $\alpha = i$ y simplificando en i, llegamos a la relación

$$(b,\ a) - (a,\ b) = (b',\ a') - (a',\ b')$$

que con (6) da de nuevo $(a,\ b) = (a',\ b')$. Luego, tenemos en ambos casos

$$(a,\ b) = (a',\ b') = (a\mathfrak{U},\ b\mathfrak{U}),$$

es decir, la aplicación $\mathfrak{U}$ es unitaria.

Examinemos el problema acerca de la transformación de coordenadas en los espacios unitarios. Sea $e_1, e_2, \ldots, e_n$ una base ortonormal de un espacio unitario $\mathfrak{L}$ y sea $\mathfrak{U}$ una aplicación unitaria cualquiera de este espacio. Puesto que una aplicación unitaria no altera las longitudes de los vectores y transforma los vectores ortogonales en ortogonales, el sistema $e_1\mathfrak{U}, e_2\mathfrak{U}, \ldots, e_n\mathfrak{U}$ será de nuevo una base ortonormal de $\mathfrak{L}$. Recíprocamente, supongamos que una aplicación lineal $\mathfrak{U}$ transforma una base ortonormal $e_1, e_2, \ldots, e_n$ en una base también ortonormal $e_1\mathcal{U}, e_2\mathcal{U}, \ldots, e_n\mathcal{U}$. Tomemos en $\mathfrak{L}$ unos vectores arbitrarios

$$a=\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_n e_n \text{ y } b=\beta_1 e_1+\beta_2 e_2+\ldots+\beta_n e_n.$$

Tenemos

$$(a, b)=(\sum \alpha_i e_i, \sum \beta_i e_i)=\alpha_1\overline{\beta}_1+\alpha_2\overline{\beta}_2+\ldots+\alpha_n\overline{\beta}_n,$$
$$(a\mathcal{U}, b\mathcal{U})=(\sum \alpha_i e_i\mathcal{U}, \sum \beta_i e_i\mathcal{U})=\alpha_1\overline{\beta}_1+\alpha_2\overline{\beta}_2+\ldots+\alpha_n\overline{\beta}_n,$$

es decir, $(a, b)=(a\mathcal{U}, b\mathcal{U})$; luego, la aplicación $\mathcal{U}$ es unitaria. Por consiguiente, *para que una aplicación lineal $\mathcal{U}$ sea unitaria es necesario y suficiente que $\mathcal{U}$ transforme una base ortonormal en una base ortonormal.*

De aquí también se desprende directamente la proposición siguiente: *la matriz del cambio de una base ortonormal por otra es unitaria y, viceversa, si una de las bases es ortonormal y la matriz del cambio es unitaria, la otra base es también ortonormal.*

Para la demostración basta observar que la matriz del cambio de un sistema de coordenadas por otro coincide con la matriz de la aplicación lineal que transforma el primer sistema en el segundo (p. 8.3).

Notemos, finalmente, que de la definición de las aplicaciones unitarias se deducen fácilmente las siguientes propiedades de las mismas: 1) *la aplicación idéntica es unitaria,* 2) *el producto de aplicaciones unitarias es unitario* y 3) *la inversa de una aplicación unitaria es una aplicación unitaria.*

19.2. Equivalencia unitaria. Ligado de un modo natural al concepto de aplicación unitaria aparece uno de los problemas principales de la teoría de espacios unitarios. Se trata de la clasificación de las aplicaciones lineales de estos espacios. Sean $\mathfrak{L}$ y $\mathfrak{L}_1$ unos espacios unitarios sobre un mismo campo principal K. Consideremos dos aplicaciones lineales $\mathcal{A}$ y $\mathcal{A}_1$ de estos espacios con la particularidad de que $\mathcal{A}$ actúa sobre $\mathfrak{L}$, mientras que $\mathcal{A}_1$ actúa sobre $\mathfrak{L}_1$. Las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ se llaman *semejantes o isomorfas* si existe una aplicación isomorfa de $\mathfrak{L}$ sobre $\mathfrak{L}_1$ que transforme la aplicación $\mathcal{A}$ en la aplicación $\mathcal{A}_1$. Puesto que todos los espacios unitarios son, salvo un isomorfismo, conocidos y se determinan por su dimensión n, podemos aceptar que $\mathfrak{L}_1$ coincide con $\mathfrak{L}$ y, que, por consiguiente, las aplicaciones $\mathcal{A}$ y $\mathcal{A}_1$ actúan sobre el mismo espacio $\mathfrak{L}$. En este caso nuestra definición significa que $\mathcal{A}$ y $\mathcal{A}_1$ son isomorfas cuando, y sólo cuando, existe una aplicación isomorfa $\mathcal{U}$ del espacio $\mathfrak{L}$ sobre sí mismo que transforma $\mathcal{A}$ en $\mathcal{A}_1$. Como hemos visto en el p. 10.2 esto equivale a la condición

$$\mathcal{A}_1=\mathcal{U}^{-1}\mathcal{A}\mathcal{U}. \quad (7)$$

Si tomamos en $\mathfrak{L}$ un sistema ortonormal de coordenadas, la relación (7) puede ser representada en la forma matricial

$$A_1 = U^{-1}AU, \tag{8}$$

donde U es una matriz unitaria y A y A_1 son las matrices de las aplicaciones lineales dadas. Unas matrices A y A_1 que satisfacen la relación (8) se llaman *unitariamente equivalentes*. Por consiguiente, *las aplicaciones lineales de un espacio unitario son isomorfas cuando, y sólo cuando, sus matrices, calculadas en una base ortonormal, son unitariamente equivalentes.*

Considerando la matriz U de la relación (8) como una matriz de cambio, llegamos a la siguiente proposición: *unas aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de un espacio unitario $\mathfrak{L}$ son isomorfas cuando, y sólo cuando, en $\mathfrak{L}$ existen dos bases ortonormales tales que la matriz de la aplicación $\mathcal{A}$ calculada en una de ellas coincide con la matriz de la aplicación $\mathcal{B}$ calculada en la otra.*

Ésta proposición es totalmente análoga a la correspondiente afirmación para espacios lineales arbitrarios que ha sido considerada en el p. 10.2, donde se ha dado también una demostración detallada.

A título de ejemplo, podemos tomar las aplicaciones normales. Está claro que para que unas aplicaciones lineales de un espacio unitario sean *unitariamente* isomorfas es necesario que sean isomorfas *linealmente*, es decir, que sean isomorfas como aplicaciones lineales de un espacio lineal. Por ello, para que unas aplicaciones lineales sean unitariamente isomorfas es necesario que sus polinomios característicos coincidan. Si las aplicaciones dadas son normales y las raíces de sus polinomios característicos son conocidas, es posible, según el punto anterior, escribir las matrices de estas aplicaciones en unas bases ortonormales del espacio convenientemente escogidas. Puesto que estas matrices resultarán idénticas, las aplicaciones serán unitariamente isomorfas. Hemos demostrado, pues, el siguiente teorema:

TEOREMA 1. *Para que unas aplicaciones normales de espacios unitarios, tanto reales como complejos, sean unitariamente isomorfas es necesario y suficiente que los polinomios característicos de estas aplicaciones coincidan.*

En una forma puramente matricial el teorema 1 puede ser enunciado del siguiente modo:

TEOREMA 1a. *Para la equivalaencia unitaria de unas matrices A y B que conmutan con sus matrices anticonjugadas complejas $\bar{A}'$ y $\bar{B}'$ es necesario y suficiente, tanto en el caso del cuerpo de los números complejos como en el caso del cuerpo de los números reales, que los polinomios característicos de las matrices A y B coincidan.*

En particular, para toda matriz compleja o real A, tal que $A\bar{A}' = A'\bar{A}$, existe una matriz unitaria U compleja o, respectivamente, real tal que la matriz $UAU^{-1} = UA\overline{U}'$ es diagonal en el caso complejo y tiene la forma (16) del p. 18.3 en el caso real.

Por consiguiente, el isomorfismo unitario de las aplicaciones normales resulta ser equivalente al isomorfismo corriente de las mismas.

19.3. Forma normal de la matriz de una aplicación unitaria. Como ya hemos señalado, las aplicaciones unitarias son un caso particular de las normales. Por esto los teoremas 1 y 1a, además de ofrecer las condiciones necesarias y suficientes para que un isomorfismo sea unitario, ofrecen también la forma normal para las matrices de las aplicaciones unitarias. El teorema que sigue indica el rasgo característico que distingue las aplicaciones unitarias de las demás aplicaciones normales:

TEOREMA 2. *El módulo de todas las raíces del polinomio característico de una aplicación unitaria es igual a la unidad.*

Consideremos primero el caso complejo. En este caso a toda raíz α del polinomio característico de una aplicación unitaria $\mathcal{U}$ le corresponde un vector propio no nulo a. De las relaciones

$$a\mathcal{U}=\alpha a \text{ y } (a\mathcal{U},\ a\mathcal{U})=(a,\ a)$$

resulta

$$(\alpha a,\ \alpha a)=\alpha\bar{\alpha}\,(a,\ a)=(a,\ a),$$

es decir, $\alpha\bar{\alpha}=|\alpha|^2=1$.

En el caso real, a toda raíz compleja $\alpha=\rho+i\sigma$ le corresponde un par a y b de vectores ortogonales no nulos tales que

$$a\mathcal{U}=\rho a-\sigma b \text{ y } b\mathcal{U}=\sigma a+\rho b.$$

De aquí tenemos

$$(a,\ a)+(b,\ b)=(a\mathcal{U},\ a\mathcal{U})+(b\mathcal{U},\ b\mathcal{U})=(\rho^2+\sigma^2)\,((a,\ a)+(b,\ b)),$$

es decir,

$$\rho^2+\sigma^2=|\alpha|^2=1.$$

Observando que entre los números reales sólo los números 1 y -1 son de módulo igual a la unidad, podemos enunciar el teorema 1a en el caso de aplicaciones unitarias en la forma siguiente:

TEOREMA 3. *Para toda matriz unitaria A existe una matriz unitaria compleja U tal que UAU^{-1} será una matriz diagonal con elementos diagonales de módulo igual a la unidad. Para toda matriz unitaria real A existe una matriz unitaria real U tal que*

$$UAU^{-1}=E_s+(-E_t)+\begin{bmatrix}\cos\varphi_1 & \operatorname{sen}\varphi_1\\ -\operatorname{sen}\varphi_1 & \cos\varphi_1\end{bmatrix}+\ \cdots$$
$$\cdots+\begin{bmatrix}\cos\varphi_m & \operatorname{sen}\varphi_m\\ -\operatorname{sen}\varphi_m & \cos\varphi_m\end{bmatrix}, \qquad (9)$$

donde E_s y E_t son matrices unidades de orden s y t, respectivamente, y $\operatorname{sen}\varphi_j \neq 0$ $(j=1, \ldots, m)$, *con la particularidad de que algunos de los números s, t y m pueden ser iguales a cero, es decir, en la fórmula* (9) *pueden no figurar los términos correspondientes.*

Consideremos, a título de ejemplo, el espacio euclídeo corriente de tres dimensiones $\mathfrak{R}$. Para toda aplicación ortogonal $\mathcal{U}$ del espacio $\mathfrak{R}$ se puede encontrar, de acuerdo con el teorema 3, un sistema e_1, e_2 y e_3 de vectores perpendiculares unitarios tal que la matriz de la aplicación $\mathcal{U}$ tomará una de las seis formas siguientes:

$$\alpha)\ \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix},\quad \beta)\ \begin{bmatrix} -1 & & \\ & 1 & \\ & & 1 \end{bmatrix},\quad \gamma)\ \begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix},$$

$$\delta)\ \begin{bmatrix} -1 & & \\ & -1 & \\ & & -1 \end{bmatrix},\quad \varepsilon)\ \begin{bmatrix} 1 & & \\ & \cos\varphi & \operatorname{sen}\varphi \\ & -\operatorname{sen}\varphi & \cos\beta \end{bmatrix}\quad \text{y}\quad \varkappa)\ \begin{bmatrix} -1 & & \\ & \cos\varphi & \operatorname{sen}\varphi \\ & -\operatorname{sen}\varphi & \cos\varphi \end{bmatrix}.$$

Es evidente que la aplicación $\mathcal{U}$ es la aplicación idéntica en el caso α), la reflexión especular respecto al plano e_2Oe_3 en el caso β), la reflexión especular respecto a la recta Oe_1 en el caso γ), la reflexión especular respecto al origen O en el caso δ), la rotación de ángulo φ alrededor del eje e_1 en el caso ε) y la rotación de ángulo φ alrededor del eje e_1 seguida de la reflexión especular respecto al plano e_2Oe_3 en el caso $\varkappa$). Los cuatro primeros casos se pueden considerar como casos particulares de los dos últimos con $\varphi=0$ y $\varphi=\pi$.

