PARTE II OPERADORES LINEALES

CAPÍTULO 7 MATRICES Y OPERADORES LINEALES

§ 56. Operadores

El elemento principal en la creación de los fundamentos del análisis matemático consiste en la introducción del concepto de función. De acuerdo con la definición, para prefijar una función es preciso indicar dos conjuntos X, Y de números reales y enunciar una regla, según la cual a todo número $x \in X$ se le pone en correspondencia el número único $y \in Y$. Esta regla representa precisamente una función unívoca de la variable real x, definida en el conjunto X.

Al realizar la idea general de la dependencia funcional, no es totalmente obligatorio exigir que X, Y sean unos conjuntos de números reales. Entendiendo por X, Y los más diversos conjuntos de elementos, llegamos a la siguiente definición que generaliza el concepto de función.

La regla, de acuerdo con la cual a todo elemento x de cierto conjunto no vacío X se le pone en correspondencia un único elemento y del conjunto no vacío Y, se llama *operador*. El resultado y de la aplicación del operador A al elemento x se designa así:

$$y = A(x), \qquad y = Ax \tag{56.1}$$

y se dice que el operador A *actúa de X en Y* o bien *aplica X en Y*.

El conjunto X se llama *campo de definición* del operador A. El elemento y de (56.1) se denomina *imagen* del elemento x y el propio x, *preimagen** del elemento y. La totalidad T_A de todas las imágenes se llama *campo de valores* (o *imagen*) del operador A. En el caso cuando cualquier elemento $y \in Y$ tiene preimagen y ésta es la única, la

* En algunas obras se emplea el término «imagen recíproca». (*N. del Tr.*)

regla (56.1) se llama *biunívoca*. El operador se denomina, además, *aplicación*, *transformación* u *operación*.

En lo que sigue consideraremos, en lo esencial, solamente los así llamados *operadores lineales*. Las peculiaridades distintivas de estos últimos consisten en lo siguiente. En primer lugar, el campo de definición de un operador lineal es siempre cierto espacio lineal o subespacio. En segundo lugar, las propiedades del operador lineal están íntimamente relacionadas con las operaciones sobre los vectores de un espacio lineal. Al estudiar los operadores lineales supondremos, generalmente, que los espacios se dan sobre un campo de números reales o complejos. Por operador se entenderá en lo sucesivo un operador lineal, siempre que no haya especificaciones especiales. En la teoría general de operadores los operadores lineales desempeñan un papel de la misma importancia que una línea recta y un plano desempeñan en el análisis matemático. A esto se debe, de hecho, la necesidad en una investigación detallada de ellos.

Sean dados los espacios lineales X, Y sobre un mismo campo P. Consideremos un operador A, cuyo campo de definición constituye el espacio X y el campo de valores es cierto conjunto de Y. El operador A se llama *lineal*, si

$$A(\alpha u + \beta v) = \alpha Au + \beta Av \tag{56.2}$$

para cualesquiera vectores $u, v \in X$ y cualesquiera números $\alpha, \beta \in P$.

Nos hemos encontrado, ya más de una vez, con los operadores lineales. De acuerdo con (9.8), como operador lineal sirve la magnitud de un segmento dirigido. Su campo de definición representa en sí el conjunto de todos los segmentos dirigidos de un eje, el campo de valores coincide con el conjunto de todos los números reales. Según se deduce de (21.2), una correspondencia isomorfa entre dos espacios lineales también será un operador lineal. Fijemos un subespacio L en un espacio lineal provisto de un producto escalar. Obtendremos dos operadores lineales, si a todo vector del espacio le asignamos o bien su proyección sobre el subespacio L o bien la perpendicular trazada desde este vector a L. La validez de esta afirmación se desprende de (30.5), (30.6).

Un operador que a todo vector x del espacio X pone en correspondencia el vector nulo del espacio Y es, evidentemente, lineal. Se llama operador *nulo* y se denota mediante el símbolo 0. Así pues,

$$0 = 0x.$$

Pondremos en correspondencia a todo vector $x \in X$ el mismo vector x. Se obtendrá un operador lineal E que actúa de X en X. Este operador se denomina *idéntico* u operador *unidad*. Por definición,

$$x = Ex.$$

Sea un operador lineal A que actúa del espacio X en el espacio Y. Construyamos un operador nuevo B, conforme a la inscripción $Bx = -Ax$. El operador obtenido B es también un operador lineal que actúa de X en Y. Se llama operador *opuesto* al operador A.

Fijemos, por fin, un número arbitrario α y a todo vector $x \in X$ le pondremos en correspondencia el vector $\alpha x \in X$. Un operador construído de este modo será, por supuesto, lineal. Se llama operador *escalar*. Cuando $\alpha = 0$ obtenemos un operador nulo, cuando $\alpha = 1$ obtenemos un operador idéntico.

A continuación expondremos el método general para obtener operadores lineales, pero ahora demos a conocer algunas de sus peculiaridades características. Según se deduce de (56.2), la correlación

$$A\left(\sum_{i=1}^{p} \alpha_i x_i\right) = \sum_{i=1}^{p} \alpha_i A x_i$$

tiene lugar para cualesquiera vectores x_i y los números α_i. De aquí se deduce, en particular, que todo operador lineal A transforma un vector nulo en vector nulo, es decir,

$$0 = A0.$$

El campo de valores T_A del operador lineal A es *un subespacio* del espacio Y. Si $z = Au$, $w = Av$, entonces el vector $\alpha z + \beta w$ será, a ciencia cierta, la imagen del vector $\alpha u + \beta v$ para cualesquiera números α, β. Por consiguiente, el vector $\alpha z + \beta w$ pertenece al campo de valores del operador A. La dimensión del subespacio T_A se llama *rango* del operador y se indica con r_A.

A la par con T_A examinemos un conjunto N_A de vectores $x \in X$ que satisfacen la igualdad

$$Ax = 0.$$

Este conjunto es también *un subespacio* y se denomina *núcleo* del operador A. La dimensión n_A del núcleo se denomina *defecto* del operador A.

El rango y el defecto no son características independientes del operador lineal A. Sea un espacio X de dimensión m. Descompongámoslo en la suma directa

$$X = N_A \dot{+} M_A, \tag{56.3}$$

donde N_A es el núcleo del operador A y M_A, cualquier subespacio complementario. Tomemos un vector arbitrario $x \in X$ y representémoslo en forma de una suma

$$x = x_N + x_M,$$

donde $x_N \in N_A$, $x_M \in M_A$. Si $y = Ax$, en virtud de la linealidad del operador A y de la condición $Ax_N = 0$, obtenemos que

$$y = Ax_M.$$

Por consiguiente, todo vector de T_A tiene por lo menos una preimagen de M_A.

Esta preimagen en M_A es en realidad *única*. Supongamos que para un vector $y \in T_A$ tenemos dos preimágenes x'_M, $x''_M \in M_A$. Ya que M_A es un subespacio, entonces $x'_M - x''_M \in M_A$. Pero en vista de que x'_M y x''_M son preimágenes de un mismo vector y, se tiene $x'_M - x''_M \in N_A$. Sólo el vector nulo es común para los subespacios M_A y N_A. Por esta razón $x'_M - x''_M = 0$, es decir, $x'_M = x''_M$.

De este modo, el operador A establece *una correspondencia biunívoca* entre los vectores de los subespacios T_A y M_A. En virtud de la linealidad del operador esta correspondencia es un isomorfismo. Por esta razón, las dimensiones de T_A y M_A coinciden y son iguales a r_A. De la descomposición (56.3) se infiere que

$$r_A + n_A = m. \tag{56.4}$$

Hemos de notar que el operador lineal A establece una correspondencia isomorfa entre el subespacio T_A y cualquier subespacio M_A de X, el cual en la suma directa con el núcleo del operador constituye todo el espacio X. Por esto podemos considerar que todo operador lineal A genera toda una familia de otros operadores lineales. En primer lugar, se trata del operador nulo definido en el núcleo N_A, es decir, del operador que actúa de N_A en 0. En segundo lugar, esto es un conjunto de operadores lineales que actúan de los subespacios M_A, complementarios al núcleo, en el subespacio T_A. Una circunstancia de gran importancia es que cada uno de los nuevos operadores coincide en su campo de definición con el operador A. Si $N_A = 0$, entonces $M_A = X$ y todo el segundo conjunto de operadores coincide con el operador A. En cambio, si $N_A = X$, entonces A es un operador nulo. Estas cuestiones las analizaremos nuevamente más adelante.

Ejercicios.

Demuéstrese que los siguientes operadores son lineales.

1. En un espacio lineal X viene dada una base. El operador A asigna a todo vector $x \in X$ su coordenada de número fijado.

2. En un espacio X provisto de un producto escalar se fija un vector x_0. El operador A asigna a todo vector $x \in X$ el producto escalar (x, x_0).

3. En un espacio V_3 se fija un vector x_0. El operador A asigna a todo vector $x \in V_3$ el producto escalar $[x, x_0]$.

4. El espacio X está formado por unos polinomios de coeficientes reales. El operador A asigna a todo polinomio su k-ésima derivada. Este operador lleva el nombre de operador de *diferenciación k-múltiple*.

5. En el espacio de los polinomios dependientes de la variable t el operador A asigna a todo polinomio $P(t)$ un polinomio $t \cdot P(t)$.

6. El espacio X está descompuesto en la suma directa de los subespacios S y T. Representemos todo vector $x \in X$ en forma de una suma $x = u + v$, donde $u \in S$, $v \in T$. El operador A asigna al vector x el vector u. Este operador lleva el nombre de operador de *proyección* sobre el subespacio S paralelamenre al subespacio T.

§ 57. Espacio lineal de los operadores

Fijemos dos espacios lineales X, Y sobre un mismo campo P y examinemos el conjunto ω_{XY} de todos los operadores lineales que actúan de X en Y. En el conjunto ω_{XY} se pueden introducir operaciones de adición de operadores y de multiplicación de un operador por los números de P, transformando de este modo ω_{XY} en un espacio lineal.

Dos operadores A, B, que actúan de X en Y, son *iguales*, si se cumple la igualdad

$$Ax = Bx$$

para todo vector $x \in X$. Es fácil comprobar que la razón de igualdad de los operadores es una razón de equivalencia. La igualdad de los operadores se designa de tal modo:

$$A = B.$$

El operador C se llama *suma* de los operadores A, B que actúan de X en Y, si se verifica la igualdad

$$Cx = Ax + Bx$$

para todo vector $x \in X$. La suma de los operadores se indica

$$C = A + B.$$

Por definición, se pueden sumar cualesquiera operadores que actúan de X en Y. Si A, B son unos operadores lineales de ω_{XY}, su suma será también un operador lineal de ω_{XY}. Para cualesquiera vectores u, $v \in X$ y cualesquiera números α, $\beta \in P$ se tiene

$$\begin{aligned} C(\alpha u + \beta v) &= A(\alpha u + \beta v) + B(\alpha u + \beta v) = \\ &= \alpha Au + \beta Av + \alpha Bu + \beta Bv = \\ &= \alpha(Au + Bu) + \beta(Av + Bv) = \alpha Cu + \beta Cv. \end{aligned}$$

La operación de adición de los operadores es *una operación algebraica*. Es, además, asociativa. En efecto, sean A, B, C tres operadores lineales arbitrarios de ω_{XY}. Entonces, para todo vector $x \in X$ se verifican las igualdades

$$\begin{aligned} ((A + B) + C)x &= (A + B)x + Cx = Ax + Bx + Cx = \\ &= Ax + (Bx + Cx) = Ax + (B + C)x = (A + (B + C))x \end{aligned}$$

Pero esto significa que

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

La operación de adición de los operadores es *conmutativa*. Si A, B son unos operadores cualesquiera de ω_{XY} y x es un vector de X, entonces

$$(A + B)\,x = Ax + Bx = Bx + Ax = (B + A)\,x,$$

es decir,

$$A + B = B + A.$$

Ahora es fácil mostrar que el conjunto ω_{XY} con la operación de adición de los operadores introducida es un grupo abeliano. Este conjunto tiene por lo menos un elemento nulo, por ejemplo, el operador nulo. Todo elemento de ω_{XY} tiene por lo menos un elemento opuesto, por ejemplo, un operador inverso. Todo lo demás proviene del teorema 7.1.

Según se deduce del mismo teorema, la operación de adición de los operadores tiene su inversa. Llamémosla *sustracción* y hagamos uso de los símbolos y las propiedades aceptadas.