19.4. Aplicaciones simétricas. Una aplicación lineal $\mathcal{A}$ de un espacio unitario $\mathfrak{L}$ se llama *hermitiana o simétrica*, si $\mathcal{A}$ coincide con su aplicación conjugada $\mathcal{A}^*$. Es decir, si la aplicación $\mathcal{A}$ es simétrica, se tiene

$$(x\mathcal{A},\ y)=(x,\ y\mathcal{A}). \tag{10}$$

Recíprocamente, si una aplicación lineal $\mathcal{A}$ satisface la condición (10) cualesquiera que sean x e y de $\mathfrak{L}$, la aplicación $\mathcal{A}$ es simétrica.

Es evidente que de la condición $\mathcal{A}^*=\mathcal{A}$ se deduce la igualdad $\mathcal{A}\mathcal{A}^*=\mathcal{A}^*\mathcal{A}$, es decir, *las aplicaciones simétricas son aplicaciones normales.*

Tomemos en $\mathfrak{L}$ una base ortonormal y sea A la matriz de una aplicación simétrica $\mathcal{A}$. La matriz de la aplicación conjugada $\mathcal{A}^*$ es igual en esta base a la matriz anticonjugada $\overline{A'}$. Tenemos, por hipótesis, $\mathcal{A}^*=\mathcal{A}$, de donde

$$\overline{A'}=A. \tag{11}$$

Recíprocamente, de (11) se deduce que $\mathcal{A}^*=\mathcal{A}$, es decir, que $\mathcal{A}$ es simétrica. Las matrices que satisfacen la relación (11) han sido llamadas en el p. 1.3 hermitianas. Por consiguiente, *en una base ortonormal a las aplicaciones simétricas les corresponden matrices hermitianas y, viceversa, a las matrices hermitianas les corresponden las aplicaciones simétricas.*

El ejemplo más sencillo de aplicación simétrica es la aplicación de tipo $\alpha\mathcal{E}$, donde α es real. El ejemplo general se obtiene con las aplicaciones cuyas matrices tienen, en una base ortonormal, la forma diagonal real.

La suma de aplicaciones simétricas y el producto de una aplicación simétrica por un número real son de nuevo aplicaciones simétricas.

En efecto, si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ son simétricas y α es un número real, se tiene

$$(\mathcal{A}+\mathcal{B})^* = \mathcal{A}^* + \mathcal{B}^* = \mathcal{A}+\mathcal{B},$$
$$(\alpha\mathcal{A})^* = \bar{\alpha}\mathcal{A}^* = \alpha\mathcal{A}.$$

El producto de dos aplicaciones simétricas es una aplicación simétrica cuando, y sólo cuando, estas aplicaciones son conmutables.

Efectivamente, de $\mathcal{A}\mathcal{B}=\mathcal{B}\mathcal{A}$, $\mathcal{A}=\mathcal{A}'$ y $\mathcal{B}=\mathcal{B}^*$ se deduce que

$$(\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^* = \mathcal{B}\mathcal{A} = \mathcal{A}\mathcal{B}.$$

Recíprocamente, si $(\mathcal{A}\mathcal{B})^* = \mathcal{A}\mathcal{B}$, $\mathcal{A}=\mathcal{A}^*$ y $\mathcal{B}=\mathcal{B}^*$, se tiene

$$\mathcal{A}\mathcal{B} = (\mathcal{A}\mathcal{B})^* = \mathcal{B}^*\mathcal{A}^* = \mathcal{B}\mathcal{A}.$$

De aquí se deduce, en particular, que las potencias de una aplicación simétrica y, en general, los polinomios de coeficientes reales en una aplicación simétrica son de nuevo aplicaciones simétricas.

Los valores propios de las aplicaciones simétricas son reales.

Efectivamente, sea $\mathcal{A}$ una aplicación simétrica, sea α un valor propio de la misma y sea a un vector propio no nulo correspondiente. Tenemos

$$(a, a\mathcal{A}) = (a, \alpha a) = \bar{\alpha}(a, a),$$
$$(a\mathcal{A}, a) = (\alpha a, a) = \alpha(a, a).$$

Pero

$$(a, a\mathcal{A}) = (a\mathcal{A}, a).$$

Comparando estos resultados, vemos que $\bar{\alpha}=\alpha$, es decir, que α es real.

Todas las raíces del polinomio característico de una matriz hermitiana son reales.

En efecto, toda matriz hermitiana A puede ser considerada como la matriz de una aplicación simétrica $\mathcal{A}$ de un espacio unitario. Las raíces del polinomio característico de la matriz A son los valores propios de la aplicación $\mathcal{A}$ y, por consiguiente, son reales.

Hemos señalado anteriormente que las aplicaciones simétricas son normales. Por lo tanto, para que unas aplicaciones simétricas sean unitariamente isomorfas es necesario y suficiente, en virtud del teorema 1, que los polinomios característicos de estas aplicaciones coincidan.

Para toda aplicación normal de un espacio unitario complejo existe, según el teorema 6 del p. 18.3, una base ortonormal en la que la matriz de la aplicación es de forma diagonal. Puesto que todos los valores propios de las aplicaciones simétricas son reales, la matriz diagonal indicada será real en el caso de aplicaciones simétricas; obtenemos así el teorema siguiente:

TEOREMA 4 *Toda aplicación simétrica de un espacio complejo unitario tiene, en una base ortonormal adecuada, una matriz diagonal real.*

La recíproca es támbién válida, ya que si la matriz de una aplicación lineal $\mathcal{A}$ en una base ortonormal es de forma diagonal real A, se tiene $\overline{A'} = A$ y, por consiguiente, $\mathcal{A}^* = \mathcal{A}$.

En términos matriciales el teorema 4 puede ser enunciado de la forma siguiente.

TEOREMA 4a. *Para toda matriz hermitiana A existe una matriz unitaria compleja U tal que la matriz UAU^{-1} es de forma diagonal real.*

Consideremos ahora el caso en que $\mathcal{A}$ es una aplicación simétrica de un espacio unitario real. Según el teorema 8 del p. 18.3, la matriz de la aplicación $\mathcal{A}$ se descompone, en una base ortonormal adecuada, en células de orden 1 ó 2. Además, las células de orden 2 aparecen sólo cuando el polinomio característico de la aplicación tiene raíces no reales. Pero los polinomios característicos de las aplicaciones simétricas no tienen raíces que no sean reales. Por consiguiente, también en el caso real la matriz de una aplicación simétrica se reduce a la forma diagonal en una base ortonormal adecuada. La recíproca, obviamente, es también válida, de modo que tiene lugar el teorema siguiente:

TEOREMA 5. *Para toda aplicación simétrica de un espacio unitario real existe una base ortonormal en la que la matriz de la aplicación adquiere la forma diagonal.*

En una base ortonormal las matrices de las aplicaciones simétricas satisfacen las relaciones $\overline{A'} = A$. En el caso real está relación se convierte en la igualdad $A' = A$. Las matrices que satisfacen esta igualdad se llaman simétricas (p. 1.3). Por consiguiente, en una base ortonormal las matrices de las aplicaciones simétricas reales son simétricas y, viceversa, las aplicaciones son simétricas si sus matrices son simétricas reales. Esta observación permite enunciar el teorema 5 del modo siguiente.

TEOREMA 5a. *Para toda matriz simétrica real A existe una matriz unitaria real U tal que la matriz UAU^{-1} es de forma diagonal.*

19.5. Aplicaciones antisimétricas. Sea $\mathfrak{L}$ un espacio unitario. Una aplicación lineal $\mathcal{A}$ se llama *antisimétrica*, si está ligada a su aplicación conjugada por la relación

$$\mathcal{A}^* = -\mathcal{A}. \tag{12}$$

Siendo a y b vectores arbitrarios de $\mathfrak{L}$, de (12) se deduce que

$$(a, b\mathcal{A}) = (a\mathcal{A}^*, b) = -(a\mathcal{A}, b).$$

Recíprocamente, si para cualesquiera a y b se tiene

$$(a, b\mathcal{A}) = -(a\mathcal{A}, b),$$

resulta que $\mathcal{A} = -\mathcal{A}^*$ y $\mathcal{A}$ es una aplicación antisimétrica.

En el caso en el que el campo principal es el cuerpo de los números complejos, las aplicaciones antisimétricas tienen una expresión muy sencilla mediante las simétricas. En efecto, sea $\mathcal{A}$ una aplicación simétrica de un espacio $\mathfrak{L}$. Entonces

$$(i\mathcal{A})^* = \bar{i}\mathcal{A}^* = -i\mathcal{A},$$

es decir, la aplicación $i\mathcal{A}$ es antisimétrica. Recíprocamente, si la aplicación $\mathcal{A}$ es antisimétrica, se tiene

$$(i\mathcal{A})^* = \bar{i}\mathcal{A}^* = i\mathcal{A},$$

y, por consiguiente, la aplicación $i\mathcal{A}$ es simétrica. En los espacios reales esta relación desaparece.

Tomemos una base ortonormal en un espacio unitario $\mathfrak{L}$ e indiquemos por A la matriz de una aplicación antisimétrica $\mathcal{A}$. Puesto que la matriz de la aplicación conjugada $\mathcal{A}^*$ es igual a $\overline{A'}$, la condición (12) da

$$\overline{A'} = -A. \tag{13}$$

Viceversa, de (13) se deduce, obviamente, que $\mathcal{A}$ es una aplicación antisimétrica. Las matrices que verifican la relación (13) se llaman matrices *hermitianas antisimétricas*. Por consiguiente, en una base ortonormal a las aplicaciones antisimétricas les corresponden las matrices hermitianas antisimétricas y, viceversa, a las matrices hermitianas antisimétricas les corresponden aplicaciones antisimétricas.

De la relación (12) se deduce directamente que la suma de aplicaciones antisimétricas y el producto de una aplicación antisimétrica por un número real son de nuevo aplicaciones antisimétricas.

Todo valor propio de una aplicación antisimétrica o bien es igual a cero o bien es un número imaginario puro.

Si α es un valor propio de una aplicación antisimétrica $\mathcal{A}$ y a es un vector propio no nulo correspondiente, se tiene

$$(a, a\mathcal{A}) = (a, \alpha a) = \bar{\alpha}(a, a),$$
$$(a\mathcal{A}, a) = (\alpha a, a) = \alpha(a, a).$$

Pero $(a, a\mathcal{A}) = -(a\mathcal{A}, a)$ y, por consiguiente, $-\alpha = \bar{\alpha}$ que es lo que se quería demostrar.

En particular, de aquí resulta que *toda raíz del polinomio característico de una matriz hermitiana antisimétrica o bien es igual a cero o bien es un número imaginario puro.*

Puesto que de la relación $\mathcal{A}^* = -\mathcal{A}$ se deduce inmediatamente la igualdad $\mathcal{A}^*\mathcal{A} = \mathcal{A}\mathcal{A}^*$, *las aplicaciones antisimétricas resultan ser un caso particular de las aplicaciones normales* y, por ello, para hallar la forma más sencilla de las matrices de las aplicaciones antisimétricas es suficiente recurrir al teorema 8 del p. 18.3. Así se obtiene el teorema siguiente:

TEOREMA 6. *En una base ortonormal adecuada de un espacio unitario real la matriz A de una aplicación antisimétrica toma la forma*

$$A = O_k \dot{+} \begin{bmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{bmatrix} \dot{+} \dots \dot{+} \begin{bmatrix} 0 & \sigma_m \\ -\sigma_m & 0 \end{bmatrix}, \tag{14}$$

donde O_k es la matriz nula de orden k.

Efectivamente, en una base ortonormal adecuada la matriz de la aplicación dada se descompone, según el teorema mencionado, en células de órdenes 1 y 2. De la relación (13) se ve que también las células aisladas deben satisfacer esta misma igualdad. Las células de orden 1 son números reales ρ_i y la relación (13) da para ellas $\rho_j = -\overline{\rho}_j = -\rho_j$, es decir, $\rho_j = 0$. En cambio, si la célula es de orden 2, de (13) resulta que debe ser de la forma

$$A_j = \begin{bmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{bmatrix}$$

que es lo que se quería demostrar.

En términos matriciales el teorema 6 se puede enunciar de modo siguiente:

TEOREMA 6a. *Para toda matriz antisimétrica real A existe una matriz unitaria real U tal que la matriz UAU^{-1} es de la forma* (14).

Para las matrices reales los conceptos de matriz unitaria y de matriz ortogonal son equivalentes y por esto en los teoremas 5a y 6a las palabras *unitaria real* pueden ser sustituidas por las palabras *unitaria ortogonal.*

19.6. Aplicaciones simétricas no negativas. Una aplicación simétrica $\mathcal{A}$ de un espacio unitario $\mathfrak{L}$ se llama *no negativa*, si para todo x de $\mathfrak{L}$ se tiene

$$(x\mathcal{A}, x) \geqslant 0. \tag{15}$$

Aquí el signo de desigualdad tiene sentido, ya que en el caso de aplicaciones simétricas el producto escalar $(x\mathcal{A}, x)$ es siempre real. Si el signo de igualdad tiene lugar en (15) sólo para el vector nulo, se dice que $\mathcal{A}$ es una aplicación *positiva* o *definida positiva.*

Una combinación lineal de aplicaciones no negativas con coeficientes reales no negativos es una aplicación no negativa.

Esto se ve directamente de la fórmula

$$(x(\alpha\mathcal{A}+\beta\mathcal{B}),\ x)=\alpha(x\mathcal{A},\ x)+\beta(x\mathcal{B},\ x).$$

El producto de cualquier aplicación lineal por su conjugada es una aplicación simétrica no negativa.

Efectivamente,

$$(x\mathcal{A}\mathcal{A}^*,\ x)=(x\mathcal{A},\ x\mathcal{A})\geqslant 0,$$
$$(x\mathcal{A}^*\mathcal{A},\ x)=(x\mathcal{A}^*,\ x\mathcal{A}^*)\geqslant 0.$$

El cuadrado de cualquier aplicación simétrica es una aplicación no negativa.

Resulta de la anterior, ya que toda aplicación simétrica es conjugada de sí misma.

Todos los valores propios de una aplicación no negativa son reales y no negativos.

Sea $\mathcal{A}$ una aplicación no negativa, sea α un valor propio de la misma y sea a un vector propio no nulo correspondiente. Entonces

$$(a\mathcal{A},\ a)=\alpha(a,\ a)\geqslant 0.$$

Por consiguiente, se tiene $\alpha\geqslant 0$.

Si los valores propios de una aplicación simétrica de un espacio unitario $\mathfrak{L}$, complejo o real, son no negativos, la aplicación es no negativa.

En $\mathfrak{L}$ existe, en virtud del p. 19.4, una base ortonormal $e_1, \ldots, e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Sean $\alpha_1, \ldots, \alpha_n$ los valores propios correspondientes y sea

$$x=\xi_1e_1+\xi_2e_2+\ldots+\xi_ne_n$$

un vector arbitrario de $\mathfrak{L}$. Entonces

$$\begin{aligned}(x\mathcal{A},\ x)&=\alpha_1\xi_1\bar{\xi}_1+\alpha_2\xi_2\bar{\xi}_2+\ldots+\alpha_n\xi_n\bar{\xi}_n=\\&=\alpha_1|\xi_1|^2+\ldots+\alpha_n|\xi_n|^2\geqslant 0\end{aligned}\qquad(16)$$

que es lo que se quería demostrar.

El determinante de la aplicación $\mathcal{A}$ es igual a $\alpha_1\alpha_2\ldots\alpha_n$. Si él es diferente de cero, todos los números α_j son mayores que el cero y la suma (16) será igual a cero en este caso sólo para $x=o$. Por consiguiente, la aplicación $\mathcal{A}$ será en este caso definida positiva. En cambio, si $|\mathcal{A}|=0$, uno de los valores propios, digamos α_1, es igual a cero. Entonces

$$(e_1\mathcal{A},\ e_1)=0(e_1,\ e_1)=0$$

y la aplicación $\mathcal{A}$ no será definida positiva.

Por consiguiente, *una aplicación simétrica no negativa es definida positiva cuando, y sólo cuando, es regular.*

Consideremos ahora la operación de extracción de la raiz cuadrada de una aplicación lineal. Se dice que una aplicación lineal $\mathcal{X}$

es raíz cuadrada de una aplicación lineal $\mathcal{A}$, si

$$\mathcal{X}^2 = \mathcal{A}. \tag{17}$$

Según sea la aplicación $\mathcal{A}$, la ecuación (17) puede no tener soluciones, puede tener sólo un número finito de soluciones y puede tener un número infinito de soluciones. Sin embargo, en el caso de aplicaciones simétricas no negativas la situación es bien determinada.

TEOREMA 7. *Para toda aplicación simétrica no negativa $\mathcal{A}$ de un espacio unitario existe una aplicación simétrica no negativa $\mathcal{B}$, y sólo una, que cumple la relación*

$$\mathcal{B}^2 = \mathcal{A}.$$

Toda aplicación lineal que conmuta con $\mathcal{A}$ es conmutable con $\mathcal{B}$.

DEMOSTRACIÓN. Tomemos en $\mathfrak{L}$ una base ortonormal $e_1, \ldots, e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Tal base existe de acuerdo con el p. 19.4. Indiquemos por $\alpha_1, \ldots, \alpha_n$ los valores propios correspondientes de la aplicación $\mathcal{A}$. Sea $\mathcal{B}$ la aplicación lineal que transforma e_j en $\sqrt{\alpha_j} e_j$ $(j = 1, \ldots, n)$, donde se toman los valores no negativos de los radicales. Puesto que $e_1, \ldots, e_n$ es una base ortonormal formada por los vectores propios de la aplicación $\mathcal{B}$ y puesto que los valores propios de ésta son iguales a $\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_n}$, es decir, son no negativos, resulta que $\mathcal{B}$ es una aplicación simétrica no negativa. Pero

$$e_j\mathcal{B}^2 = \alpha_j e_j = e_j\mathcal{A} \qquad (j = 1, \ldots, n).$$

Por lo tanto, $\mathcal{B}^2 = \mathcal{A}$. Hemos demostrado que la raíz cuadrada requerida de $\mathcal{A}$ existe.

Demostremos la última afirmación del teorema. Sea $\mathcal{X}$ una aplicación lineal que conmuta con $\mathcal{A}$. Tomemos los vectores coordenados $e_1, \ldots, e_n$ en tal orden que los valores propios iguales, si es que existen, correspondan a vectores coordenados adyacentes. Entonces las matrices de las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ serán, respectivamente, de la forma

$$A = \begin{bmatrix} \alpha_1 E_1 & & & & \\ & \alpha_2 E_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \ \ \alpha_s E_s \end{bmatrix} \quad \text{y}$$

$$B = \begin{bmatrix} \sqrt{\alpha_1} E_1 & & & & \\ & \sqrt{\alpha_2} E_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \ \ \sqrt{\alpha_s} E_s \end{bmatrix},$$

donde $\alpha_1, \ldots, \alpha_s$ son *distintos* valores propios de la aplicación $\mathcal{A}$ y $E_1, \ldots, E_s$ son matrices unidades. Representemos la matriz de la aplicación $\mathcal{X}$ en la forma celular correspondiente

$$X = \begin{vmatrix} X_{11} & X_{12} & \ldots & X_{1s} \\ X_{21} & X_{22} & \ldots & X_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ X_{s1} & X_{s2} & \ldots & X_{ss} \end{vmatrix}.$$

De la condición $AX = XA$ tenemos

$$\alpha_j X_{jk} = X_{jk}\alpha_k \qquad (j,\ k = 1,\ 2,\ \ldots,\ s),$$

es decir,

$$(\alpha_j - \alpha_k) X_{jk} = O.$$

Puesto que $\alpha_j \neq \alpha_k$ para $j \neq k$, tenemos $X_{jk} = O$ $(j \neq k)$. Por consiguiente,

$$X = \begin{vmatrix} X_{11} & O & \ldots & O \\ O & X_{22} & \ldots & O \\ \cdot & \cdot & \cdot & \cdot \\ O & O & \ldots & X_{ss} \end{vmatrix};$$

pero entonces

$$BX = \begin{bmatrix} \sqrt{\alpha_1}X_{11} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \sqrt{\alpha_s}X_{ss} \end{bmatrix} = XB$$

que es lo que se quería demostrar.

Resta demostrar la unicidad. Sea $\mathcal{C}$ otra aplicación simétrica no negativa tal que $\mathcal{C}^2 = \mathcal{A}$. Correspondientemente a la descomposición celular de la matriz A señalada anteriormente, el espacio $\mathfrak{L}$ se descompondrá en la suma directa de los subespacios invariantes $\mathfrak{L}_j (j = 1, \ldots, s)$. Puesto que $\mathcal{C}\mathcal{A} = \mathcal{A}\mathcal{C}$, los subespacios $\mathfrak{L}_j$ serán también invariantes, de acuerdo con lo demostrado, respecto a $\mathcal{C}$. Como $\mathcal{C}$ es simétrica, en cada uno de los subespacios $\mathfrak{L}_j$ existe una base ortonormal formada por los vectores propios de $\mathcal{C}$. Sean $\gamma_{j1}, \ldots, \gamma_{jp}$ los valores propios correspondientes. Indiquemos por $\mathcal{A}_j$, $\mathcal{B}_j$ y $\mathcal{C}_j$ las aplicaciones inducidas en el subespacio $\mathfrak{L}_j$ por las aplicaciones $\mathcal{A}$, $\mathcal{B}$ y $\mathcal{C}$, respectivamente, tenemos

$$\mathcal{A}_j = \alpha_j \mathcal{E}_j, \quad \mathcal{B}_j = \sqrt{\alpha_j}\mathcal{E}_j \quad \text{y} \quad \mathcal{C}_j^2 = \mathcal{A}_j,$$

de donde

$$\gamma_{j1}^2 = \ldots = \gamma_{jp}^2 = \alpha_j.$$

Puesto que todos los números $\gamma_{j1}, \ldots, \gamma_{jp}$ son no negativos, de aquí se deduce que

$$\gamma_{j1} = \ldots = \gamma_{jp} = \sqrt{\alpha_j}$$

y, por consiguiente, $\mathscr{C}_j = \sqrt{\alpha_j}\mathscr{E}_j = \mathscr{B}_j$, es decir, $\mathscr{C} = \mathscr{B}$.