El operador C se llama *producto del operador* A, que actúa de X en Y, *por el número* λ del campo P, si se verifica la igualdad

$$Cx = \lambda \cdot Ax$$

para todo vector $x \in X$. Este producto se designa con

$$C = \lambda A.$$

El producto de un operador lineal de ω_{XY} por un número es también un operador lineal de ω_{XY}. En efecto, para cualesquiera vectores u, $v \in X$ y cualesquiera números α, $\beta \in P$ tenemos

$$C\,(\alpha u + \beta v) = \lambda A\,(\alpha u + \beta v) = \lambda\,(\alpha A u + \beta A v) =$$
$$= \alpha\,(\lambda A u) + \beta\,(\lambda A v) = \alpha C u + \beta C v.$$

No es difícil convencerse de que en la operación de adición de los operadores y en la de multiplicación de un operador por un número se cumplen todas las propiedades que determinan un espacio lineal. Por consiguiente, el conjunto ω_{XY} de todos los operadores lineales, que actúan del espacio lineal X en el espacio lineal Y, forma un espacio lineal nuevo. De aquí se infere que desde el punto de vista de las operaciones de multiplicación de un operador por un número, de adición y sustracción de los operadores, se cumplen todas las reglas para las transformaciones equivalentes de las expresiones algebraicas operacionales. En lo sucesivo dichas reglas ya no serán un objeto de especificación especial.

Hemos de notar que nunca usamos la relación mutua de los espacios lineales X, Y. Pueden ser tanto diferentes como coincidentes. El conjunto ω_{XX} de los operadores lineales, que actúan del espacio X en el mismo espacio X, será uno de los fundamentales objetos de nuestra investigación. Los operadores citados se llamarán *operadores lineales en* X.

Ejercicios.

1. Demuéstrese que al multiplicar un operador por un número no nulo, el rango y el defecto del operador no varían.
2. Demuéstrese que el rango de la suma de unos operadores no es superior a la suma de rangos de los sumandos.
3. Demuéstrese que un conjunto de operadores lineales de ω_{XY}, cuyos campos de valores pertenecen a un mismo subespacio, forma de por sí un subespacio lineal.
4. Demuéstrese que un sistema de dos operadores no nulos de ω_{XY}, cuyos campos de valores son distintos, es linealmente independiente.
5. Demuéstrese que un espacio de operadores lineales que actúan en V_1, es unidimensional.

§ 58. Anillo de los operadores

Consideremos tres espacios lineales X, Y, Z sobre un mismo campo P. Sea A un operador que actúa de X en Y y sea B un operador que actúa de Y en Z.

El operador C que actúa de X en Z se llama *producto del operador B por el operador A*, si se verifica la igualdad

$$Cx = B\,(Ax)$$

para todo vector $x \in X$. El producto de los operadores B y A se designa por

$$C = BA.$$

El producto de los operadores lineales es también un operador lineal. Para cualesquiera vectores u, $v \in X$ y cualesquiera números α, $\beta \in P$ se tiene

$$C\,(\alpha u + \beta v) = B\,(A\,(\alpha u + \beta v)) = B\,(\alpha Au + \beta Av) =$$
$$= \alpha B\,(Au) + \beta B\,(Av) = \alpha Cu + \beta Cv.$$

La multiplicación de los operadores no es una operación algebraica, dado que el producto no está definido para todo par de operadores. No obstante, siendo factible, la operación de multiplicación de los operadores posee unas propiedades bien naturales. A saber:

$$\begin{aligned}
&1)\ (AB)\,C = A\,(BC),\\
&2)\ \lambda\,(AB) = (\lambda B)\,A = B\,(\lambda A),\\
&3)\ (A + B)\,C = AC + BC,\\
&4)\ A\,(B + C) = AB + AC
\end{aligned} \tag{58.1}$$

para cualesquiera operadores A, B, C y todo número λ de P, si, por supuesto, las expresiones correspondientes están definidas.

La demostración de todas estas propiedades se efectúa de una manera igual, razón por la cual nos limitaremos a estudiar solamente la primera de las propiedades. Sean X, Y, Z, U unos espacios lineales fijados y sean A, B, C cualesquiera operadores lineales, de los cuales A actúa de X en Y, B actúa de Y en Z y C, de Z en U. Observemos ante todo que en la igualdad 1 están definidos ambos operadores, $(AB)\,C$ y $A\,(BC)$. Para todo vector $x \in X$ tenemos

$$((AB)\,C)\,x = AB\,(Cx) = A\,(B\,(Cx)),$$
$$(A\,(BC))\,x = A\,(BCx) = A\,(B\,(Cx)),$$

de donde se deduce la validez de la igualdad 1.

Consideremos otra vez el conjunto ω_{XX} de los operadores lineales que actúan en el espacio X. Para cualesquiera dos operadores de ω_{XX} se han definido tanto la suma como el producto. De acuerdo con las propiedades 3, 4, ambas operaciones están relacionadas por una ley distributiva. Por esto el conjunto ω_{XX} de operadores lineales representa en sí *un anillo*. En lo sucesivo mostraremos que el anillo de los operadores es *no conmutativo*. Desde luego, por casualidad puede ocurrir que para algún par de operadores A, B se cumpla la correlación $AB = BA$. Estos operadores se llamarán *conmutables*. En particular, un operador idéntico es conmutable con cualquier operador.

En el anillo de los operadores lineales, al igual que en todo otro anillo, el producto de cualquier operador por un operador nulo es también un operador nulo. La ley distributiva relaciona con la multiplicación no sólo la suma de operadores, sino también la diferencia entre éstos. Un anillo de los operadores lineales es a la vez un espacio lineal, por lo cual, para la diferencia entre los operadores resulta lícita la fórmula

$$A - B = A + (-1)\,B.$$

La propiedad 2 de (58.1) muestra la relación entre la multiplicación de los operadores en un anillo y la multiplicación por un número. Desde luego, quedan en vigor también todas las correlaciones que provienen de las propiedades de los espacios lineales.

Ejercicios.

1. Denotemos mediante D el operador de diferenciación y mediante T, el operador de multiplicación por t, en un espacio de polinomios que dependen de la variable t. Demuéstrese que $DT \neq TD$. Hállese el operador $DT - TD$.
2. Fijemos cierto operador B del espacio ω_{XX}. Demuéstrese que el conjunto de los operadores A, para los cuales $BA = 0$, forma en ω_{XX} un subespacio.
3. Demuéstrese que el rango de un producto de los operadores no es superior al rango de cada uno de los factores.
4. Demuéstrese que el defecto de un producto de operadores no es inferior al defecto de cada uno de los factores.
5. Demuéstrese que en el anillo ω_{XX} de operadores lineales hay divisores de cero.

§ 59. Grupo de operadores regulares

Los operadores lineales que actúan en el espacio X forman un grupo abeliano de adición. Pero entre tales operadores pueden indicarse unos conjuntos que representan en sí los grupos de multiplicación. Estos grupos están ligados con los llamados operadores regulares.

Un operador que actúa en un espacio lineal se llama *regular*, si su núcleo consta sólo de un vector nulo. Un operador no regular se llama *degenerado*. Serán regulares, por ejemplo, el operador idéntico y el operador escalar, si no es nulo. A veces, con el operador A, que actúa en el espacio X, se puede ligar cierto operador regular, incluso cuando A sea degenerado. En efecto, sea T_A el campo de valores del operador A y sea N_A su núcleo. Si T_A y N_A no tienen vectores no nulos comunes, entonces, de acuerdo con (56.4), tenemos

$$X = N_A \dot{+} T_A.$$

Como ya se ha observado, el operador A engendra un conjunto de otros operadores que actúan de cualquier subespacio, complementario al núcleo N_A, en el subespacio de valores T_A. En el caso considerado el operador A engendra un operador que actúa de T_A en T_A. Este operador será regular, puesto que transforma en cero sólo el vector nulo de T_A.

Los operadores regulares poseen una serie de peculiaridades remarcables. Para los operadores de este tipo el defecto es igual a cero, por lo cual de la fórmula (56.4) se infiere que el rango del operador regular coincide con la dimensión del espacio. Si un operador regular A actúa en el espacio X, el campo de valores T_A coincide con X. De este modo, todo vector de X es una imagen de cierto vector de X. Esta propiedad del operador regular es *equivalente* a su definición.

Una propiedad importante del operador regular consiste en la unicidad de la preimagen de todo vector del espacio. Efectivamente, supongamos que para cierto vector y existen dos preimágenes u, v. Esto significa que

$$Au = y, \qquad Av = y.$$

Pero en este caso

$$A(u - v) = 0.$$

Por definición de operador regular, el núcleo se compone sólo del vector nulo. Por esto, $u - v = 0$, esto es, $u = v$. La propiedad demostrada es también *equivalente* a la definición de operador regular. Esta propiedad ya se ha mencionado, de hecho, en el § 56.

Un producto de cualquier número finito de operadores regulares es también un operador regular. Evidentemente, basta demostrar esta afirmación para dos operadores. Sean A, B cualesquiera operadores regulares que actúan en el mismo espacio X. Consideremos la ecuación

$$BAx = 0. \tag{59.1}$$

De acuerdo con la definición de la multiplicación de operadores, esta ecuación significa que

$$B\,(Ax) = 0.$$

El operador B es regular, por lo cual de la última ecuación se deduce que $Ax = 0$. Pero A es también un operador regular y de aquí proviene que $x = 0$. Así pues, la ecuación (59.1) se satisface solamente por el vector nulo, es decir, el operador BA es regular.

Una suma de operadores regulares ya no será obligatoriamente un operador regular. Si A es un operador regular, lo será también el operador $(-1)\,A$. Pero la suma de estos operadores es un operador nulo que es degenerado.

Examinemos el conjunto de operadores regulares que actúan en un mismo espacio lineal. En dicho conjunto la multiplicación de los operadores es una operación *algebraica* y, además, *asociativa*. Entre los operadores regulares figura también el operador idéntico E, que desempeña el papel de la unidad. Efectivamente, es fácil comprobar que para todo operador A que actúa en el espacio X se verifica

$$AE = EA = A.$$

Si probamos que para todo operador regular A existe un operador regular que, siendo multiplicado por A, da un operador idéntico, esto será el testimonio de que el conjunto de todos los operadores regulares forma un grupo de multiplicación.

Sea A un operador regular. Como se sabe, para todo vector $y \in X$ existe un vector, y sólo uno, $x \in X$, relacionado con y mediante la expresión

$$y = Ax. \tag{59.2}$$

Por consiguiente, a todo vector $y \in X$ se puede poner en correspondencia el único vector $x \in X$, para el cual y es una imagen suya. La correspondencia construida es un cierto operador. Se llama operador *inverso* del operador A y se designa mediante el símbolo A^{-1}. Si se verifica la desigualdad (59.2), entonces

$$x = A^{-1}y. \tag{59.3}$$

Demostremos que el operador inverso es lineal y regular.

El producto está definido para cualesquiera operadores, no sólo para los lineales. Por esto, de la definición de operador inverso se desprende que

$$A^{-1}A = AA^{-1} = E. \tag{59.4}$$

Para demostrar estas igualdades, basta aplicar a ambos miembros de (59.2) el operador A^{-1} y a cada uno de los miembros de (59.3), el operador A.

Tomemos unos vectores cualesquiera $u, v \in X$ y cualesquiera números $\alpha, \beta \in P$ y consideremos un vector

$$z = A^{-1}(\alpha u + \beta v) - \alpha A^{-1}u - \beta A^{-1}v.$$

Apliquemos ahora a ambos miembros de la igualdad el operador A. Tomando en consideración la linealidad del operador A y las correlaciones (59.4), concluimos que $Az = \mathbf{0}$. Puesto que el operador A es regular, esto quiere decir $z = 0$. Por consiguiente,

$$A^{-1}(\alpha u + \beta v) = \alpha A^{-1}u + \beta A^{-1}v,$$

es decir, el operador A^{-1} es lineal.

Es fácil mostrar que el operador A^{-1} es regular. Para todo vector y del núcleo del operador A^{-1} tenemos

$$A^{-1}y = 0.$$

Apliquemos a ambos miembros de esta igualdad el operador A. Como A es un operador lineal, entonces $A\mathbf{0} = \mathbf{0}$. Teniendo presentes las correlaciones (59.4), concluimos que $y = 0$. Así pues, el núcleo del operador A^{-1} se compone sólo de un vector nulo, es decir, A^{-1} es el operador regular.

De este modo, el conjunto de operadores regulares representa en sí *un grupo de multiplicación*. Un poco más adelante mostraremos que este grupo es *no conmutativo*.