Demostremos, como ejemplo de una aplicación directa del teorema 7, que *el producto de aplicaciones simétricas no negativas conmutables es una aplicación no negativa.*

En efecto, sean $\mathscr{A}_1$ y $\mathscr{A}_2$ unas aplicaciones simétricas no negativas que conmutan. Indiquemos por $\mathscr{B}_1$ y $\mathscr{B}_2$ sus raíces cuadradas que, según el teorema 7, se pueden escoger de manera que conmuten y que sean simétricas y no negativas. Tenemos entonces

$$(\mathscr{B}_1\mathscr{B}_2)^2 = \mathscr{B}_1\mathscr{B}_2\mathscr{B}_1\mathscr{B}_2 = \mathscr{B}_1^2\mathscr{B}_2^2 = \mathscr{A}_1\mathscr{A}_2,$$

es decir, $\mathscr{A}_1\mathscr{A}_2$ es igual al cuadrado de una aplicación simétrica $\mathscr{B}_1\mathscr{B}_2$. Luego, $\mathscr{A}_1\mathscr{A}_2$ es no negativa. Hemos demostrado la proposición.

De ella se desprende, en particular, que los polinomios con coeficientes reales no negativos en una aplicación simétrica no negativa son también aplicaciones simétricas no negativas.

Ejemplos y problemas

1. Sea e_1, e_2 y e_3 una base ortonormal de un espacio unitario $\mathfrak{L}$. Hállense las matrices de las aplicaciones unitarias que transforman los vectores e_1 y e_2 en los vectores $\frac{2}{3}e_1 + \frac{2}{3}e_2 - \frac{1}{3}e_3$ y $\frac{2}{3}e_1 - \frac{1}{3}e_2 + \frac{2}{3}e_3$.

2. Si $a_1, \ldots, a_m$ y $b_1, \ldots, b_m$ son dos sistemas ortonormales de vectores de un espacio unitario $\mathfrak{L}$ de n dimensiones ($m \leqslant n$), existe una aplicación unitaria de $\mathfrak{L}$ que transforma el primer sistema en el segundo.

3. Para que un sistema de vectores $a_1, \ldots, a_m$ de un espacio unitario pueda ser transformado por una aplicación unitaria en otro sistema $b_1, \ldots, b_m$ es necesario y suficiente que las matrices de Gram de estos sistemas coincidan (véase el problema 5, pág. 215).

4. Demuéstrese, empleando la forma normal de Jordan y el proceso de ortonormalización de Gram—Schmidt, que la matriz de toda aplicación lineal de un espacio unitario puede ser reducida a la forma triangular en un sistema de coordenadas ortonormal adecuado (teorema de Schur).

5. Demuéstrese que siendo $\mathscr{A}$ una aplicación arbitraria de un espacio unitario $\mathfrak{L}$ que conserva los valores de los productos escalares, la aplicación $\mathscr{A}$ es lineal y es, por consiguiente, una aplicación unitaria del espacio $\mathfrak{L}$.

6. En un espacio lineal existen, salvo un isomorfismo, sólo dos funciones lineales. En un espacio unitario las funciones lineales son, salvo un isomorfismo, de la forma $\alpha(x, e)$, donde e es un vector unitario fijo.

7. En una base ortonormal de un espacio euclídeo las matrices de las aplicaciones $\mathscr{A}$, $\mathscr{B}$ y $\mathscr{C}$ son iguales respectivamente a

$$\begin{bmatrix} 5 & 2 & 4 \\ 2 & 2 & 2 \\ 4 & 2 & 5 \end{bmatrix}, \quad \begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1 & 3 & 1 \\ -1 & 2 & 2 \\ 1 & 0 & 2 \end{bmatrix}.$$

Demuéstrese que $\mathcal{A}$ es definida positiva, que $\mathcal{B}$ es no negativa y que $\mathcal{C}$, aun no siendo simétrica, es tal que $(x\mathcal{C}, x) \geqslant 0$ cualquiera que sea x.

8. Hállese la raíz cuadrada simétrica no negativa de la aplicación $\mathcal{B}$ del problema anterior.

9. La matriz de una aplicación simétrica no negativa, calculada en una base ortonormal, se llama *hermitiana no negativa*. Pruébese que una matriz hermitiana es no negativa cuando, y sólo cuando, se alternan los signos de los coeficientes de su polinomio característico. Además, si uno de los coeficientes es igual a cero, también tienen que ser iguales a cero los coeficientes de los términos de grado menor.

10. Demuéstrese que de toda aplicación normal $\mathcal{A}$ se puede extraer la raíz de cualquier grado positivo n, es decir, demuéstrese que para toda aplicación normal $\mathcal{A}$ existe una aplicación normal $\mathcal{X}$ que satisface la relación $\mathcal{X}^n = \mathcal{A}$. ¿Cuál es el número máximo de tales aplicaciones $\mathcal{X}$?

11. En un espacio unitario $\mathfrak{L}$ se ha tomado una base ortonormal. Demuéstrese que en esta base la matriz de toda aplicación simétrica no negativa $\mathcal{A}$ de rango 1 puede ser representada en la forma

$$A = \overline{[x]'}\,[x],$$

donde $[x]$ es la fila coordenada de un vector x convenientemente escogido.

12. Toda aplicación simétrica no negativa es una suma de aplicaciones simétricas no negativas de rango 1.

13. Si unas matrices hermitianas de elementos α_{ij} y β_{ij} son no negativas, la matriz de elementos $\gamma_{ij} = \alpha_{ij}\beta_{ij}$ es también no negativa $(i, j = 1, 2, \ldots, n)$.

§ 20. Descomposición de aplicaciones generales

Las aplicaciones unitarias, simétricas y antisimétricas tienen una estructura geométrica muy clara. Por esto al estudiar las aplicaciones lineales generales de espacios euclídeos o unitarios resulta natural preguntarse si es posible expresar de algún modo simple estas aplicaciones en términos de las aplicaciones especiales mencionadas. Algunos de estos métodos, que son de importancia principal, se consideran precisamente en este parrágrafo.

20.1. Descomposición en partes simétrica y antisimétrica. Sea $\mathfrak{L}$ un espacio unitario complejo y sea $\mathcal{A}$ una aplicación lineal del mismo. Designemos

$$\mathcal{B} = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*) \quad \text{y} \quad \mathcal{C} = \frac{1}{2i}(\mathcal{A} - \mathcal{A}^*). \tag{1}$$

Tenemos

$$\mathcal{B}^* = \frac{1}{2}(\mathcal{A}^* + \mathcal{A}) = \mathcal{B} \quad \text{y} \quad \mathcal{C}^* = -\frac{1}{2i}(\mathcal{A}^* - \mathcal{A}) = \mathcal{C}.$$

Por consiguiente, $\mathcal{B}$ y $\mathcal{C}$ son simétricas. De (1) resulta

$$\mathcal{A} = \mathcal{B} + i\mathcal{C}. \tag{2}$$

Es decir, *toda aplicación lineal $\mathcal{A}$ de un espacio unitario complejo puede ser representada en la forma* (2), *donde $\mathcal{B}$ y $\mathcal{C}$ son aplicaciones simétricas*. Esta representación es unívoca, ya que de (2) se deduce que

$$\mathcal{A}^* = \mathcal{B}^* - i\mathcal{C}^* = \mathcal{B} - i\mathcal{C},$$

de donde obtenemos para $\mathcal{B}$ y $\mathcal{C}$ de nuevo las expresiones (1).

Si el campo principal es real, la descomposición (2) no sirve. En este caso se procede del modo siguiente. Sea $\mathcal{A}$ una aplicación arbitraria. Pongamos

$$\mathcal{B}=\frac{1}{2}(\mathcal{A}+\mathcal{A}^*) \quad \text{y} \quad \mathcal{C}=\frac{1}{2}(\mathcal{A}-\mathcal{A}^*). \tag{3}$$

De aquí resulta

$$\mathcal{A}=\mathcal{B}+\mathcal{C}. \tag{4}$$

Puesto que

$$\mathcal{B}^*=\frac{1}{2}(\mathcal{A}^*+\mathcal{A})=\mathcal{B} \quad \text{y} \quad \mathcal{C}^*=\frac{1}{2}(\mathcal{A}^*-\mathcal{A})=-\mathcal{C},$$

en la descomposición (4) $\mathcal{B}$ es una aplicación simétrica y $\mathcal{C}$ es una aplicación antisimétrica. La descomposición (4) es unívoca, ya que de ella se deduce que $\mathcal{A}^*=\mathcal{B}-\mathcal{C}$ y de aquí obtenemos para $\mathcal{B}$ y $\mathcal{C}$ de nuevo las expresiones (3). Por consiguiente, *toda aplicación lineal puede ser representada como la suma de una aplicación simétrica y una aplicación antisimétrica. Esta representación es unívoca.*

Es evidente que la descomposición (4) sirve cualquiera que sea el campo principal. La aplicación $\mathcal{B}$ se llama parte *simétrica* y la aplicación $\mathcal{C}$ se llama parte *antisimétrica* de la aplicación $\mathcal{A}$.

Desde el punto de vista del cálculo de matrices, la descomposición (2) significa que toda matriz cuadrada A se puede representar en la forma $B+iC$, donde B y C son matrices hermitianas, mientras que la descomposición (4) significa que toda matriz cuadrada A puede ser representada en la forma $B+C$, donde B es una matriz simétrica y C es una matriz antisimétrica.

20.2. Descomposición polar. Desde el punto de vista geométrico es de mucho mayor interés la representación de una aplicación lineal como el producto de unas aplicaciones simétrica y unitaria. La posibilidad de tal representación se basa en el lema siguiente.

LEMA. *Si las aplicaciones lineales $\mathcal{A}$ y $\mathcal{B}$ de un espacio unitario $\mathfrak{L}$ alteran igualmente las longitudes de los vectores*[1], *es decir, si para cualquier a*

$$(a\mathcal{A},\ a\mathcal{A})=(a\mathcal{B},\ a\mathcal{B}), \tag{5}$$

existe una aplicación unitaria $\mathcal{U}$ del espacio $\mathfrak{L}$ tal que $\mathcal{A}\mathcal{U}=\mathcal{B}$.

Consideremos el dominio de valores de la aplicación $\mathcal{A}$, es decir, el conjunto de vectores de tipo $x\mathcal{A}$, donde x recorre todo el espacio $\mathfrak{L}$. Indiquemos por $\mathfrak{A}$ este dominio. Análogamente, indiquemos por $\mathfrak{B}$ el dominio de valores de la aplicación $\mathcal{B}$. $\mathfrak{A}$ y $\mathfrak{B}$ son, según el p. 10.1, unos subespacios lineales. Queremos, ante todo, establecer una correspondencia isomorfa entre $\mathfrak{A}$ y $\mathfrak{B}$. Sea a un vector de $\mathfrak{A}$.

[1] Las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ que poseen esta propiedad se llaman *métricamente iguales.*

Busquemos en $\mathfrak{L}$ un vector x tal que $x\mathcal{A}=a$ y tomemos $x\mathcal{B}=b$. Convengamos en decir que b es la imagen de a y en indicarla por $a\mathcal{V}^0$. Probemos que a determina unívocamente a b. En efecto, lo contrario puede acontecer sólo si el vector x no se determina unívocamente por la condición $x\mathcal{A}=a$. Sin embargo, si x_1 es otro vector tal que $x_1\mathcal{A}=a$, tenemos $(x-x_1)\mathcal{A}=o$. De aquí se deduce, en virtud de (5), que

$$((x-x_1)\mathcal{B},\ (x-x_1)\mathcal{B})=((x-x_1)\mathcal{A},\ (x-x_1)\mathcal{A})=0,$$

es decir, que $(x-x_1)\mathcal{B}=o$ ó $x_1\mathcal{B}=x\mathcal{B}$ que es lo que se quería demostrar. Hemos probado de esta forma que $\mathcal{V}^0$ es una aplicación unívoca de $\mathfrak{A}$ en $\mathfrak{B}$. Sin embargo, es fácil ver que $\mathcal{V}^0$ es una aplicación *biyectiva* de $\mathfrak{A}$ sobre $\mathfrak{B}$. Efectivamente, siendo b un vector de $\mathfrak{B}$, existe en $\mathfrak{L}$ un vector x tal que $x\mathcal{B}=b$. Tomando entonces $x\mathcal{A}=a$, tendremos $a\mathcal{V}^0=b$ que es lo que se quería demostrar.

De la misma definición de la correspondencia $\mathcal{V}^0$ se desprende que para todo vector x de $\mathfrak{L}$ es válida la igualdad

$$x\mathcal{A}\mathcal{V}^0=x\mathcal{B}. \tag{6}$$

Empleando esta igualdad se puede probar que $\mathcal{V}^0$ es una aplicación *isomorfa* de $\mathfrak{A}$ sobre $\mathfrak{B}$. Efectivamente, sean a_1 y a_2 unos vectores de $\mathfrak{A}$. Busquemos en $\mathfrak{L}$ unos vectores x_1 y x_2 tales que $x_1\mathcal{A}=a_1$ y $x_2\mathcal{A}=a_2$. Tenemos entonces

$$\begin{aligned}(\alpha a_1+\beta a_2)\mathcal{V}^0&=(\alpha x_1\mathcal{A}+\beta x_2\mathcal{A})\mathcal{V}^0=(\alpha x_1+\beta x_2)\mathcal{A}\mathcal{V}^0=(\alpha x_1+\beta x_2)\mathcal{B}=\\&=\alpha(x_1\mathcal{B})+\beta(x_2\mathcal{B})=\alpha(x_1\mathcal{A}\mathcal{V}^0)+\beta(x_2\mathcal{A}\mathcal{V}^0)=\alpha a_1\mathcal{V}^0+\beta a_2\mathcal{V}^0,\end{aligned} \tag{7}$$

es decir, la aplicación $\mathcal{V}^0$ conserva las operaciones de adición y de multiplicación por número. Además,

$$\begin{aligned}(a_1\mathcal{V}^0,\ a_1\mathcal{V}^0)&=(x_1\mathcal{A}\mathcal{V}^0,\ x_1\mathcal{A}\mathcal{V}^0)=(x_1\mathcal{B},\ x_1\mathcal{B})=\\&=(x_1\mathcal{A},\ x_1\mathcal{A})=(a_1,\ a_1),\end{aligned} \tag{8}$$

es decir, la aplicación $\mathcal{V}^0$ conserva las longitudes de los vectores. Por consiguiente, $\mathcal{V}^0$ es un isomorfismo.

Tenemos definida la aplicación $\mathcal{V}^0$ sólo para los vectores de $\mathfrak{A}$. Queremos ahora definirla también en todos los demás vectores del espacio $\mathfrak{L}$. Con este fin consideremos los subespacios ortogonales $\mathfrak{A}^\perp$ y $\mathfrak{B}^\perp$. Para $\mathfrak{L}$ son válidas, de acuerdo con el p. 17.5, las descomposiciones directas

$$\mathfrak{L}=\mathfrak{A}\dotplus\mathfrak{A}^\perp=\mathfrak{B}\dotplus\mathfrak{B}^\perp.$$

Los subespacios $\mathfrak{A}$ y $\mathfrak{B}$ son isomorfos y, por lo tanto, tienen la misma dimensión. Pero entonces los complementos ortogonales $\mathfrak{A}^\perp$ y $\mathfrak{B}^\perp$ también tienen la misma dimensión. Como los espacios unitarios de una misma dimensión son isomorfos, debe existir una aplicación biyectiva de $\mathfrak{A}^\perp$ sobre $\mathfrak{B}^\perp$ que conserva las operaciones de adición y de multiplicación por número y que no altera las longitudes de los vectores. Indiquemos esta aplicación por $\mathcal{W}$. Luego, si a' y a'' son unos vectores de $\mathfrak{A}^{\perp}$, los vectores $a'\mathcal{W}$ y $a''\mathcal{W}$ pertenecen a $\mathfrak{B}^{\perp}$ y

$$(\alpha a' + \beta a'')\mathcal{W} = \alpha(a'\mathcal{W}) + \beta(a''\mathcal{W}), \tag{9}$$

$$(a'\mathcal{W},\ a''\mathcal{W}) = (a',\ a''). \tag{10}$$

Definamos ahora una aplicación $\mathcal{U}$ del espacio $\mathfrak{L}$ del modo siguiente. Sea x un vector cualquiera de $\mathfrak{L}$. Puesto que $\mathfrak{L} = \mathfrak{A} \dot{+} \mathfrak{A}^{\perp}$, el vector x se puede representar unívocamente en la forma

$$x = x' + x'' \quad (x' \in \mathfrak{A} \quad \text{y} \quad x'' \in \mathfrak{A}^{\perp}). \tag{11}$$

Tomemos, por definición,

$$x\mathcal{U} = x'\mathcal{V} + x''\mathcal{W}. \tag{12}$$

La aplicación $\mathcal{U}$ es lineal, ya que si

$$y = y' + y'' \quad (y' \in \mathfrak{A} \quad \text{e} \quad y'' \in \mathfrak{A}^{\perp}),$$

de (7), (9) y (11) resulta:

$$(\alpha x + \beta y)\mathcal{U} = (\alpha x' + \beta y')\mathcal{V} + (\alpha x'' + \beta y'')\mathcal{W} = \alpha(x\mathcal{U}) + \beta(y\mathcal{U}).$$

La aplicación $\mathcal{U}$ es unitaria, ya que debido a (8), (10) y (11) se tiene

$$(x\mathcal{U},\ x\mathcal{U}) = (x'\mathcal{V} + x''\mathcal{W},\ x'\mathcal{V} + x''\mathcal{W}) =$$
$$= (x'\mathcal{V},\ x'\mathcal{V}) + (x''\mathcal{W},\ x''\mathcal{W}) = (x',\ x') + (x'' + x'') = (x,\ x).$$

Para todo x de $\mathfrak{L}$ es válida la relación

$$x\mathcal{A}\mathcal{U} = x\mathcal{B}.$$

Efectivamente, $x\mathcal{A}$ pertenece a $\mathfrak{A}$; luego, el vector x'' de la descomposición (11) es igual a cero y, por consiguiente,

$$x\mathcal{A}\mathcal{U} = x\mathcal{A}\mathcal{V}.$$

Teniendo en cuenta (6), esto da $x\mathcal{A}\mathcal{U} = x\mathcal{B}$. De aquí resulta $\mathcal{A}\mathcal{U} = \mathcal{B}$ y el lema queda demostrado.

TEOREMA 1 *Toda aplicación lineal $\mathcal{A}$ de un espacio unitario $\mathfrak{L}$ admite una descomposición polar*

$$\mathcal{A} = \mathcal{D}\mathcal{U}, \tag{13}$$

donde $\mathcal{D}$ es una aplicación simétrica no negativa y $\mathcal{U}$ es una aplicación unitaria del espacio $\mathfrak{L}$. La aplicación $\mathcal{D}$ se determina unívocamente; si $\mathcal{A}$ es una aplicación regular, la aplicación $\mathcal{U}$ tambien se determina unívocamente.

La aplicación $\mathcal{A}\mathcal{A}^*$ es, según el p. 19.6, simétrcia y no negativa. Indiquemos por $\mathcal{D}$ la raíz cuadrada simétrica no negativa de $\mathcal{A}\mathcal{A}^*$. Es decir,

$$\mathcal{D}^2 = \mathcal{A}\mathcal{A}^*.$$

Tenemos cualquiera que sea el vector x

$$(x\mathcal{A},\ x\mathcal{A}) = (x,\ x\mathcal{A}\mathcal{A}^*) = (x,\ x\mathcal{D}^2) = (x\mathcal{D},\ x\mathcal{D}),$$

es decir, las aplicaciónes $\mathcal{A}$ y $\mathcal{D}$ alteran igualmente las longitudes de los vectores. Basándonos en el lema, deducimos de aquí que existe una aplicación unitaria $\mathcal{U}$ tal que

$$\mathcal{D}\mathcal{U} = \mathcal{A}.$$

Con esto queda demostrada la existencia de la descomposición polar. Resta examinar su unicidad. De (13) resulta $\mathcal{A}^* = \mathcal{U}^*\mathcal{D} = \mathcal{U}^{-1}\mathcal{D}$ y $\mathcal{A}\mathcal{A}^* = \mathcal{D}\mathcal{U}\mathcal{U}^{-1}\mathcal{D} = \mathcal{D}^2$. Por consiguiente, la aplicación $\mathcal{D}$ es no negativa y simétrica y su cuadrado es igual a la aplicación $\mathcal{A}\mathcal{A}^*$. Según el teorema 7 del p. 19.6, estas condiciones determinan unívocamente la aplicación $\mathcal{D}$. Si $\mathcal{A}$ es una aplicación regular, también $\mathcal{D}$ es regular y de (13) resulta entonces que $\mathcal{U} = \mathcal{D}^{-1}\mathcal{A}$, es decir, la aplicación $\mathcal{U}$ también se determina unívocamente.

El significado geométrico del teorema 1 es muy sencillo. Indica precisamente que la acción de toda aplicación lineal del espacio $\mathfrak{L}$ puede ser representada de la forma siguiente: primero el espacio $\mathfrak{L}$ se dilata en n direcciones recíprocamente ortogonales con un coeficiente de dilatación concreto, real y no negativo, en cada una de las direcciones y después gira alrededor del origen de coordenadas[1]. Si la aplicación es regular, todos los coeficientes de dilatación son estrictamente positivos. En el caso de una aplicación singular algunos de los coeficientes resultan iguales a cero y en lugar de la dilatación en las direcciones correspondientes tiene lugar la proyección del espacio.

Observemos también que en la demostración de la existencia de la descomposición polar nos hemos basado en el producto $\mathcal{A}\mathcal{A}^*$. Si en lugar de él tomamos el producto $\mathcal{A}^*\mathcal{A}$, obtendremos la descomposición de tipo

$$\mathcal{A} = \mathcal{U}\mathcal{D}_1,$$

donde $\mathcal{U}$ es unitaria y $\mathcal{D}_1$ es una aplicación simétrica no negativa.

Tomemos en el espacio $\mathfrak{L}$ un sistema ortonormal de coordenadas. A las aplicaciones unitarias les corresponden entonces matrices unitarias y a las aplicaciones simétricas les corresponden matrices hermitianas y el teorema 1 se convierte en la proposición siguiente: *toda matriz cuadrada puede ser representada en la forma de un producto de una matriz hermitiana y otra unitaria.*

---

[1] El giro se entiende en el sentido de una aplicación unitaria.

Suponiendo que el campo principal es el cuerpo de los números reales, obtenemos que *toda matriz cuadrada real puede ser representada como el producto de una matriz simétrica real y otra ortogonal real.*

En el teorema 1 se afirma que $\mathcal{D}$ es una aplicación simétrica *no negativa.* De acuerdo con esto en las dos últimas proposiciones a las palabras *hermitiana* y *simétrica* se puede agregar: *de valores propios no negativos.*

20.3. Aplicación de Cayley. Comparando las propiedades de las aplicaciones unitarias con las propiedades de las aplicaciones simétricas, podemos notar que ambas clases de aplicaciones están ligadas estrechamente. En forma explícita esta relación queda expresada en las así llamadas fórmulas de Cayley.