Con ayuda de los operadores regulares pueden construirse también unos grupos conmutativos. Sea A un operador arbitrario que actúa en el espacio X. Para cualquier número positivo entero p hallemos el p-ésimo *grado* del operador A mediante la igualdad

$$A^p = \overbrace{A \cdot A \ldots A}^{p}, \tag{59.5}$$

donde en el segundo miembro están contenidos p factores. En virtud de la asociatividad de la operación de multiplicación, el operador A^p se define unívocamente. Desde luego, este operador es lineal.

Para cualesquiera números positivos y enteros p, r de (59.5) se deduce que

$$A^pA^r = A^{p+r}. \tag{59.6}$$

Si se considera, por definición, que

$$A^0 = E$$

para todo operador A, entonces la fórmula (59.6) tendrá lugar para cualesquiera números no negativos y enteros p, r.

Supongamos que A es un operador regular, entonces para todo r no negativo será regular también el operador A^r. Por consiguiente, para él existe un operador inverso. De acuerdo con las fórmulas (7.2), (59.5), tenemos

$$(A^r)^{-1} = (A^{-1})^r = \overbrace{A^{-1}A^{-1} \ldots A^{-1}}^{r}. \qquad (59.7)$$

Consideraremos también, por definición, que

$$A^{-r} = (A^r)^{-1}.$$

Teniendo en cuenta las fórmulas (59.5), (59.7) y tomando en consideración que $AA^{-1} = A^{-1}A$, no es difícil demostrar la correlación

$$A^p A^{-r} = A^{-r} A^p$$

para cualesquiera p, r no negativos y enteros. Esto significa que la fórmula (59.6) es válida para cualesquiera números enteros p, r.

Tomemos ahora un operador regular A y formemos el conjunto ω_A de los operadores del tipo A^p para todos los p enteros. En este conjunto la multiplicación de los operadores es una operación *algebraica* y, como se deduce de (59.6), *conmutativa*. Todo operador A^p tiene su inverso, igual a A^{-p}. En el conjunto ω_A figura también el operador idéntico E. Por consiguiente, el conjunto ω_A representa en sí *un grupo conmutativo de multiplicación*. Este grupo se denomina *cíclico*, generado por el operador A.

Ejercicios.

1. Demuéstrese que si para dos operadores lineales A, B de ω_{XX} se verifica la correlación $AB = E$, entonces ambos operadores son regulares.

2. Demuéstrese que para que los operadores A, B, de ω_{XX} sean regulares, es necesario y suficiente que sean regulares los operadores AB y BA.

3. Demuéstrese que si el operador A es regular y el número $\alpha \neq 0$, entonces el operador αA es también regular y $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$.

4. Demuéstrese que $T_A \subset N_A$ cuando, y sólo cuando, $A^2 = 0$.

5. Demuéstrese que para cualquier operador A se cumplen las correlaciones

$$N_A \subseteq N_{A^2} \subseteq N_{A^3} \ldots, \quad T_A \supseteq T_{A^2} \supseteq T_{A^3} \ldots$$

6. Demuéstrese que el operador P es un operador de proyección cuando, y sólo cuando, $P^2 = P$. ¿Qué representan en sí los subespacios N_P y T?

7. Demuéstrese que si P es un operador de proyección, entonces $E - P$ es también un operador de proyección.

8. Demuéstrese que si el operador A satisface la igualdad $A^m = 0$ para algún número positivo entero m, entonces el operador $\alpha E - A$ será regular para cualquier número $\alpha \neq 0$.

9. Demuéstrese que el operador lineal A, para el cual $E + \alpha_1 A + \alpha_2 A^2 + \ldots + \alpha_n A^n = 0$, es regular.

10. Demuéstrese que si A es un operador regular, entonces o bien todos los operadores en el grupo cíclico ω_A son distintos o bien cierta potencia del operador A coincide con el operador idéntico.

§ 60. Matriz del operador

Demos a conocer un método general para construir el operador lineal que actúa del espacio m-dimensional X en el espacio n-dimensional Y. Supongamos que a los vectores de la base $e_1, e_2, \ldots, e_m$ del espacio X les están asignados unos vectores $f_1, f_2, \ldots, f_m$ del espacio Y. En este caso *existe* un operador lineal A *y es*, además, *único*, que actúa de X en Y y que transforma todo vector e_k en el vector correspondiente f_k.

Supongamos que el operador buscado A existe. Tomemos un vector arbitrario $x \in X$ y representémoslo en forma de un desarrollo

$$x = \xi_1 e_1 + \xi_2 e_2 + \ldots + \xi_m e_m.$$

Entonces

$$Ax = A\left(\sum_{k=1}^{m} \xi_k e_k\right) = \sum_{k=1}^{m} \xi_k A e_k = \sum_{k=1}^{m} \xi_k f_k.$$

El segundo miembro de las correlaciones se determina unívocamente por el vector x y las imágenes de la base. Por eso la igualdad obtenida demuestra la unicidad del operador A, si éste existe. Por otra parte, podemos definir el operador A precisamente mediante esta igualdad, es decir, poner

$$Ax = \sum_{k=1}^{m} \xi_k f_k.$$

El operador obtenido, como es fácil de comprobar, es un operador lineal que actúa de X en Y y transforma, a la vez, todo vector e_k en el vector correspondiente f_k. El campo de valores T_A del operador A coincide con la cápsula lineal del sistema de vectores $f_1, f_2, \ldots, f_m$.

Ahora podemos enunciar una deducción importante: *el operador lineal A que actúa del espacio X en el espacio Y está enteramente definido mediante la totalidad de imágenes*

$$Ae_1, Ae_2, \ldots, Ae_m$$

para cualquier base fijada

$$e_1, e_2, \ldots, e_m$$

del espacio X.

Fijemos en el espacio X la base $e_1, e_2, \ldots, e_m$ y en el espacio Y, la base $q_1, q_2, \ldots, q_n$. El vector e_1 se transforma por el operador A en cierto vector Ae_1 del espacio Y, el cual, como todo vector de este espacio, puede ser desarrollado por vectores básicos

$$Ae_1 = a_{11}q_1 + a_{21}q_2 + a_{n1}q_n.$$

Análogamente,

$$Ae_2 = a_{12}q_1 + a_{22}q_2 + \ldots a_{n2}q_n,$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$Ae_m = a_{1m}q_1 + a_{2m}q_2 + \ldots + a_{nm}q_n.$$

Los coeficientes a_{ij} de estas correlaciones determinan una matriz A_{qe} de n filas y m columnas

$$A_{qe} = \begin{pmatrix} a_{11}a_{12}\ldots a_{1m} \\ a_{21}a_{22}\ldots a_{2m} \\ \ldots\ldots\ldots \\ a_{n1}a_{n2}\ldots a_{nm} \end{pmatrix}$$

que se denomina *matriz del operador A en bases elegidas*.

Como columnas de la matriz del operador sirven las coordenadas de los vectores $Ae_1, Ae_2, \ldots, Ae_m$ respecto de la base $q_1, q_2, \ldots, q_n$. Con el fin de determinar el elemento a_{ij} de la matriz del operador A hace falta aplicar el operador al vector e_j y tomar la i-ésima coordenada en la imagen Ae_j. Si, para abreviar, designamos mediante $\{x\}_i$ la i-ésima coordenada del vector x, entonces $a_{ij} = \{Ae_j\}_i$. En lo sucesivo haremos uso del método descrito para determinar los elementos de la matriz del operador.

Consideremos un vector arbitrario $x \in X$ y su imagen $y = Ax$. Aclaremos de qué modo se expresan las coordenadas del vector y en términos de las coordenadas del vector x y los elementos de la matriz del operador. Sea

$$x = \sum_{j=1}^{m} \xi_j e_j, \qquad y = \sum_{i=1}^{n} \eta_i q_i. \tag{60.1}$$

Calculamos

$$Ax = A\Big(\sum_{j=1}^{m} \xi_j e_j\Big) = \sum_{j=1}^{m} \xi_j A e_j = \sum_{j=1}^{m} \xi_j \sum_{i=1}^{n} a_{ij} q_i = \sum_{i=1}^{n} \Big(\sum_{j=1}^{m} \xi_j a_{ij}\Big) q_i.$$

Al comparar el segundo miembro de estas igualdades con el desarrollo (60.1) para el vector y, concluimos que deben cumplirse las igualdades

$$\sum_{j=1}^{m} a_{ij}\xi_j = \eta_i$$

para $i = 1, 2, \ldots, n$, es decir

$$\begin{aligned} a_{11}\xi_1 + a_{12}\xi_2 + \ldots + a_{1m}\xi_m &= \eta_1, \\ a_{21}\xi_1 + a_{22}\xi_2 + \ldots + a_{2m}\xi_m &= \eta_2, \\ \ldots\ldots\ldots\ldots\ldots \\ a_{n1}\xi_1 + a_{n2}\xi_2 + \ldots + a_{nm}\xi_m &= \eta_n. \end{aligned} \tag{60.2}$$

De este modo, todo operador lineal genera, cuando están fijadas las bases en los espacios X, Y, las correlaciones (60.2) que relacionan entre sí las coordenadas de la imagen y las de la preimagen. Con el fin de determinar las coordenadas de la imagen según las coordenadas de la preimagen, basta calcular los primeros miembros de estas correlaciones. Para determinar las coordenadas de la preimagen según las coordenadas conocidas del vector y hemos de resolver el sistema de ecuaciones algebraicas lineales (60.2) respecto de las incógnitas $\xi_1, \xi_2, \ldots, \xi_m$. La matriz de este sistema coincide con la matriz del operador.

Las correlaciones (60.2) establecen una relación profunda entre los operadores lineales y sistemas de ecuaciones algebraicas lineales. En particular, de (60.2) proviene que el rango del operador coincide con el de la matriz del operador y la dimensión del núcleo coincide con el número de soluciones fundamentales del sistema homogéneo reducido. De este hecho se deduce trivialmente la fórmula (56.4) y una serie de otras fórmulas.

A la relación existente entre los sistemas de ecuaciones algebraicas lineales y los operadores lineales volveremos con frecuencia. Pero primeramente demostremos que entre los operadores y las matrices que, de hecho, determinan los sistemas del tipo (60.2) existe una correspondencia biunívoca. Ya hemos mostrado que todo operador A determina cierta matriz A_{qe}, siendo fijadas las bases. Tomemos ahora una matriz arbitraria A_{qe} de dimensiones $n \times m$. Siendo fijadas las bases en los espacios X, Y, las correlaciones (60.2) a todo vector $x \in X$ le ponen en correspondencia cierto vector $y \in Y$. Es fácil comprobar que esta correspondencia es un operador lineal. Construyamos la matriz del operador dado en las mismas bases. Todas las coordenadas del vector e_j son nulas, a excepción de la j-ésima coordenada que es igual a uno. De (60.2) se deduce que las coordenadas del vector Ae_j coinciden con los elementos de la j-ésima columna de la matriz A_{qe} y, por ende, $\{Ae_j\}_i = a_{ij}$. Por lo tanto, la matriz del operador construido coincide con la inicial A_{qe}.

Así pues, toda matriz $n \times m$ es una matriz de cierto operador lineal que actúa del espacio m-dimensional X en el espacio n-dimensional Y, para las bases fijadas en dichos espacios. De este modo *se establece una correspondencia biunívoca, para unas bases fijadas cualesquiera, entre los operadores lineales y las matrices rectangulares.* En este caso, tanto los espacios lineales como las matrices se consideran, desde luego, sobre el mismo campo P.

He aquí algunos de los ejemplos. Sea 0 un operador nulo. Tenemos

$$\{0e_j\}_i = \{0\}_i = 0.$$

Por consiguiente, todos los elementos de la matriz del operador nulo son iguales a cero. Tal matriz se llama *nula* y se designa por el símbolo 0.

Tomemos ahora un operador idéntico E. Para este operador encontramos

$$\{Ee_j\}_i=\{e_j\}_i=\begin{cases}1, & \text{si } i=j,\\ 0, & \text{si } i\neq j.\end{cases}$$

Por esta razón la matriz del operador idéntico tiene la forma siguiente. Es una matriz cuadrada en cuya diagonal principal se disponen las unidades, mientras que los demás lugares son ocupados por ceros. La matriz de un operador idéntico se denomina matriz *unidad* y se designa con la letra E.