TEOREMA 2 (aplicación de Cayley). *Si $\mathcal{A}$ es una aplicación simétrica de un espacio unitario complejo, las aplicaciones $\mathcal{A} \pm i\mathcal{E}$ son invertibles; la aplicación $\mathcal{U}$ definida mediante la fórmula*

$$\mathcal{U} = (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1} \tag{14}$$

es unitaria, no tiene valores propios iguales a la unidad y, además, $\mathcal{A}$ se expresa mediante $\mathcal{U}$ por la fórmula

$$\mathcal{A} = -i(\mathcal{U} + \mathcal{E})(\mathcal{U} - \mathcal{E})^{-1}. \tag{15}$$

Recíprocamente, si $\mathcal{U}$ es una aplicación unitaria que no tiene valores propios iguales a la unidad, la aplicación $\mathcal{U} - \mathcal{E}$ es invertible, la aplicación $\mathcal{A}$ calculada mediante la fórmula (15) *es simétrica y $\mathcal{U}$ se expresa mediante $\mathcal{A}$ en la forma* (14).

DEMOSTRACIÓN. Sea $\mathcal{A}$ una aplicación simétrica de un espacio unitario. Los números $\pm i$ no pueden ser valores propios de la aplicación $\mathcal{A}$, ya que todos los valores propios de las aplicaciones simétricas son reales (p. 19.4). Esto significa que las aplicaciones $\mathcal{A} \pm i\mathcal{E}$ son regulares. Puesto que las aplicaciones $\mathcal{A} + i\mathcal{E}$ y $\mathcal{A} - i\mathcal{E}$ conmutan, resulta que ellas conmutan también con las aplicaciones $(\mathcal{A} + i\mathcal{E})^{-1}$ y $(\mathcal{A} - i\mathcal{E})^{-1}$. Para la aplicación $\mathcal{U}$ definida mediante la fórmula (14) existe la conjugada que es igual a

$$\mathcal{U}^* = (\mathcal{A} + i\mathcal{E})^{*-1}(\mathcal{A} - i\mathcal{E})^* = (\mathcal{A} - i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E}).$$

Tenemos de aquí

$$\begin{aligned}\mathcal{U}\mathcal{U}^* &= (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1}(\mathcal{A} - i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E}) = \\ &= (\mathcal{A} - i\mathcal{E})(\mathcal{A} + i\mathcal{E})^{-1}(\mathcal{A} + i\mathcal{E})(\mathcal{A} - i\mathcal{E})^{-1} = \mathcal{E},\end{aligned}$$

es decir, $\mathcal{U}$ es unitaria. Probemos que $\mathcal{U} - \mathcal{E}$ es invertible. Para ello restemos de ambos miembros de la igualdad (14) la aplicación $\mathcal{E}$ y multipliquemos los resultados por $\mathcal{A} + i\mathcal{E}$. Después de efectuar transformaciones evidentes, tendremos

$$(\mathcal{U} - \mathcal{E})(\mathcal{A} + i\mathcal{E}) = -2i\mathcal{E}, \tag{16}$$

es decir,

$$(\mathcal{U}-\mathcal{E})^{-1}=-\frac{1}{2i}(\mathcal{A}+i\mathcal{E}).$$

Por consiguiente, $\mathcal{U}$ no tiene valores propios iguales a la unidad. Además, de (16) resulta:

$$(\mathcal{U}-\mathcal{E})\mathcal{A}=-2i\mathcal{E}-i(\mathcal{U}-\mathcal{E})=-i(\mathcal{U}+\mathcal{E}),$$

es decir,

$$\mathcal{A}=-i(\mathcal{U}+\mathcal{E})(\mathcal{U}-\mathcal{E})^{-1}.$$

Hemos demostrado la primera parte del teorema. La demostración de la recíproca es análoga totalmente a la demostración realizada.

Las fórmulas de Cayley establecen una correspondencia biyectiva entre todas las aplicaciones simétricas de un espacio unitario $\mathfrak{L}$ y aquellas aplicaciones unitarias $\mathcal{U}$ del mismo para los cuales 1 no es valor propio. Análogamente, las fórmulas

$$\mathcal{U}=(i\mathcal{E}+\mathcal{A})(i\mathcal{E}-\mathcal{A})^{-1}, \tag{14'}$$

$$\mathcal{A}=i(\mathcal{U}-\mathcal{E})(\mathcal{U}+\mathcal{E})^{-1} \tag{15'}$$

ofrecen una correspondencia biyectiva entre las aplicaciones simétricas del espacio $\mathfrak{L}$ y aquellas aplicaciones unitarias $\mathcal{U}$ para las cuales -1 no es valor propio.

Las aplicaciones (14) y (15), así como las aplicaciones (14′) y (15′), son posibles gracias a que en el campo principal existe el número i. Si el campo principal es real, las fórmulas indicadas no son válidas. Sin embargo, es fácil modificar estas fórmulas de manera que sean válidas para cualquier campo. Tiene lugar el teorema siguiente:

TEOREMA 3. *Sea $\mathcal{A}$ una aplicación antisimétrica de un espacio unitario $\mathfrak{L}$. Entonces las aplicaciones $\mathcal{A}\pm\mathcal{E}$ son invertibles, la aplicación*

$$\mathcal{U}=(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1} \tag{17}$$

es unitaria, no tiene valores propios iguales a la unidad y, ademas,

$$\mathcal{A}=-(\mathcal{U}+\mathcal{E})(\mathcal{U}-\mathcal{E})^{-1}. \tag{18}$$

Recíprocamente, si $\mathcal{U}$ es una aplicación unitaria y la unidad no es valor propio de la misma, la aplicación $\mathcal{A}$ definida mediante la fórmula (18) *es antisimétrica y $\mathcal{U}$ se expresa mediante $\mathcal{A}$ en la forma* (17).

En efecto, si $\mathcal{A}$ es antisimétrica, los números ± 1 no pueden ser sus valores propios, ya que todos los valores propios de las aplicaciones antisimétricas o bien son iguales a cero o bien son imaginarios puros (p. 19.5). Por esto las aplicaciones $\mathcal{A}\pm\mathcal{E}$ son regulares. Puesto que $(\mathcal{A}+\mathcal{E})(\mathcal{A}-\mathcal{E})=(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})$ tenemos $(\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1}=(\mathcal{A}+\mathcal{E})^{-1}(\mathcal{A}-\mathcal{E})$. De (17) resulta

$$\mathcal{U}^{*}=(\mathcal{A}^{*}+\mathcal{E})^{-1}(\mathcal{A}^{*}-\mathcal{E})=(-\mathcal{A}+\mathcal{E})^{-1}(-\mathcal{A}-\mathcal{E})=$$
$$=(\mathcal{A}-\mathcal{E})^{-1}(\mathcal{A}+\mathcal{E}),$$

de donde

$$\mathcal{U}\mathcal{U}^* = (\mathcal{A}-\mathcal{E})(\mathcal{A}+\mathcal{E})^{-1}(\mathcal{A}-\mathcal{E})^{-1}(\mathcal{A}+\mathcal{E}) = \mathcal{E},$$

es decir, la aplicación $\mathcal{U}$ es unitaria. Los razonamientos ulteriores son totalmente análogos a los realizados en la demostración del teorema 2 y, por ello, los omitimos.

Para concluir, observemos que los resultados de los últimos parágrafos descubren cierta semejanza entre las propiedades de las aplicaciones lineales de los espacios unitarios y las propiedades de los números complejos. Convengamos en aceptar que las aplicaciones lineales son, en cierto sentido, análogas a los números complejos y que las aplicaciones conjugadas son análogas a los números conjugados. Entonces las aplicaciones simétricas, que se caracterizan por la condición $\mathcal{A}^* = \mathcal{A}$, serán análogas a los números complejos que satisfacen la relación $\bar{z} = z$, es decir, a los números reales; las aplicaciones antisimétricas, que se caracterizan por la condición $\mathcal{A}^* = \mathcal{A}$, serán análogas a los números complejos que satisfacen a la relación $\bar{z} = -z$, es decir, a los números imaginarios puros; las aplicaciones unitarias con la propiedad $\mathcal{U}\mathcal{U}^* = \mathcal{E}$ serán análogas a los números complejos z para los cuales $z\bar{z} = 1$, es decir, $|z| = 1$. La descomposición $\mathcal{A} = \mathcal{B} + i\mathcal{C}$ del punto 20.1 corresponderá a la representación de un número complejo z en la forma cartesiana $z = x + iy$ y la descomposición polar $\mathcal{A} = \mathcal{D}\mathcal{U}$ corresponderá a la representación de un número complejo en la forma trigonométrica $z = \rho(\cos\varphi + i\,\mathrm{sen}\,\varphi)$, etc.

20.4. Descomposición espectral. Desde el punto de vista geométrico uno de los tipos más sencillos de las aplicaciones lineales es la proyección de los vectores sobre un subespacio. Algunas de las propiedades de estas aplicaciones proyectivas serán ahora consideradas.

Sea $\mathfrak{A}$ un subespacio lineal de un espacio unitario $\mathfrak{R}$. El conjunto de vectores ortogonales a $\mathfrak{A}$ es el subespacio ortogonal $\mathfrak{A}^{\perp}$ y $\mathfrak{R}$ es la suma directa de $\mathfrak{A}$ y $\mathfrak{A}^{\perp}$. Luego, todo vector a de $\mathfrak{R}$ puede ser representado unívocamente en la forma

$$a = a' + a'' \quad (a' \in \mathfrak{A} \quad \text{y} \quad a'' \in \mathfrak{A}^{\perp}). \tag{19}$$

El vector a' se llama *proyección* del vector a sobre el subespacio $\mathfrak{A}$. Poniendo en correspondencia a todo vector su proyección sobre $\mathfrak{A}$, obtenemos una aplicación del espacio $\mathfrak{R}$ que se llama *proyectiva* y se indica por $\mathcal{P}_{\mathfrak{A}}$. Es decir, se toma por definición

$$a\mathcal{P}_{\mathfrak{A}} = a'.$$

A veces, para abreviar la notación, omitiremos el índice $\mathfrak{A}$ y en lugar de $\mathcal{P}_{\mathfrak{A}}$ escribiremos $\mathcal{P}$.

Las aplicaciones proyectivas son lineales, ya que si para un vector a tiene lugar la descomposición (19) y para otro vector b la descomposición

$$b = b' + b'' \quad (b' \in \mathfrak{A} \quad \text{y} \quad b'' \in \mathfrak{A}^{\perp}),$$

resulta

$$\alpha a + \beta b = (\alpha a' + \beta b') + (\alpha a'' + \beta b''),$$

es decir, $(\alpha a+\beta b)\mathcal{P}=\alpha a'+\beta b'$, de donde

$$(\alpha a+\beta b)\,\mathcal{P}=\alpha\cdot a\mathcal{P}+\beta\cdot b\mathcal{P}.$$

Las aplicaciones proyectivas son aplicaciones simétricas no negativas. Efectivamente, de la descomposición (19) tenemos

$$(a\mathcal{P},\ a)=(a',\ a'+a'')=(a',\ a')\geqslant 0,$$

es decir, $\mathcal{P}$ es no negativa. Por otra parte, si

$$b=b'+b''\quad (b'\in\mathfrak{A}\quad \text{y}\quad b''\in\mathfrak{A}^{\perp}),$$

tenemos

$$(a\mathcal{P},\ b)=(a',\ b'+b'')=(a',\ b')=(a'+a'',\ b')=(a,\ b\mathcal{P}),$$

es decir, la aplicación $\mathcal{P}$ es simétrica.