Nos encontraremos frecuentemente con un tipo más de matrices. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ unos números arbitrarios del campo P. Construyamos una matriz cuadrada Λ la que tiene dichos números dispuestos por la diagonal principal, mientras que en los restantes lugares se disponen ceros, es decir,

$$\Lambda=\begin{pmatrix}\lambda_1 & & 0\\ & \lambda_2 & \\ & & \ddots & \\ 0 & & & \lambda_n\end{pmatrix}.$$

Las matrices de tal índole se llaman *diagonales*. Si todos los elementos diagonales son iguales entre sí, la matriz se denomina *escalar*. En particular, la matriz unidad es escalar. Llamemos también diagonales las matrices rectangulares, construidas de modo análogo. Si nos referimos a las correlaciones (60.2), estableceremos fácilmente cómo actúa el operador lineal con la matriz Λ. Este operador "estira" la i-ésima coordenada de cualquier vector λ_i veces para todo i.

Ejercicios.

1. En un espacio de polinomios de grado no superior a n se halla fijada la base $1, t, t^2, \ldots, t^n$. ¿Qué forma tiene en esta base la matriz del operador de diferenciación?

2. En el espacio X se halla dado el operador P de proyección sobre el subespacio S, paralelamente al subespacio T. Fijamos en X cualquier base formada como la reunión de los subespacios S y T. ¿Qué forma tienen en dicha base las matrices de los operadores P y $E-P$?

3. Supongamos que el operador lineal A actúa de X en Y. Designemos mediante M_A un subespacio en X, complementario al núcleo N_A, y mediante R_A un subespacio en Y complementario a T_A. ¿Cómo variará la matriz del operador A, si al elegir las bases en X, Y, empleamos las bases de unos o de todos los subespacios indicados?

§ 61. Operaciones sobre las matrices

Ya se ha mostrado que, siendo fijadas las bases en los espacios, todo operador lineal se define unívocamente por medio de su matriz. Por esto, las operaciones con los operadores examinadas más arriba conducen a las operaciones bien determinadas sobre las matrices. En los problemas que actualmente son de interés para nosotros, la elección de una base no juega ningún papel, razón por la cual los operadores y sus matrices se designan mediante las mismas letras, omitiendo todos los índices concernientes a las bases.

Supongamos que dos operadores iguales actúan del espacio m-dimensional X en el espacio n-dimensional Y. Puesto que los operadores iguales se ponen de manifiesto de una manera igual, cualquiera que sea la situación, tendrán una misma matriz. Esto nos ofrece un fundamento para enunciar la siguiente definición.

Las matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} se llaman *iguales*, si

$$a_{ij} = b_{ij}$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La igualdad de las matrices se indica del modo siguiente:

$$A = B.$$

Supongamos ahora que del espacio X en el espacio Y actúan dos operadores A, B. Consideraremos el operador $C = A + B$. Designemos los elementos de las matrices de estos operadores con c_{ij}, a_{ij}, b_{ij}, respectivamente. Con arreglo a lo dicho anteriormente, $c_{ij} = \{Ce_j\}_i$. Teniendo presentes la definición de la suma de operadores y las propiedades de las coordenadas de los vectores respecto a las operaciones sobre ellos, obtendremos

$$c_{ij} = \{Ce_j\}_i = \{(A + B)\, e_j\}_i = \{Ae_j + Be_j\}_i = \\ = \{Ae_j\}_i + \{Be_j\}_i = a_{ij} + b_{ij}.$$

Por esta razón:

Se llama *suma* de dos matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} una matriz C de las mismas dimensiones con los elementos c_{ij}, si

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

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La suma de las matrices se designa con

$$C = A + B.$$

Se llama *diferencia* de dos matrices A, B de dimensiones iguales $n \times m$ con los elementos a_{ij}, b_{ij} una matriz C de las mismas dimensiones con los elementos c_{ij}, si

$$c_{ij} = a_{ij} - b_{ij}$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. La diferencia entre las matrices se indica con

$$C = A - B.$$

Consideremos un operador A que actúa de X en Y y el operador $C = \lambda A$ para cierto número λ. Si a_{ij}, c_{ij} son los elementos de las matrices de dichos operadores, entonces

$$c_{ij} = \{Ce_j\}_i = \{\lambda A e_j\}_i = \lambda \{Ae_j\}_i = \lambda a_{ij},$$

y llegamos a la siguiente definición:

Se llama *producto de la matriz* A de dimensiones $n \times m$ con los elementos a_{ij} *por el número* λ una matriz C de las mismas dimensiones con los elementos c_{ij}, si

$$c_{ij} = \lambda a_{ij}.$$

para $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. El producto de una matriz por un número se designa así

$$C = \lambda A.$$

Sean dados un espacio m-dimensional X y un espacio n-dimensional Y sobre un mismo campo P. Según lo demostrado más arriba, siendo fijadas las bases en X, Y, entre el conjunto ω_{XY} de todos los operadores que actúan de X en Y y el conjunto de todas las matrices de dimensiones $n \times m$ con los elementos del campo P tiene lugar una correspondencia biunívoca. Puesto que las operaciones sobre las matrices se introducían en concordancia con las operaciones sobre los operadores, el conjunto de las matrices $n \times m$, al igual que el conjunto ω_{XY}, representa en sí un espacio lineal.

Es fácil mostrar una de las bases del espacio de matrices. Ésta será, por ejemplo, el sistema de matrices $A^{(kp)}$ para $k = 1, 2, \ldots, n$, $p = 1, 2, \ldots, m$, donde los elementos $a_{ij}^{(kp)}$ de la matriz $A^{(kp)}$ se definen por las siguientes igualdades:

$$a_{ij}^{(kp)} = \begin{cases} 1, & \text{si } i = k, \; j = p, \\ 0 & \text{en todos los demás casos.} \end{cases}$$

En el espacio ω_{XY} de base sirve un sistema de operadores con matrices $A^{(kp)}$. De aquí concluimos que *un espacio lineal de operadores que actúan de* X *en* Y *es un espacio de dimensión finita y su dimensión es igual al producto* mn.

Supongamos dados tres espacios lineales X, Y, Z, con la particularidad de que el operador A actúa de X en Y, el B, de Y en Z. Sean las dimensiones de los espacios citados m, n, p, respectivamente. Convengamos en considerar que en X, Y, Z están fijadas las bases $e_1, \ldots, e_m$, $q_1, \ldots, q_n$, $r_1, \ldots, r_p$. El operador A tiene la matriz $n \times m$ con los elementos a_{ij} y

$$Ae_j = \sum_{s=1}^{n} a_{sj} q_s.$$

El operador B tiene la matriz $p \times n$ con los elementos b_{ij} y

$$Bq_s = \sum_{k=1}^{p} b_{ks} r_k.$$

Al investigar la matriz del operador $C = BA$, llegamos a la conclusión que debe tener dimensiones $p \times m$, y sus elementos son como siguen:

$$c_{ij} = \{Ce_j\}_i = \{BAe_j\}_i = \{B(\sum_{s=1}^{n} a_{sj} q_s)\}_i =$$
$$= \{\sum_{s=1}^{n} a_{sj} Bq_s\}_i = \{\sum_{s=1}^{n} a_{sj} \sum_{k=1}^{p} b_{ks} r_k\}_i =$$
$$= \{\sum_{k=1}^{p} (\sum_{s=1}^{n} b_{ks} a_{sj}) r_k\}_i = \sum_{s=1}^{n} b_{is} a_{sj}.$$

La fórmula obtenida nos dicta la siguiente definición.

El producto de la matriz B de dimensiones $p \times n$ con los elementos b_{ij} *por la matriz* A de dimensiones $n \times m$ con los elementos a_{ij} es la matriz C de dimensiones $p \times m$ con los elementos c_{ij}, si

$$c_{ij} = \sum_{s=1}^{n} b_{is} a_{sj} \tag{61.1}$$

para $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, m$. El producto de las matrices se designa con

$$C = BA.$$

De este modo, el producto se ha definido sólo para aquellas matrices, en las cuales el número de columnas del factor izquierdo equivale al número de filas del factor derecho. El elemento de la matriz del producto que se dispone en la intersección de la i-ésima fila y la j-ésima columna es igual a la suma de productos de todos los elementos de la i-ésima fila del factor izquierdo por los elementos correspondientes de la j-ésima columna del factor derecho.

Recordemos una vez más que entre los operadores lineales y las matrices tiene lugar una correspondencia biunívoca. Las operaciones sobre las matrices se introducían conforme a las operaciones con los operadores. Por esta razón, la multiplicación de matrices está ligada, mediante la correlación (58.1), con la sumación y la multiplicación de la matriz por un número.

Ya se ha observado que el anillo de operadores y el grupo de todos los operadores regulares, que actúan en un espacio lineal, son no conmutativos. Para demostrar esta afirmación es suficiente, evidentemente, hallar dos matrices cuadradas A, B tales que sea $AB \neq BA$.

Tomemos, por ejemplo,

$$A=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \qquad B=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}.$$

Calculamos:

$$AB=\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}, \qquad BA=\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix},$$

y el carácter no conmutativo de la multiplicación queda demostrado.

La operación de multiplicación de las matrices permite anotar cómodamente las correlaciones del tipo (60.2). Designemos con x_e la matriz de dimensiones $m \times 1$, compuesta por las coordenadas del vector x, y mediante y_q, la matriz de dimensiones $n \times 1$, compuesta por las coordenadas del vector y. Entonces, las correlaciones (60.2) serán equivalentes a una igualdad matricial

$$A_{qe}x_e = y_q. \tag{61.2}$$

Se llama igualdad *coordenada*, correspondiente a la igualdad *operacional*

$$Ax = y.$$

Dicha igualdad liga, en forma matricial, las coordenadas de la preimagen con las de la imagen a través de la matriz del operador.

Resulta importante observar que desde el punto de vista de la notación, las igualdades coordenada y operacional parecen ser completamente análogas, siempre que, desde luego, se omitan los índices, mientras que *el símbolo* Ax se entiende como el producto de A por x. Puesto que las notaciones y las propiedades de las operaciones sobre matrices y operadores coinciden, cualquier transformación de una igualdad operacional lleva a la misma transformación de la igualdad coordenada. Por esto, formalmente no importa de qué correlaciones se trata: de las matriciales o de las operacionales.

En adelante, de hecho, no haremos distinciones entre las igualdades operacionales y coordenadas. Más aún, *todos los conceptos y hechos nuevos referentes a los operadores, los extenderemos sin reservas a las matrices.*

Ejercicios.

1. Demuéstrese que las operaciones sobre las matrices están vinculadas a la operación de transposición mediante las siguientes correlaciones:

$$(\alpha A)' = \alpha A', \qquad (A+B)' = A' + B',$$
$$(AB)' = B'A', \qquad (A')' = A.$$

2. Demuéstrese que todo operador lineal de rango r puede representarse en forma de una suma de r operadores lineales y no puede representarse en forma de una suma de un número menor de operadores de rango 1.

3. Demuéstrese que una matriz de dimensiones $n \times m$ es de rango 1, si, y sólo si, puede representarse en forma de un producto de dos matrices no nulas de dimensiones $n \times 1$ y $1 \times m$.

4. Supongamos que para las matrices fijadas A, B se cumple la igualdad

$$AC = BC$$

para cualquier matriz C. Demuéstrese que $A = B$.

5. Hállese la forma general de una matriz cuadrada que conmutable con una matriz diagonal dada.

6. Demuéstrese que para que una matriz sea escalar, es necesario y suficiente que sea conmutable con todas las matrices cuadradas.

7. La suma de elementos diagonales de una matriz A se denomina *traza* de la matriz A y se designa con $\operatorname{tr} A$. Demuéstrese que

$$\operatorname{tr} A = \operatorname{tr} A', \qquad \operatorname{tr}(\alpha A) = \alpha \cdot \operatorname{tr} A,$$
$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B, \qquad \operatorname{tr}(BA) = \operatorname{tr}(AB).$$

8. Demuéstrese que una matriz real A es nula, si, y sólo si, $\operatorname{tr}(AA') = 0$.

§ 62. Matrices y determinantes

Las matrices desempeñan un papel esencial en la investigación de los operadores lineales. Como medio auxiliar en las investigaciones se utiliza con frecuencia un determinante. Consideraremos ahora algunas cuestiones relacionadas con las matrices y los determinantes.

Supongamos que un operador regular A actúa en el espacio X. Su rango coincide con la dimensión de X. Según se deduce de las fórmulas (60.2), esto significa que el rango del sistema de columnas de la matriz del operador coincide con su número. Esto es posible cuando, y sólo cuando el determinante de la matriz sea distinto de cero. Así pues,

Un operador que actúa en un espacio lineal será regular cuando, y sólo cuando, el determinante de su matriz sea distinto de cero.

La propiedad obtenida del operador regular sirve de base para las definiciones siguientes.

Una matriz cuadrada se llama *regular*, si su determinante es distinto de cero y se llama *degenerada*, en el caso contrario.

Naturalmente, apoyándose en las propiedades correspondientes de los operadores regulares, se puede decir, por ejemplo, que un producto de matrices regulares es, nuevamente, una matriz regular; todas las matrices regulares forman un grupo de multiplicación; cada matriz regular engendra un grupo cíclico, etc. La conexión existente con los operadores regulares permite afirmar que cualquier matriz regular A tiene, y además, una sola matriz A^{-1} tal, que

$$A^{-1}A = AA^{-1} = E. \qquad (62.1)$$

La matriz A^{-1} se llama *inversa* de la matriz A.

Empleando el concepto de determinante, podemos indicar la forma explícita de los elementos de la matriz inversa a través de los menores de la matriz A. Sirven de base para la resolución de este problema las fórmulas (40.5)—(40.9). Teniendo presente la fórmula (61.1) para un elemento del producto de dos matrices, concluimos que a las ecuaciones (62.1) les satisface la matriz

$$A^{-1}=\begin{pmatrix} \frac{A_{11}}{d} & \frac{A_{21}}{d} & \dots & \frac{A_{m1}}{d} \\ \frac{A_{12}}{d} & \frac{A_{22}}{d} & \dots & \frac{A_{m2}}{d} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1m}}{d} & \frac{A_{2m}}{d} & \dots & \frac{A_{mm}}{d} \end{pmatrix}.$$

Aquí d es el determinante de la matriz A; A_{ij}, el complemento algebraico de su elemento a_{ij}. Debido a la unicidad de la matriz inversa, ésta, puede tener sólo esta forma.

Introduzcamos unas designaciones abreviadas para los menores de una matriz arbitraria A. El menor, dispuesto en las filas i_1, i_2, ..., i_p y en las columnas j_1, j_2, ..., j_p, se designará mediante

$$A\begin{pmatrix} i_1, & i_2, & \dots, & i_p \\ j_1, & j_2, & \dots, & j_p \end{pmatrix}.$$

Consideraremos, además, que la coincidencia de algunos índices en la fila superior (inferior) de la designación del menor significa que son coincidentes las filas (columnas) del mismo menor.

TEOREMA 62.1. *(fórmula Binet—Cauchy). Supongamos que una matriz cuadrada C de orden n es igual al producto de dos matrices rectangulares A y B, cuyas dimensiones son $n \times m$ y $m \times n$, respectivamente, con la particularidad de que $m \geqslant n$.* Entonces

$$C\begin{pmatrix} 1 & 2 & \dots & n \\ 1 & 2 & \dots & n \end{pmatrix} = \sum_{1\leqslant k_1<k_2< \dots <k_n\leqslant m} A\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \dots & k_n \\ 1 & 2 & \dots & n \end{pmatrix}. \tag{62.2}$$

DEMOSTRACIÓN. Denotemos con a_{ij}, b_{ij}, c_{ij} los elementos de las matrices A, B, C. Por definición del producto de las matrices, tenemos

$$c_{ij}=\sum_{s=1}^{m} a_{is}b_{sj}.$$

Sustituyendo, los elementos de la matriz C por sus expresiones y sirviéndose de la propiedad de linealidad del determinante en relación a los vectores columna, encontramos

$$\det\begin{pmatrix} c_{11} & \dots & c_{1n} \\ c_{n1} & \dots & c_{nn} \end{pmatrix} = \det\begin{pmatrix} \sum\limits_{s_1=1}^{m} a_{1s_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{1s_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{1s_n}b_{s_nn} \\ \dots & \dots & \dots & \dots \\ \sum\limits_{s_1=1}^{m} a_{ns_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{ns_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{ns_n}b_{s_nn} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{1s_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{1s_n}b_{s_nn} \\ \dots & \dots & \dots & \dots \\ a_{ns_1}b_{s_11} & \sum\limits_{s_2=1}^{m} a_{ns_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{ns_n}b_{s_nn} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & a_{1s_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{1s_n}b_{s_nn} \\ \dots & \dots & \dots & \dots \\ a_{ns_1}b_{s_11} & a_{ns_2}b_{s_22} & \dots & \sum\limits_{s_n=1}^{m} a_{ns_n}b_{s_nn} \end{pmatrix} = \dots =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \dots \sum_{n=1}^{m} \det \begin{pmatrix} a_{1s_1}b_{s_11} & a_{1s_2}b_{s_22} & \dots & a_{1s_n}b_{s_nn} \\ \dots & \dots & \dots & \dots \\ a_{ns_1}b_{s_11} & a_{ns_2}b_{s_22} & \dots & a_{ns_n}b_{s_nn} \end{pmatrix} =$$

$$= \sum_{s_1=1}^{m} \sum_{s_2=1}^{m} \dots \sum_{s_n=1}^{m} A\begin{pmatrix} 1 & 2 & \dots & n \\ s_1 & s_2 & \dots & s_n \end{pmatrix} b_{s_11}b_{s_22}\dots b_{s_nn}. \qquad (62.3)$$

Cada uno de los índices $s_1, s_2, \dots, s_n$ no depende de los otros y puede tomar cualesquiera valores desde 1 hasta m, por lo que la expresión obtenida representa en sí la suma de m^n sumandos. En dicha suma serán nulos aquellos sumandos cuyos valores tengan aunque sea dos índices iguales, puesto que serán iguales a cero los menores correspondientes de la matriz A. Todos los demás sumandos pueden dividirse en grupos con $n!$ sumandos en cada uno de ellos, reuniendo en cada grupo todos los sumandos, cuyos valores de los índices forman una misma totalidad de números.

Designemos mediante $k_1, k_2, \dots, k_n$ la disposición de los valores de los índices $s_1, s_2, \dots, s_n$, ordenada en forma creciente. Sea

$$\varepsilon(s_1, s_2, \dots, s_n) = (-1)^N,$$

donde N es el número de transposiciones necesarias para transformar la permutación $s_1, s_2, \dots, s_n$ a la $k_1, k_2, \dots, k_n$. En este caso, dentro de los límites de un grupo de los valores de los índices $s_1, s_2, \dots, s_n$, la suma de los sumandos correspondientes de (62.3) será igual a

$$\sum \varepsilon(s_1, s_2, \dots, s_n) A\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} b_{s_11}b_{s_22}\dots b_{s_nn} =$$

$$= A\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} \sum \varepsilon(s_1, s_2, \dots, s_n) b_{s_11}b_{s_22}\dots b_{s_nn} =$$

$$= A\begin{pmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{pmatrix} B\begin{pmatrix} k_1 & k_2 & \dots & k_n \\ 1 & 2 & \dots & n \end{pmatrix}.$$

De la correlación obtenida deducimos la fórmula (62.2).

COROLARIO. *El determinante del producto de dos matrices cuadradas es igual al producto de los determinantes de los factores.*

La suma en la fórmula (62.2) constará, en el caso dado, de un sumando, por lo cual

$$C\begin{pmatrix}1 & 2 \dots n\\ 1 & 2 \dots n\end{pmatrix}=A\begin{pmatrix}1 & 2 \dots n\\ 1 & 2 \dots n\end{pmatrix}B\begin{pmatrix}1 & 2 \dots n\\ 1 & 2 \dots n\end{pmatrix}$$

o bien, lo que es igual,

$$\det C = \det A \cdot \det B.$$

COROLARIO. *Supongamos que una matriz cuadrada C de orden n es igual al producto de dos matrices rectangulares A y B cuyas dimensiones son $n \times m$ y $m \times n$, respectivamente, con la particularidad de que $m < n$. En este caso,* $\det C = 0$.

En efecto, agreguemos a las matrices A y B $n - m$ últimas columnas nulas y, correspondientemente, $n - m$ filas a cada una de las matrices. Las matrices obtenidas se convierten en las cuadradas de orden n, mientras que sus determinantes serán nulos. El producto de estas matrices nos da la matriz C. Por ello, de acuerdo con el primer corolario, $\det C = 0$.

Ejercicios.

1. Demuéstrese que para cualquier matriz regular A se verifica la igualdad $(A^{-1})' = (A')^{-1}$.
2. Demuéstrese que para cualquier matriz regular A es válida la igualdad $\det (A^{-1}) = (\det A)^{-1}$.
3. Demuéstrese que para cualquier matriz cuadrada A de orden n se verifica la igualdad $\det (\alpha A) = \alpha^n \cdot \det A$.
4. Demuéstrese que si para las matrices cuadradas A, B se verifica la igualdad $AB = E$, entonces A es regular y $B = A^{-1}$.
5. Escríbase la fórmula del tipo (62.2) para un menor arbitrario del producto de dos matrices.
6. Demuéstrese que para cualquier matriz real A todos los menores principales de las matrices $A'A$ y AA' son no negativos.
7. Demuéstrese que el rango de un producto de matrices no es superior al rango de cada uno de los factores.
8. Demuéstrese que en una operación de multiplicación por una matriz regular el rango no varía.

§ 63. Paso a la otra base

Siendo fijadas las bases en los espacios, la igualdad coordenada permite investigar totalmente la acción de un operador lineal. Evidentemente, cuanto más simple es la forma de la matriz de un operador, tanto más eficaz será la realización de dicha investigación. Generalmente las matrices de los operadores dependen de las bases y nuestra tarea inmediata consiste en aclarar esta dependencia.

Sean $e_1, e_2, \ldots, e_m$ y $f_1, f_2, \ldots, f_m$ dos bases de un mismo espacio m-dimensional X. Los vectores $f_1, f_2, \ldots, f_m$ se definen unívocamente mediante sus descomposiciones

$$\begin{aligned} f_1 &= p_{11}e_1 + p_{21}e_2 + \ldots + p_{m1}e_m, \\ f_2 &= p_{12}e_1 + p_{22}e_2 + \ldots + p_{m2}e_m, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f_m &= p_{1m}e_1 + p_{2m}e_2 + \ldots + p_{mm}e_m \end{aligned} \tag{63.1}$$

según los vectores $e_1, e_2, \ldots, e_m$. Los coeficientes p_{ij} determinan la matriz

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix},$$

la cual se llama *matriz de la transformación de coordenadas* al pasar de la base $e_1, e_2, \ldots, e_m$ a la base $f_1, f_2, \ldots, f_m$.

Tomemos un vector arbitrario $x \in X$ y descompongámoslo según los vectores de ambas bases. Sea

$$x = \sum_{i=1}^{m} \xi_i e_i = \sum_{i=1}^{m} \eta_i f_i.$$

De acuerdo con (63.1) tenemos

$$\sum_{i=1}^{m} \xi_i e_i = \sum_{i=1}^{m} \eta_i f_i = \sum_{i=1}^{m} \eta_i \sum_{j=1}^{m} p_{ji} e_j =$$
$$= \sum_{j=1}^{m} \Big(\sum_{i=1}^{m} \eta_i p_{ji} \Big) e_j = \sum_{i=1}^{m} \Big(\sum_{j=1}^{m} \eta_j p_{ij} \Big) e_i.$$

Comparando los coeficientes de e_i en el primero y segundo miembros de las correlaciones obtenidas, encontramos

$$\xi_i = \sum_{j=1}^{m} p_{ij}\eta_j \tag{63.2}$$

para $i = 1, 2, \ldots, m$. Estas fórmulas se denominan *fórmulas de transformación de las coordenadas*. Designemos, como hasta ahora, mediante x_e y x_f las matrices de dimensiones $m \times 1$, formadas por las coordenadas del vector x en las bases correspondientes. Las fórmulas (63.2) muestran que

$$x_e = Px_f. \tag{63.3}$$

La matriz de la transformación de coordenadas debe ser *regular*, puesto que en el caso contrario tendrá lugar la dependencia lineal entre sus columnas y, por tanto, entre los vectores $f_1, f_2, \ldots, f_m$. Por supuesto, cualquier matriz regular es una matriz de cierta transformación de coordenadas definida mediante la igualdad (63.3). Al multiplicar a la izquierda la igualdad (63.3) por la matriz P^{-1}, obtendremos

$$x_f = P^{-1}x_e.$$

Supongamos ahora que en el espacio lineal X vienen dadas tres bases $e_1, e_2, \ldots, e_m, f_1, \ldots, f_m$ y $r_1, \ldots, r_m$. El paso de la primera base a la tercera puede realizarse con ayuda de dos procedimientos: o bien directamente de la primera a la tercera o bien primero de la primera a la segunda, y después de la segunda a la tercera. No es difícil establecer la conexión entre las matrices correspondientes de la transformación de coordenadas. De acuerdo con (63.3), tenemos:

$$x_e = Px_f, \qquad x_f = Rx_r, \qquad x_e = Sx_r.$$

De las primeras dos correlaciones se desprende

$$x_e = Px_f = P\,(Rx_r) = (PR)\,x_r,$$

de donde proviene que

$$S = PR.$$

De este modo, cuando las coordenadas se transforman de manera consecutiva, la matriz de la transformación resultante será igual al producto de matrices de las transformaciones intermedias.