De la fórmula (19) se desprende también otra propiedad importante de las aplicaciones proyectivas. Es evidente que para cualquier a tenemos

$$a\mathcal{P}^2=a'\mathcal{P}=a'=a\mathcal{P},$$

de donde

$$\mathcal{P}^2=\mathcal{P}. \tag{20}$$

Las aplicaciones que coinciden con su cuadrado se llaman *idempotentes*. Por consiguiente, de (20) resulta que *todas las aplicaciones proyectivas son idempotentes.*

Recíprocamente, las propiedades de simetría y de idempotencia caracterizan plenamente las aplicaciones proyectivas: *toda aplicación simétrica idempotente $\mathcal{P}$ es la proyección sobre el dominio de valores de $\mathcal{P}$.*

Sea $\mathfrak{A}$ el dominio de valores de $\mathcal{P}$. Para todo a es válida la descomposición

$$a=a\mathcal{P}+a\,(\mathcal{E}-\mathcal{P}). \tag{21}$$

El sumando $a\mathcal{P}$ pertenece, por definición, a $\mathfrak{A}$. El segundo sumando es ortogonal a $\mathfrak{A}$, ya que todo vector de $\mathfrak{A}$ es de la forma $x\mathcal{P}$, donde x es un vector de $\mathfrak{L}$, y debido a que la aplicación $\mathcal{P}$ es simétrica e idempotente tenemos

$$(x\mathcal{P},\ a\,(\mathcal{E}-\mathcal{P}))=(x,\ a\,(\mathcal{E}-\mathcal{P})\,\mathcal{P})=(x,\ a\mathcal{P}-a\mathcal{P})=0.$$

Por lo tanto, la descomposición (21) muestra que $a\mathcal{P}$ es la proyección del vector a sobre $\mathfrak{A}$ que es lo que se quería demostrar.

Hallemos la forma más sencilla de la matriz de una aplicación proyectiva $\mathcal{P}_{\mathfrak{A}}$. Tomemos en $\mathfrak{A}$ y en $\mathfrak{A}^{\perp}$ unas bases ortonormales $e_1,\ \ldots,\ e_m$ y $e_{m+1},\ \ldots,\ e_n$. El sistema unido $e_1,\ \ldots,\ e_m,\ e_{m+1},\ \ldots,\ e_n$ será una base ortonormal del espacio $\mathfrak{L}$. Las igualdades

$$e_j\mathcal{P}=e_j\quad \text{y}\quad e_\alpha\mathcal{P}=o\quad (j=1,\ \ldots,\ m;\ \alpha=m+1,\ \ldots,\ n)$$

muestran que la matriz de la aplicación $\mathcal{P}_{\mathfrak{A}}$ en esta base es de la forma

$$P=\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}. \tag{22}$$

Recíprocamente, si en un sistema ortonormal de coordenadas la matriz de una aplicación lineal $\mathcal{P}$ se reduce a la forma (22), es evidente que $\mathcal{P}$ es una aplicación proyectiva.

Sean $\mathcal{P}$ y Q las operaciones de proyección de un espacio $\mathfrak{L}$ sobre unos subespacios $\mathfrak{A}$ y $\mathfrak{B}$. Surge la pregunta: ¿cómo la posición de los subespacios $\mathfrak{A}$ y $\mathfrak{B}$ en $\mathfrak{L}$ influye en las propiedades de las aplicaciones $\mathcal{P}$ y Q? Por ejemplo, ¿qué puede decirse sobre $\mathcal{P}$ y Q, si $\mathfrak{A}$ y $\mathfrak{B}$ son ortogonales o si $\mathfrak{A}$ pertenece a $\mathfrak{B}$, etc.? Para responder a estas preguntas introduciremos, ante todo, la definición siguiente: *dos aplicaciones proyectivas $\mathcal{P}$ y Q se llaman ortogonales, si $\mathcal{P}Q=\mathcal{O}$*. Puesto que las aplicaciones proyectivas son simétricas, tenemos de aquí

$$Q\mathcal{P}=Q^*\mathcal{P}^*=(\mathcal{P}Q)^*=\mathcal{O},$$

es decir, si $\mathcal{P}$ es ortogonal a Q, Q es ortogonal a $\mathcal{P}$.

Para que unas aplicaciones proyectivas $\mathcal{P}$ y Q sean ortogonales es necesario y suficiente que sean ortogonales los subespacios correspondientes $\mathfrak{A}$ y $\mathfrak{B}$.

Sea $\mathcal{P}Q=\mathcal{O}$; entonces para $a\in\mathfrak{A}$ y $b\in\mathfrak{B}$ tenemos

$$(a,\ b)=(a\mathcal{P},\ bQ)=(a\mathcal{P}Q,\ b)=0,$$

es decir, $\mathfrak{A}$ es ortogonal a $\mathfrak{B}$. Recíprocamente, si $\mathfrak{A}$ es ortogonal a $\mathfrak{B}$, tenemos para cualquier vector x de $\mathfrak{L}$

$$x\mathcal{P}\in\mathfrak{A},\ x(\mathcal{P}Q)=(x\mathcal{P})Q=o \quad \text{y} \quad \mathcal{P}Q=\mathcal{O}$$

que es lo que se quería demostrar.

Para el estudio detallado de las propiedades de las aplicaciones lineales suelen emplearse las representaciones matriciales de las mismas. Pero si la representación matricial es, por cualquier razón, incómoda, se trata de expresar la aplicación lineal dada mediante aplicacionas de carácter más simple. En el caso de las aplicaciones normales, para estas aplicaciones elementales se pueden tomar las aplicaciones proyectivas.

Una descomposición de tipo

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots\alpha_s\mathcal{P}_s \tag{23}$$

se llama descomposición espectral de la aplicación $\mathcal{A}$, si

a) *los números $\alpha_1, \ldots, \alpha_s$ son diferentes;*

b) $\mathcal{P}_j^*=\mathcal{P}_j\neq\mathcal{O} \quad (j=1,\ 2,\ \ldots,\ s);$

c) $\mathcal{P}_j^2=\mathcal{P}_j \quad (j=1,\ 2,\ \ldots,\ s);$

d) $\mathcal{P}_j\mathcal{P}_k=\mathcal{O} \quad (j\neq k;\ j,\ k=1,\ 2,\ \ldots,\ s);$

e) $\mathcal{P}_1+\mathcal{P}_2+\ldots+\mathcal{P}_s=\mathcal{E}.$

Las condiciones b), c) y d) significan que $\mathcal{P}_1, \ldots, \mathcal{P}_s$ son aplicaciones proyectivas recíprocamente ortogonales.

Está claro que *sólo las aplicaciones normales admiten la descomposición espectral*. Efectivamente, de (23) se deduce:

$$\mathcal{A}^*=\bar{\alpha}_1\mathcal{P}_1+\bar{\alpha}_2\mathcal{P}_2+\ldots+\bar{\alpha}_s\mathcal{P}_s,$$
$$\mathcal{A}\mathcal{A}^*=\left(\sum\alpha_j\mathcal{P}_j\right)\left(\sum\bar{\alpha}_k\mathcal{P}_k\right)=\sum\bar{\alpha}_j\alpha_j\mathcal{P}_j=\mathcal{A}^*\mathcal{A}.$$

Recíprocamente, *toda aplicación normal de un espacio unitario complejo admite una descomposición espectral.*

Sea $\mathcal{A}$ una aplicación normal de un espacio unitario complejo $\mathfrak{L}$. Hemos visto que en $\mathfrak{L}$ existe una base ortonormal $e_1, e_2, \ldots, e_n$ formada por los vectores propios de la aplicación $\mathcal{A}$. Ordenemos estos vectores de manera que aquellos que corresponden a valores propios iguales se encuentren al lado. Supongamos, por ejemplo, que $e_1, \ldots, e_{m_1}$ corresponden al valor propio α_1, que $e_{m_1+1}, \ldots, e_{m_2}$ corresponden al valor propio α_2, etc. Indiquemos por $\mathfrak{L}_j$ el subespacio tendido sobre los vectores $e_{m_{j-1}+1}, \ldots, e_{m_j}$ correspondientes al valor propio α_i ($i=1, 2, \ldots, s$). Puesto que todos los vectores coordenados de $\mathfrak{L}_i$ son vectores propios correspondientes a un mismo valor propio α_i, todos los vectores de $\mathfrak{L}_i$ serán vectores propios con valores propios iguales a α_i. Tenemos

$$\mathfrak{L}=\mathfrak{L}_1+\mathfrak{L}_2+\ldots+\mathfrak{L}_s, \tag{24}$$

donde los subespacios $\mathfrak{L}_1, \ldots, \mathfrak{L}_s$ son recíprocamente ortogonales. Indiquemos por $\mathcal{P}_i$ la aplicación de proyección sobre el subespacio $\mathfrak{L}_i$. Pero la ortogonalidad de los subespacios $\mathfrak{L}_i$ implica la ortogonalidad de las correspondientes aplicaciones proyectivas. Además, de la igualdad (24) se deduce que $\mathcal{P}_1+\mathcal{P}_2+ +\ldots+\mathcal{P}_s=\mathcal{P}$, donde $\mathcal{P}$ es la proyección sobre $\mathfrak{L}$. Como la proyección sobre $\mathfrak{L}$ es la aplicación idéntica, tenemos

$$\mathcal{E}=\mathcal{P}_1+\mathcal{P}_2+\ldots+\mathcal{P}_s. \tag{25}$$

Vemos, por consiguiente, que las aplicaciones $\mathcal{P}_1, \ldots, \mathcal{P}_s$ poseen las propiedades de b) a e). Probemos finalmente que

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s.$$

Sea a un vector cualquiera de $\mathfrak{L}$. De la igualdad (25) resulta

$$a=a\mathcal{P}_1+a\mathcal{P}_2+\ldots+a\mathcal{P}_s. \tag{26}$$

El vector $a\mathcal{P}_i$ pertenece a $\mathfrak{L}_i$ y todos los vectores de $\mathfrak{L}_i$ son vectores propios correspondientes al valor propio α_i; por ello $a\mathcal{P}_i\mathcal{A}=\alpha_i\cdot a\mathcal{P}_i$. Multiplicando (26) por $\mathcal{A}$, obtenemos

$$a\mathcal{A}=\alpha_1 a\mathcal{P}_1+\ldots+\alpha_s a\mathcal{P}_s=a(\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s),$$

de donde $\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s$ que es lo que se quería demostrar.

Si

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s \tag{27}$$

es una descomposición espectral de una aplicación $\mathcal{A}$, entonces $\alpha_1, \ldots, \alpha_s$ es el conjunto de todos los valores propios de esta aplicación.

En efecto, cualquier aplicación $\mathcal{P}_j$ es, por hipótesis, diferente de la aplicación $\mathcal{O}$. Por consiguiente, existe un vector a tal que $a\mathcal{P}_j \neq o$. Pero como las aplicaciones $\mathcal{P}_1, \ldots, \mathcal{P}_s$ son ortogonales, obtenemos entonces

$$a\mathcal{P}_j\mathcal{A}=\alpha_1 a\mathcal{P}_j\mathcal{P}_1+\alpha_2 a\mathcal{P}_j\mathcal{P}_2+\ldots+\alpha_s a\mathcal{P}_j\mathcal{P}_s=\alpha_j a\mathcal{P}_j,$$

es decir, $a\mathcal{P}_j$ es un vector propio correspondiente al valor propio α_j.