Examinemos otra vez el operador lineal A que actúa de X en Y. Elijamos en el espacio X dos bases $e_1, \ldots, e_m$ y $f_1, \ldots, f_m$, y en el espacio Y otras dos bases $q_1, \ldots, q_n$ y $t_1, \ldots, t_n$. En las primeras dos bases a un mismo operador A le corresponde la igualdad coordenada

$$y_q = A_{qe}x_e, \tag{63.4}$$

y en las otras dos bases, la igualdad

$$y_t = A_{tf}x_f. \tag{63.5}$$

En concordancia con estos pares de bases, para un mismo operador A tenemos dos matrices A_{qe} y A_{tf}.

Designemos con P la matriz de la transformación de coordenadas al pasar de la base $e_1, \ldots, e_m$ a la base $f_1, \ldots, f_m$ y con Q, la matriz de la transformación de coordenadas al pasar de $q_1, \ldots, q_n$ a $t_1, \ldots$ $\ldots, t_n$. Se tiene

$$x_e = Px_f, \qquad y_q = Qy_t. \tag{63.6}$$

Sustituyendo estas expresiones para x_e, y_q en (63.4), obtenemos

$$Qy_t = A_{qe}Px_f,$$

de donde se deduce que

$$y_t = (Q^{-1}A_{qe}P)\,x_f.$$

Al comparar la igualdad obtenida con (63.5), concluimos que

$$A_{tf} = Q^{-1}A_{qe}P. \tag{63.7}$$

Esto es precisamente la correlación buscada que liga las matrices de un mismo operador en diferentes bases.

Ejercicios.

1. Demuéstrese que al pasar a otras bases el rango de la matriz del operador no cambia.
2. Demuéstrese que el determinante de una matriz de un operador que actúa en un espacio lineal no depende de cómo se elige la base.
3. ¿Qué correspondencia puede establecerse entre unos operadores regulares, que actúan en el espacio X, y las transformaciones de coordenadas en el mismo espacio?
4. Llamemos homónimas dos bases de un espacio real, si el determinante de su matriz de la transformación de coordenadas es positiva. Demuéstrese uqe todas las bases pueden ser repartidas entre dos clases de bases homónimas.
5. Llamemos izquierda una clase de las bases homónimas y derecha, la otra. Compárense dichas clases con las descritas en el § 34.

§ 64. Matrices equivalentes y matrices semejantes

A todo operador lineal A que actúa del espacio X en el espacio Y le corresponde un conjunto de sus matrices que se define por la posibilidad de elegir diferentes bases en X e Y. La estructura de este conjunto puede ser sustancialmente diferente en dependencia de si son o no coincidentes X e Y.

Dos matrices rectangulares A y B de dimensiones iguales se denominan *equivalentes*, si existen dos matrices cuadradas regulares R y S tales que se verifique

$$B = RAS.$$

De (63.7) se deduce que dos matrices, correspondientes a un mismo operador lineal, siendo diferente la elección de las bases en X e Y, serán siempre equivalentes entre sí. No es difícil ver que la afirmación contraria es también cierta. A saber, dos matrices equivalentes corresponden siempre a un mismo operador lineal en las bases adecuadamente elegidas. De este modo, a todo operador lineal que aplica X en Y le corresponde una clase de matrices equivalentes.

TEOREMA 64.1 *Para que dos matrices rectangulares de dimensiones iguales sean equivalentes, es necesario y suficiente que tengan un mismo rango.*

DEMOSTRACIÓN. Al multiplicar una matriz cualquiera por otras matrices regulares, su rango no varía, razón por la cual las matrices equivalentes tienen rangos iguales. Supongamos ahora que dos matrices de dimensiones iguales tienen un mismo rango. Demonstraremos que estas matrices son equivalentes. Demostraren.os, además, que cada matriz de rango r es equivalente a la matriz

$$I_r = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \underbrace{0 \;\; 0 \;\; \ldots \;\; 0}_{r} & & & & 0 & \ldots & 0 \end{pmatrix}.$$

Sea dada una matriz rectangular de dimensiones $n \times m$. Define cierto operador lineal A que aplica el espacio X de base $e_1, e_2, \ldots, e_m$ en el espacio Y de base $q_1, q_2, \ldots, q_n$. Designemos mediante r el número de vectores linealmente independientes entre las imágenes de los vectores de la base $Ae_1, Ae_2, \ldots, Ae_m$. Sin perturbar la generalidad, podemos considerar que los vectores $Ae_1, Ae_2, \ldots, Ae_r$ son linealmente independientes, puesto que se puede conseguirlo numerando adecuadamente los vectores de la base. Los restantes vectores, $Ae_{r+1}, \ldots, Ae_m$, se expresan linealmente en términos de los primeros

$$Ae_k = \sum_{j=1}^{r} c_{kj} Ae_j \tag{64.1}$$

para $k = r + 1, \ldots, m$. Definamos la base nueva $f_1, f_2, \ldots, f_m$ en X de la manera siguiente:

$$f_k = \begin{cases} e_k, & k = 1, 2, \ldots, r, \\ e_k - \sum\limits_{j=1}^{r} c_{kj} e_j, & k = r+1, \ldots, m \end{cases} \tag{64.2}$$

En este caso, en virtud de (64.1), tenemos

$$Af_k = 0 \tag{64.3}$$

para $k = r + 1, \ldots, m$. Hagamos, luego,

$$Af_j = t_j \tag{64.4}$$

para $j = 1, 2, \ldots, r$. Los vectores $t_1, t_2, \ldots, t_r$ son, por hipótesis, linealmente independientes. Completémoslos con ciertos vectores $t_{r+1}, \ldots, t_n$ hasta obtener una base en Y y examinemos la matriz del operador A en las bases nuevas $f_1, \ldots, f_m$ y $t_1, \ldots, t_n$. Los coeficientes de la k-ésima columna de dicha matriz coinciden con las coordenadas del vector Af_k en la base $t_1, \ldots, t_n$. De conformidad con las correlaciones (64.3), (64.4), la matriz del operador A coincidirá con I_r.

La matriz inicial y la I_r corresponden a un mismo operador, por lo cual son equivalentes. Consecuentemente, todas las matrices de un mismo rango son equivalentes a la matriz I_r y por esta razón son equivalentes entre sí.

En el transcurso de la demostración del teorema hemos respondido a una pregunta muy importante: "*¿Cómo se deben elegir las bases en los espacios X e Y, para que la matriz del operador lineal tenga una forma más simple?*" Además, hemos mostrado la forma explícita de esta matriz más simple.

La respuesta, tan sencilla y eficaz, resultó ser posible debido a que las bases en X e Y podían escogerse *independientemente* una de la otra. Supongamos ahora que el operador A actúa en el espacio X. Por supuesto, podríamos considerar nuevamente las imágenes y preimágenes en diferentes bases, no obstante esto no parece ser natural aquí, puesto que tanto las imágenes como las preimágenes pertenecen a un mismo espacio. El empleo de diferentes bases complicaría considerablemente la investigación del modo con que actúa el operador contra los vectores del espacio X. Si la base es una, las matrices P y Q en (63.6) coinciden. Por consiguiente, a todo operador lineal que actúa en un espacio lineal le corresponde una clase de matrices relacionadas por las correlaciones

$$B = P^{-1}AP \tag{64.5}$$

para diferentes matrices regulares P. Las matrices de tal índole se llaman *semejantes* y la matriz P se denomina *matriz de la transformación de semejanza*.

La cuestión referente a las condiciones bajo las cuales dos matrices pueden ser semejantes se resuelve con unas dificultades bastante grandes y la respuesta se obtendrá más adelante. También es complicada la cuestión acerca de la forma de la matriz más simple entre todas las matrices semejantes. Los dos capítulos siguientes están dedicados precisamente a las investigaciones de estos problemas.

Ejercicios.

1. Demuéstrese que el criterio de equivalencia de las matrices y el criterio de semejanza son unas correlaciones de equivalencia.
2. Demuéstrese que las matrices semejantes poseen una traza y un determinante iguales.
3. Demuéstrese que en una misma transformación de semejanza un grupo cíclico de matrices regulares pasa en un grupo cíclico.
4. Demuéstrese que en una misma transformación de semejanza un subespacio lineal de matrices pasa en un subespacio lineal.
5. En un conjunto de matrices cuadradas del mismo orden examinemos un operador consistente en la transformación de semejanza de dichas matrices con una matriz fijada de la transformación de semejanza. Demuéstrese que este operador es lineal.
6. Demuéstrese que el conjunto de todos los operadores de una transformación de semejanza definidos sobre un mismo conjunto de matrices cuadradas del mismo orden forma un grupo de multiplicación.

CAPÍTULO 8 POLINOMIO CARACTERÍSTICO

§ 65. Valores propios y vectores propios

Supongamos que un operador lineal A actúa en el espacio X. Esto significa que a todo vector $x \in X$ se le pone en correspondencia un vector $y = Ax$ del mismo espacio X. Puede ocurrir que para cierto vector no nulo x la imagen y la preimagen son colineales. Según veremos en lo sucesivo, una situación semejante permite simplificar considerablemente la investigación del operador.

El número λ se denomina *valor propio* y el vector no nulo x, *vector propio* del operador lineal A, siempre que estén ligados entre sí mediante la correlación $Ax = \lambda x$.

Hemos de notar que si x es un vector propio correspondiente al valor propio λ, todo vector colineal αx será también un vector propio para $\alpha \neq 0$. Si al valor propio λ le corresponden dos vectores propios x, y, entonces todo vector no nulo del tipo $\alpha x + \beta y$ también será un vector propio. Por definición, el vector nulo no es propio. Por esto el conjunto X_λ de *todos* los vectores propios que son combinaciones lineales de cualquier número de vectores propios dados, correspondientes a un mismo valor propio λ, no será un subespacio. En el caso de ampliar X_λ, al agregarle un vector nulo, X_λ se convertirá en un subespacio. Este último se denominará subespacio *propio* del operador A correspondiente al valor propio λ.

No es difícil comprender que todos los vectores no nulos del espacio X serán vectores propios de los operadores 0, E y αE. Cada uno de estos operadores tiene sólo un valor propio que es igual a 0, 1 y α, respectivamente, y, consecuentemente, por lo menos un subespacio propio que coincide con todo el espacio X. El operador proyector P cuenta con dos totalidades de vectores propios: todos los vectores pertenecientes al campo de valores del operador P y todos los vectores pertenecientes al campo de valores del operador $E - P$. A la primera totalidad de vectores propios le corresponde el valor propio $\lambda = 1$; a la segunda, el valor propio $\lambda = 0$. En efecto, como $P^2 = P$, tenemos

$$P\,(Px) = P^2x = Px = 1\cdot Px,$$
$$P\,((E - P)\,x) = (P - P^2)\,x = (P - P)\,x = 0 = 0\cdot(E - P)\,x.$$

Por consiguiente, el operador proyector tiene al menos dos subespacios propios.

TEOREMA 65.1. *Un sistema de vectores propios $x_1, x_2, \ldots, x_m$ del operador A, que corresponden a los valores propios distintos dos a dos $\lambda_1, \lambda_2, \ldots, \lambda_m$, es linealmente independiente.*

DEMOSTRACION. Los vectores propios son no nulos por definición, razón por la cual el teorema es justo, a ciencia cierta, para $m = 1$. Supongamos que es válido para cualquier sistema de $m - 1$ vectores propios y no es válido para los vectores $x_1, x_2, \ldots, x_m$. En este caso el sistema de dichos vectores será linealmente independiente, es decir, para ciertos números $\alpha_1, \alpha_2, \ldots, \alpha_m$, no iguales a cero simultáneamente, se verifica la igualdad

$$\alpha_1x_1 + \alpha_2x_2 + \ldots + \alpha_mx_m = 0. \tag{65.1}$$

Supongamos que $\alpha_1 \neq 0$. Aplicando A a (65.1), obtendremos

$$\alpha_1\lambda_1x_1 + \alpha_2\lambda_2x_2 + \ldots + \alpha_m\lambda_mx_m = 0. \tag{65.2}$$

Al multiplicar (65.1) por λ_m y al restarla de (65.2), hallamos

$$\alpha_1\,(\lambda_1 - \lambda_m)\,x_1 + \alpha_2\,(\lambda_2 - \lambda_m)\,x_2 + \ldots$$
$$\ldots + \alpha_{m-1}\,(\lambda_{m-1} - \lambda_m)\,x_{m-1} = 0.$$

Conforme a la suposición inductiva, de aquí se deduce que todos lo coeficientes de los vectores $x_1, x_2, \ldots, x_{m-1}$ son nulos. En particus lar, $\alpha_1\,(\lambda_1 - \lambda_m) = 0$, lo que contradice la condición $\lambda_1 \neq \lambda_m$ y la suposición $\alpha_1 \neq 0$. Por consiguiente, el sistema de vectores $x_1, x_2, \ldots, x_m$ es linealmente independiente.

COROLARIO. *Todo operador lineal que actúa en un espacio m-dimensional no puede tener más de m valores propios distintos dos a dos.*

Es de mayor interés el caso en que el operador A en un espacio m-dimensional tiene m valores propios distintos dos a dos. De acuerdo con el teorema 65.1, en este caso podemos escoger una base del espacio compuesta íntegramente de los vectores propios del operador A.

El operador A que actúa en el espacio m-dimensional X se llama *operador de estructura simple*, si tiene m vectores propios linealmente independientes.

El hecho de que entre todos los operadores lineales destacamos los de estructura simple se explica de una manera sencilla. Estos operadores, y sólo ellos, tienen en cierta base las matrices *diagonales*. Efectivamente, sean $x_1, x_2, \ldots, x_m$ los vectores propios del operador A linealmente independientes. Al tomarlos en calidad de los vectores básicos del espacio X, construyamos en la base citada la matriz del operador A. Tenemos

$$Ax_1 = \lambda_1 x_1,$$
$$Ax_2 = \lambda_2 x_2,$$
$$\cdot \cdot \cdot \cdot \cdot \cdot$$
$$Ax_m = \lambda_m x_m.$$

Recordemos que los elementos de las columnas de la matriz del operador coinciden con las coordenadas de las imágenes de los vectores de la base. Por eso, la matriz A_λ del operador A tendrá en la base de los vectores propios la forma siguiente:

$$A_\lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \lambda_m \end{pmatrix}.$$

Ahora, si el operador A tiene en cierta base $x_1, x_2, \dots, x_m$ una matriz diagonal con ciertos números $\lambda_1, \lambda_2, \dots, \lambda_m$ (*no forzosamente distintos*) en la diagonal principal, entonces $x_1, x_2, \dots, x_m$ serán vectores propios del operador A correspondientes a los valores propios $\lambda_1, \lambda_2, \dots, \lambda_m$.

De este modo, los operadores de estructura simple, y sólo ellos, tienen en cierta base unas matrices diagonales. Esta base puede ser compuesta *solamente* de los vectores propios del operador A. La acción de cualquier operador de estructura simple se reduce siempre al "alargamiento" de las coordenadas del vector en la base dada. Si todos los operadores lineales tuvieran estructura simple, el problema referente a la elección de la base en la que la matriz del operador tenga una forma más sencilla sería resuelto por completo. No obstante, con los operadores de estructura simple no se agotan todos los operadores lineales.

Ejercicios.

1. Supongamos que el operador A tiene un vector propio x correspondiente al valor propio λ. Demuéstrese que para el operador

$$\alpha_0 E + \alpha_1 A + \dots + \alpha_n A^n,$$

donde $\alpha_0, \alpha_1, \dots, \alpha_n$ son ciertos números, el vector x será también propio, correspondiente, esta vez, al valor propio $\alpha_0 + \alpha_1\lambda + \dots + \alpha_n\lambda^n$.

2. Demuéstrese que los operadores A y $A - \alpha E$ tienen los mismos vectores propios, cualesquiera que sean el operador A y el número α.

3. Demuéstrese que el operador A es regular cuando, y sólo cuando, no tiene valores propios nulos.

4. Demuéstrese que los operadores A y A^{-1} tienen los mismos vectores propios, cualquiera que sea el operador regular A. ¿De qué modo están relacionados entre sí los valores propios de estos operadores?

5. Demuéstrese que si el operador A es de estructura simple, el operador

$$\alpha_0 E + \alpha_1 A + \ldots + \alpha_n A^n$$

es también de estructura simple.

6. Demuéstrese que un operador de diferenciación que actúa en un espacio de polinomios no es operador de estructura simple. Hállense los vectores propios y los valores propios de este operador.

7. Examínemos el operador de una transformación de semejanza con la matriz diagonal. Demuéstrese que dicho operador es de estructura simple. Hállense todos sus vectores propios y valores propios.

§ 66. Polinomio característico

No todo operador lineal tiene aunque sea un solo vector propio. Supongamos, por ejemplo, que un operador actúa en el espacio V_2 y realiza el giro de cada segmento dirigido alrededor del origen de coordenadas a 90° en el sentido contrahorario. Es evidente que en este caso la imagen y la preimagen nunca serán colineales y el operador no tendrá ni un solo vector propio. Para investigar la cuestión de existencia de los vectores propios, deduzcamos primero una ecuación la que satisfacen todos los valores propios del operador lineal.

Supongamos que el operador lineal A actúa en el espacio m-dimensional X prefijado sobre el campo P. Si el operador dispone de valor propio λ, correspondiente al vector propio x, entonces, por definición, queda cumplida la correlación $Ax = \lambda x$, o bien, lo que es igual,

$$(\lambda E - A)\, x = 0. \qquad (66.1)$$

El vector x es no nulo, por lo cual de (66.1) se deduce que el operador $\lambda E - A$ es degenerado. De este modo, los valores propios del operador A son aquellos números λ de P, y sólo ellos, para los cuales el operador $\lambda E - A$ es degenerado.

Fijemos en el espacio X cierta base $e_1, e_2, \ldots, e_m$ y designemos mediante A_e la matriz del operador A en dicha base. El operador $\lambda E - A$ es degenerado cuando, y sólo cuando, sea degenerada su matriz $\lambda E - A_e$, es decir, cuando

$$\det (\lambda E - A_e) = 0. \qquad (66.2)$$

La determinación de valores propios no ha sido ligada con la elección de la base en el espacio X. Por esto los números λ del campo P, que satisfacen la ecuación (66.2), tampoco deben depender de la base. En realidad, *no depende de la elección de la base* el primer miembro de (66.2), cualquiera que sea λ, aunque formalmente esta dependencia se ha observado. Supongamos que en cierta otra base $f_1, f_2, \ldots, f_m$ el operador A tiene la matriz A_f. Conforme a (64.5), las matrices A_e y A_f están ligadas entre sí mediante la correlación

$$A_f = Q^{-1} A_e Q$$

donde Q es una matriz regular. Ahora, con cualquier λ de P encontramos

$$\det(\lambda E - A_f) = \det(\lambda Q^{-1}EQ - Q^{-1}A_e Q) = \det(Q^{-1}(\lambda E - A_e Q) =$$
$$= \det Q^{-1} \det(\lambda E - A_e) \det Q = (\det Q)^{-1} \det(\lambda E - A_e) \det Q =$$
$$= \det(\lambda E - A_e).$$

Al tomar en consideración la expresión del determinante de la matriz en términos de los elementos de ésta, es fácil entender que el primer miembro de (66.2) puede ser representado en la forma:

$$\det(\lambda E - A_e) = a_0 + a_1\lambda + \ldots + a_{m-1}\lambda^{m-1} + a_m\lambda^m. \qquad (66.3)$$

Los coeficientes $a_0, \ldots, a_m$ se calculan de tal o cual manera según los elementos de la matriz A_e y no dependen de λ. La potencia máxima de λ sólo figura en el producto de los elementos diagonales de la matriz $\lambda E - A_e$ y, por ello,

$$a_m = 1.$$

He aquí la expresión explícita para dos coeficientes más. A saber,

$$a_0 = (-1)^m \det A_e, \qquad a_{m-1} = -\operatorname{tr} A_e.$$

Se puede suponer, en general, que al representar el determinante $\det(\lambda E - A_e)$ en potencias de λ, empleando para ello diferentes métodos, obtendremos las expresiones de un tipo análogo al segundo miembro de (66.3), mas con distintos coeficientes a_i. Sin embargo, en lo que sigue se mostrará que la suposición citada no tiene lugar. Los coeficientes en el segundo miembro de (66.3) no dependen de cómo se realiza su cálculo. Teniendo en cuenta que el determinante $\det(\lambda E - A_e)$ no depende de la base, llegamos a que todos los coeficientes $a_0, \ldots, a_{m-1}$ son, en realidad, las características del operador A. La función

$$f(\lambda) = a_0 + a_1\lambda + \ldots + a_{m-1}\lambda^{m-1} + \lambda^m \qquad (66.4)$$

se denomina *polinomio característico del operador* A.

A todo operador lineal se le atribuye un polinomio característico. Lo recíproco es también cierto. Todo polinomio del tipo (66.4) es característico para cierto operador lineal. A título del último puede servir, por ejemplo, un operador cuya matriz A_e tiene en alguna base la forma siguiente:

$$A_e = \begin{pmatrix} -a_{m-1} & -a_{m-2} & \ldots & -a_1 & -a_0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \qquad (66.5)$$

Es fácil convencerse de esto por comprobación inmediata, recurriendo al teorema de Laplace para calcular el determinante det $(\lambda E - A_f)$. Las matrices del tipo (66.5) se llaman *matrices de Frobenius*.

Para que el número λ del campo P sea el valor propio del operador A, es necesario y suficiente que satisfaga la ecuación

$$a_0 + a_1\lambda + \ldots + a_{m-1}\lambda^{m-1} + \lambda^m = 0,$$

es decir, que sea una raíz del polinomio característico. No en cualquier campo P ni mucho menos, todo polínomio con los coeficientes de P tiene aunque sea una sola raíz de P. Como ejemplo puede indicarse el polinomio $\lambda^2 + 1$, que no tiene raíces ni en el campo de números racionales ni en el campo de números reales.

El campo P se llama *algebraicamente cerrado*, si todo polinomio con los coeficientes de P tiene al menos una raíz de P.

Así pues, si un operador lineal actúa en un espacio dado sobre un campo algebraicamente cerrado, tiene sin falta por lo menos un solo vector propio. Pueden construirse los diversos ejemplos de los campos algebraicamente cerrados, no obstante, el mayor valor práctico lo tiene sólo uno de ellos, a saber, el campo de números complejos. Nuestras investigaciones más próximas están dedicadas a la demostración de que dicho campo es algebraicamente cerrado.

Ejercicios.

1. Hállese el polinomio característico para los operadores nulo e idéntico.
2. Hállese el polinomio característico para el operador de diferenciación.
3. ¿Será señal de igualdad de operadores la coincidencia de los polinomios característicos?
4. Demuéstrese que los operadores con las matrices A y A' tienen polinomios característicos iguales.
5. Supongamos que en cierta base el operador tiene la matriz (66.5). Hállense en la misma base las coordenadas de los vectores propios.
6. Demuéstrese que un operador con la matriz (66.5) tiene estructura simple cuando, y sólo cuando, el polinomio característico cuenta con m raíces distintas dos a dos.

§ 67. Anillo de polinomios

En algunos ejercicios y ejemplos ya hemos atraído la atención del lector a las propiedades algebraicas de los polinomios. En relación con el estudio del polinomio característico estas investigaciones serán continuadas.

Sea dado un campo arbitrario P. Consideraremos un conjunto de polinomios, es decir, de funciones del tipo

$$f(z) = a_0 + a_1z + \ldots + a_nz^n, \tag{67.1}$$

dependientes del argumento z, que toma los valores de P, y que tienen los coeficientes $a_0, \ldots, a_n$ de P. Convengamos en que $f(z)$ es un polinomio de *grado* n, siempre que $a_n \neq 0$ y los coeficientes de números mayores sean nulos. El único polinomio que no tiene grado determinado es aquel cuyos coeficientes son todos iguales a cero. Llamémoslo polinomio *nulo* y lo designaremos con el símbolo 0.