Recíprocamente, sea a un vector propio no nulo de la aplicación $\mathcal{A}$ correspondiente al valor propio β. De la propiedad e) se deduce que

$$a=a\mathcal{P}_1+a\mathcal{P}_2+\ldots+a\mathcal{P}_s. \tag{28}$$

La condición (27) da:

$$a\mathcal{A}=\alpha_1 a\mathcal{P}_1+\alpha_2 a\mathcal{P}_2+\ldots+\alpha_s a\mathcal{P}_s.$$

Puesto que $a\mathcal{A}=\beta a$, tenemos

$$\alpha_1 a\mathcal{P}_1+\ldots+\alpha_s a\mathcal{P}_s=\beta a\mathcal{P}_1+\ldots+\beta a\mathcal{P}_s.$$

Multiplicando esta relación por $\mathcal{P}_j$ y empleando las propiedades c) y d) encontramos

$$\alpha_j a\mathcal{P}_j=\beta a\mathcal{P}_j \quad \text{y} \quad (\alpha_j-\beta)\, a\mathcal{P}_j=o \; (j=1, \ldots, s).$$

El vector a es diferente de cero y por ello al menos un sumando $a\mathcal{P}_j$ de (28) es también diferente de cero. Pero de la igualdad $(\alpha_j-\beta)\,a\mathcal{P}_j=o$ resulta entonces $\beta=\alpha_j$ que es lo que se quería demostrar.

Si $\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s$ es una descomposición espectral de una aplicación $\mathcal{A}$ y si $f(\lambda)=\beta_0+\beta_1\lambda+\ldots+\beta_m\lambda^m$ es un polinomio cualquiera, se tiene

$$f(\mathcal{A})=f(\alpha_1)\mathcal{P}_1+\ldots+f(\alpha_s)\mathcal{P}_s. \tag{29}$$

Tenemos

$$\mathcal{E}=\mathcal{P}_1+\ldots+\mathcal{P}_s,$$
$$\mathcal{A}=\alpha_1\mathcal{P}_1+\ldots+\alpha_s\mathcal{P}_s,$$
$$\mathcal{A}^2=\sum\alpha_j\mathcal{P}_j\cdot\sum\alpha_j\mathcal{P}_j=\sum\alpha_j\alpha_k\mathcal{P}_j\mathcal{P}_k=\alpha_1^2\mathcal{P}_1+\ldots+\alpha_s^2\mathcal{P}_s,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\mathcal{A}^m=\sum\alpha_j^{m-1}\mathcal{P}_j\cdot\sum\alpha_j\mathcal{P}_j=\sum\alpha_j^{m-1}\alpha_k\mathcal{P}_j\mathcal{P}_k=\alpha_1^m\mathcal{P}_1+\ldots+\alpha_s^m\mathcal{P}_s.$$

Multiplicando estas igualdades por los números $\beta_0, \beta_1, \ldots, \beta_m$, respectivamente, y sumándolas obtenemos la relación requerida.

Si

$$\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s$$

es una descomposición espectral de una aplicación $\mathcal{A}$, se tiene

$$\mathcal{P}_i=\frac{(\mathcal{A}-\alpha_1\mathcal{E})\ldots(\mathcal{A}-\alpha_{i-1}\mathcal{E})(\mathcal{A}-\alpha_{i+1}\mathcal{E})\ldots(\mathcal{A}-\alpha_s\mathcal{E})}{(\alpha_i-\alpha_1)\ldots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\ldots(\alpha_i-\alpha_s)}.$$

En efecto, consideremos el polinomio

$$\varphi_i(\lambda)=\frac{(\lambda-\alpha_1)\ldots(\lambda-\alpha_{i-1})(\lambda-\alpha_{i+1})\ldots(\lambda-\alpha_s)}{(\alpha_i-\alpha_1)\ldots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\ldots(\alpha_i-\alpha_s)}.$$

De acuerdo con (29) tenemos

$$\varphi_i(\mathcal{A})=\varphi_i(\alpha_1)\mathcal{P}_1+\varphi_i(\alpha_2)\mathcal{P}_2+\ldots+\varphi_i(\alpha_s)\mathcal{P}_s.$$

Pero $\varphi_i(\alpha_j)=0$ $(i\neq j)$ y $\varphi_i(\alpha_i)=1$; por consiguiente, $\varphi_i(\mathcal{A})=\mathcal{P}_i$.

La última propiedad significa que *toda aplicación normal de un espacio unitario admite una descomposición espectral, y sólo una.*

Efectivamente, ya hemos demostrado la existencia de la descomposición y por ello sólo se trata de demostrar su unicidad. Los coeficientes $\alpha_1, \ldots, \alpha_s$ de la descomposición espectral son los valores propios de la aplicación $\mathcal{A}$ y, por consiguiente, se determinan unívocamente por la aplicación $\mathcal{A}$. Pero si conocemos estos coeficientes, conocemos al mismo tiempo los polinomios $\varphi_i(\lambda)$ y, por lo tanto, conocemos las aplicaciones proyectivas $\mathcal{P}_i$ que es lo que se quería demostrar.

Indiquemos algunas otras propiedades de las descomposiciones espectrales. El conjunto de los coeficientes $\alpha_1, \ldots, \alpha_s$ de la descomposición espectral se denomina *espectro* de la misma. Se llama *espectro de una aplicación lineal* el espectro de su descomposición espectral. Hemos demostrado anteriormente que el espectro de una aplicación normal coincide con el conjunto de sus valores propios.

Para que una aplicación normal sea simétrica, respectivamente antisimétrica o unitaria, es necesario y suficiente que su espectro sea real, respectivamente imaginario puro o formado por números de módulo unidad.

Sea $\mathcal{A}=\alpha_1\mathcal{P}_1+\alpha_2\mathcal{P}_2+\ldots+\alpha_s\mathcal{P}_s$ la descomposición espectral de la aplicación $\mathcal{A}$. Tenemos entonces que $\mathcal{A}^*=\bar{\alpha}_1\mathcal{P}_1+\bar{\alpha}_2\mathcal{P}_2+\ldots+\bar{\alpha}_s\mathcal{P}_s$ es la descomposición espectral de la aplicación conjugada. Puesto que toda aplicación normal admite sólo una descomposición espectral, resulta que la condición $\mathcal{A}=\mathcal{A}^*$ equivale a las igualdades $\alpha_i=\bar{\alpha}_i$ $(i=1, \ldots, s)$, es decir, equivale a que el espectro sea real. Hemos demostrado la primera afirmación. Las otras dos se demuestran análogamente.

Ejemplos y problemas

1. Descompóngase en parte simétrica y antisimétrica las matrices

$$\begin{bmatrix} 2 & 7 & 0 \\ 6 & 2 & 2 \\ 2 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 4 & 0 & 0 \\ 0 & 0 & 9 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} -1 & 2 & -2 \\ 4 & -2 & -4 \\ 4 & 4 & 2 \end{bmatrix}.$$

y hállense las descomposiciones polares de las mismas.

2. Sea $\mathfrak{L}$ un espacio euclídeo de dos dimensiones y sea $\mathcal{A}$ una aplicación lineal del mismo que tiene dos vectores propios unitarios no ortogonales a_1 y a_2 correspondientes a los valores propios α_1 y α_2 ($\alpha_1 \neq \alpha_2$). Hállese la descomposición polar de la aplicación $\mathcal{A}$, si $(a_1, a_2) = \cos\varphi$, donde φ es un valor dado.

3. Indíquese el enunciado matricial de la aplicación de Cayley (teoremas 2 y 3).

4. Demuéstrese que toda matriz ortogonal U de orden tres, que no contiene la unidad entre sus valores propios, es de la forma

$$U = \frac{1}{1+\alpha^2+\beta^2+\gamma^2} \begin{bmatrix} \alpha^2+\beta^2-\gamma^2-1 & 2(\alpha+\beta\gamma) & 2(\beta-\alpha\gamma) \\ 2(-\alpha+\beta\gamma) & \alpha^2-\beta^2+\gamma^2-1 & 2(\gamma+\alpha\beta) \\ 2(-\beta-\alpha\gamma) & 2(-\gamma+\alpha\beta) & -\alpha^2+\beta^2+\gamma^2-1 \end{bmatrix}.$$

5. Las matrices cuadradas complejas de orden n forman, respecto a las operaciones de adición y de multiplicación por número, un espacio lineal $\mathfrak{L}$ de dimensión n^2. Las matrices E_{ij}, en la i-ésima fila y j-ésima columna de las cuales aparece la unidad, mientras que en las demás posiciones aparece el cero, constituyen una base del espacio $\mathfrak{L}$. Hagamos el espacio $\mathfrak{L}$ unitario, aceptando que las matrices E_{ij} ofrecen una base ortonormal en $\mathfrak{L}$, es decir, aceptando que el producto escalar de dos matrices A y B se calcula mediante la fórmula $(A, B) = \sum\sum \alpha_{ij}\overline{\beta}_{ij}$, donde α_{ij} son los elementos de la matriz A y β_{ij} son los elementos de la matriz B. La multiplicación de todas las matrices de $\mathfrak{L}$ por una matriz cualquiera X a la derecha ofrece una aplicación lineal del espacio $\mathfrak{L}$. Demuéstrese que:

a) las matrices unitarias tienen en $\mathfrak{L}$ la longitud igual a $\sqrt{n}$;

b) las multiplicaciones por matrices conjugadas transpuestas X y $\overline{X}'$ originan en $\mathfrak{L}$ aplicaciones conjugadas;

c) la multiplicación de las matrices de $\mathfrak{L}$ por una matriz unitaria origina en $\mathfrak{L}$ una aplicación unitaria:

d) la multiplicación por una matriz hermitiana origina en $\mathfrak{L}$ una aplicación simétrica, mientras que la multiplicación por una matriz hermitiana antisimétrica origina una aplicación antisimétrica.

6. Demuéstrese que la suma de dos aplicaciones proyectivas $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ es una aplicación proyectiva cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathfrak{G}$. Además, en este caso $\mathcal{P}_{\mathfrak{A}} + \mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}+\mathfrak{B}}$.

7. El producto de dos aplicaciones proyectivas $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ es una aplicación proyectiva cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}$ y $\mathcal{P}_{\mathfrak{B}}$ conmutan. Además, en este caso $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}\cap\mathfrak{B}}$.

8. Un subespacio $\mathfrak{A}$ está contenido en otro $\mathfrak{B}$ cuando, y sólo cuando, $\mathcal{P}_{\mathfrak{A}}\mathcal{P}_{\mathfrak{B}} = \mathcal{P}_{\mathfrak{A}}$.

9. Demuéstrese la fórmula $\mathcal{P}_{\mathfrak{A}\perp} = \mathcal{E} - \mathcal{P}_{\mathfrak{A}}$.

10. Un subespacio $\mathfrak{A}$ es invariante respecto a una aplicación arbitraria $\mathcal{A}$ cuando, y sólo cuando, la correspondiente aplicación proyectiva $\mathcal{P} = \mathcal{P}_{\mathfrak{A}}$ verifica la relación $\mathcal{P}\mathcal{A}\mathcal{P} = \mathcal{P}\mathcal{A}$.

11. Si las aplicaciones $\mathcal{A}$ y $\mathcal{B}$ con las descomposiciones espectrales $\mathcal{A} = \sum \alpha_j \mathcal{P}_j$ y $\mathcal{B} = \sum \beta_k Q_k$ conmutan, toda aplicación $\mathcal{P}_j$ es conmutable con cualquier aplicación Q_k.

12. Si una aplicación normal $\mathcal{A}$ conmuta con cualquier aplicación lineal $\mathcal{B}$, la aplicación $\mathcal{A}$ conmuta también con $\mathcal{B}^*$.

13. Toda aplicación unitaria $\mathcal{U}$ de un espacio unitario complejo puede ser representada en la forma $\mathcal{U} = e^{i\mathcal{A}}$, donde $\mathcal{A}$ es una aplicación simétrica. Recíprocamente, la aplicación $e^{i\mathcal{A}}$ es unitaria cualquiera que sea la aplicación simétrica $\mathcal{A}$.