Dos polinomios se considerarán *iguales*, si son iguales todos sus coeficientes de los argumentos de potencias iguales.

Sean dados, ahora, los polinomios $f(z)$ y $g(z)$ de grados n y s respectivamente. Denotemos

$$\begin{aligned} f(z) &= a_0 + a_1 z + \ldots a_{n-1} z^{n-1} + a_n z^n, \\ g(z) &= b_0 + b_1 z + \ldots + b_{s-1} z^{s-1} + b_s z^s \end{aligned} \tag{67.2}$$

y supongamos, para concretar, que $n \geqslant s$. Se denomina suma $f(z) + g(z)$ de los polinomios $f(z)$ y $g(z)$ el polinomio

$$f(z) + g(z) = c_0 + c_1 z + \ldots + c_{n-1} z^{n-1} + c_n z^n,$$

donde $c_i = a_i + b_i$ para todo $i \leqslant s$, y $c_i = a_i$ para todo $i > s$. La potencia de la suma de los polinomios es igual a n, si $n > s$, pero será inferior a n, para $n = s$, siempre que $b_n = -a_n$.

Se llama *producto* $f(z) \cdot g(z)$ *de los polinomios* $f(z)$ *y* $g(z)$ el polinomio

$$f(z) \cdot g(z) = d_0 + d_1 z + \ldots + d_{n+s-1} z^{n+s-1} + d_{n+s} z^{n+s},$$

donde

$$d_t = \sum_{k+l=t} a_k b_l$$

para $t = 0, 1, \ldots, n+s$. El coeficiente d_t es la suma de productos de aquellos coeficientes de los polinomios $f(z)$ y $g(z)$ cuyos índices, siendo sumados, dan t. Por ejemplo,

$$d_0 = a_0 b_0, \qquad d_{n+s} = a_n b_s.$$

De la última igualdad se desprende que $d_{n+s} \neq 0$, por lo cual la potencia del producto de polinomios no nulos es igual a la suma de potencias de los factores. Por consiguiente, el producto de los polinomios no nulos es un polinomio no nulo.

Como caso particular del producto de los polinomios interviene *el producto* $\alpha f(z)$ *del polinomio* $f(z)$ *por el número* α, puesto que un número no nulo puede considerarse como un polinomio de grado nulo.

El conjunto de polinomios con las operaciones introducidas más arriba representa en sí *un anillo conmutativo*. No nos detendremos en la comprobación de la validez de todos los axiomas.

TEOREMA 67.1. *Para cualquier polinomio* $f(z)$ *y el polinomio no nulo* $g(z)$ pueden hallarse los únicos polinomios $q(z)$ y $r(z)$ tales que se verifica la ecuación

$$f(z) = g(z)\, q(z) + r(z), \tag{67.3}$$

con la particularidad de que el grado de $r(z)$ es inferior al de $g(z)$ o bien $r(z) = 0$.

DEMOSTRACIÓN. Supongamos que los polinomios $f(z)$ y $g(z)$ son de grado n y s. Si $n < s$ o bien $f(z) = 0$, entonces en la descomposición (67.3) podemos hacer $q(z) = 0$, $r(z) = f(z)$. Supongamos, por esto, que $n \geqslant s$.

Representemos los polinomios $f(z)$ y $g(z)$ conforme a (67.2) y hagamos

$$f(z) - \frac{a_n}{b_s} z^{n-s} g(z) = f_1(z). \tag{67.4}$$

Supongamos que el grado del polinomio $f_1(z)$ es igual a n_1, y su coeficiente mayor es $a_{n_1}^{(1)}$. Es obvio que $n_1 < n$. Si $n_1 \geqslant s$, hagamos

$$f_1(z) - \frac{a_{n_1}^{(1)}}{b_s} z^{n_1-s} g(z) = f_2(z). \tag{67.5}$$

Designaremos mediante n_2 el grado y mediante $a_{n_2}^{(2)}$, el coeficiente mayor del polinomio $f_2(z)$. Si $n_2 \geqslant s$, hagamos otra vez

$$f_2(z) - \frac{a_{n_2}^{(2)}}{b_s} z^{n_2-s} g(z) = f_3(z), \tag{67.6}$$

etc.

Los grados de los polinomios $f_1(z)$, $f_2(z)$, ... van decreciendo. Por ello, realizados un número finito de pasos, llegaremos a la igualdad

$$f_{k-1}(z) - \frac{a_{n_{k-1}}^{(k-1)}}{b_s} z^{n_{k-1}-s} g(z) = f_k(z), \tag{67.7}$$

en la cual el polinomio $f_k(z)$ o bien es nulo o bien su grado n_k es inferior a s. En este momento el proceso se para.

Sumando ahora todas las igualdades del tipo (67.4) — (67.7), obtendremos

$$f(z) - \left(\frac{a_n}{b_s} z^{n-s} + \frac{a_{n_1}^{(1)}}{b_s} z^{n_1-s} + \ldots + \frac{a_{n_{k-1}}^{(k-1)}}{b_s} z^{n_{k-1}-s}\right) g(z) = f_k(z).$$

Esto es testimonio de que los polinomios

$$q(z) = \frac{a_n}{b_s} z^{-s} + \frac{a_n^{(1)}}{b_s} z^{n_1-s} + \ldots + \frac{a_{n_{k-1}}^{(k-1)}}{b_s} z^{n_{k-1}-s}, \quad r(z) = f_k(z)$$

satisfacen la igualdad (67.3), con la particularidad de que o bien $r(z) = 0$, o bien el grado del polinomio $r(z)$ es inferior al grado de $g(z)$.

Demostraremos ahora que los polinomios $q(z)$ y $r(z)$, que satisfacen la condición del teorema, son únicos. Supongamos que existen además los polinomios $q'(z)$ y $r'(z)$, para los cuales

$$f(z) = g(z)\, q'(z) + r'(z),$$

con la particularidad de que o bien $r'(z) = 0$, o bien el grado de $r'(z)$ es inferior al grado de $g(z)$. Entonces

$$g(z)(q(z) - q'(z)) = r'(z) - r(z). \qquad (67.8)$$

El polinomio en el segundo miembro de esta igualdad o bien es nulo o bien su grado es inferior al grado de $g(z)$. Mientras tanto, el polinomio en el primer miembro tiene, cuando $q(z) - q'(z) \neq 0$, un grado que no es inferior al grado de $g(z)$. Por esta razón, la igualdad (67.8) se verifica sólo en el caso en que

$$q(z) = q'(z), \qquad r(z) = r'(z).$$

El teorema queda demostrado por completo.

El polinomio $q(z)$ se denomina *cociente* de la división de $f(z)$ por $g(z)$, y $r(z)$ es *el resto* de la división. Si el resto es igual a cero, diremos que $f(z)$ se divide por $g(z)$ y el propio polinomio $g(z)$ se llamará *divisor* del polinomio $f(z)$.

Consideraremos la división de un polinomio arbitrario no nulo $f(z)$ por un polinomio de primer grado $z - a$. Tenemos

$$f(z) = (z - a)\, q(z) + r(z). \qquad (67.9)$$

Puesto que el grado de $r(z)$ debe ser inferior al grado del polinomio $z - a$, entonces $r(z)$ es un polinomio de grado nulo, es decir, una constante. Esta constante se puede determinar con facilidad. Sustituyamos en el primero y segundo miembros de la correlación (67.9) $z = a$ y encontramos que $r(z) = f(a)$. Así,

$$f(z) = (z - a)\, q(z) + f(a). \qquad (67.10)$$

Para que el polinomio $f(z)$ se divida por el polinomio $z - a$, es necesario y suficiente que $f(a) = 0$. Los números a, para los cuales $f(a) = 0$, suelen llamarse *raíces* del polinomio $f(z)$. De este modo, la búsqueda de todos los divisores lineales del polinomio es equivalente a la búsqueda de todas sus raíces.

El empleo de la fórmula (67.10) permite hacer la siguiente deducción. Para todo número a de P un polinomio $f(z)$ de grado n puede ser representado de un modo único en forma de la descomposición en potencias $(z - a)$:

$$f(z) = A_0 + A_1 (z - a) + \ldots + A_{n-1} (z - a)^{n-1} + \\ + A_n (z - a)^n, \qquad (67.11)$$

donde $A_0, \ldots, A_n$ son los números de P.

La existencia de aunque sea una sola descomposición (67.11) se establece de un modo relativamente sencillo. Al dividir $f(z)$ por $(z - a)$, obtendremos el cociente $q_1(z)$ y el resto A_0, relacionados mediante la igualdad

$$f(z) = (z - a)\, q_1(z) + A_0. \qquad (67.12)$$

Si el grado de $q_1(z)$ es nulo, la descomposición (67.11) queda obtenida. En cambio, si el grado de $q_1(z)$ es distinto de cero, entonces al dividir $q_1(z)$ por $(z-a)$, tendremos

$$q_1(z) = (z-a)\,q_2(z) + A_1. \tag{67.13}$$

Reuniendo (67.12), (67.13), encontramos

$$f(z) = (z-a)^2\,q_2(z) + A_1(z-a) + A_0.$$

Cuando sea necesario, dividimos otra vez $q_2(z)$ por $(z-a)$, etc. Como los grados de los cocientes $q_1(z), q_2(z), \ldots$ decrecen de manera sucesiva, el proceso se parará dentro de n pasos, dando por resultado la descomposición (67.11).

Supongamos ahora que la descomposición del mismo tipo se ha obtenido de algún otro modo y tiene los coeficientes $A'_0, \ldots, A'_n$. Al designar

$$q'_i(z) = A'_i + A'_{i+1}(z-a) + \ldots + A'_n(z-a)^{n-i}$$

para $i = 0, 1, \ldots, n$. Concluimos que

$$q'_i(z) = (z-a)\,q'_{i+1}(z) + A'_i. \tag{67.14}$$

En este caso, naturalmente, $q'_0(z) = f(z)$. Comprando (67.12) y (67.14) para $i = 0$, y tomando en consideración la unicidad del cociente y del resto, llegamos a la conclusión de que $A_0 = A'_0$, $q_1(z) = q'_1(z)$. De modo análogo se demuestra que los otros coeficientes son también iguales.

Ejercicios.

1. Demuéstrese que en el anillo de polinomios no hay divisores de cero.
2. Supongamos que para ciertos polinomios se verifica la igualdad $f(z)\,\varphi(z) = g(z)\,\varphi(z)$. Demuéstrese que si $\varphi(z) \neq 0$, entonces $f(z) = g(z)$.
3. Demuéstrese que los polinomios no nulos $f(z)$ y $g(z)$ se dividen uno por otro cuando, y sólo cuando, $g(z) = \alpha f(z)$ para el número α no nulo.
4. Supongamos que cada uno de los polinomios $f_1(z), \ldots, f_k(z)$ se divide por $\varphi(z)$. Demuéstrese que el polinomio $f_1(z)\,g_1(z) + \ldots + f_k(z)\,g_k(z)$, donde $g_1(z), \ldots, g_k(z)$ son unos polinomios arbitrarios, también se divide por $\varphi(z)$.
5. Demuéstrese que en las descomposiciones (67.1), (67.11) para un mismo polinomio $f(z)$ los coeficientes a_n y A_n coinciden.

§ 68. Teorema fundamental del álgebra

Procedamos a demostrar una de las más importantes afirmaciones, es decir, el teorema acerca de que el campo de números complejos es algebraicamente cerrado. Este teorema es de amplio uso en las más diversas ramas de las matemáticas. En particular, en este teorema está basada toda la teoría ulterior de los operadores lineales. Conforme a la tradición establecida, se llamará *teorema fundamental del álgebra*.

Así pues, hemos de mostrar que todo polinomio de grado $n \geqslant 1$ con coeficientes complejos tiene al menos una raíz que es, en el caso general, compleja. Consideraremos al principio unos polinomios del tipo especial. A saber,

$$f(z) = a - z^n. \tag{68.1}$$

Representemos los números complejos z en la así llamada *forma triangular*

$$z = r(\cos\varphi + i \operatorname{sen}\varphi).$$

Aquí, r es un número no negativo, llamado *módulo* del número z, y φ es un número real llamado *argumento* del número z. Está claro que para todo número z el módulo está definido unívocamente. Para los números z no nulos el módulo está definido, salvo un número múltiple de 2π; para $z = 0$ el argumento no está definido. Formando el producto de dos números complejos

$$z = r(\cos\varphi + i \operatorname{sen}\varphi), \qquad v = \rho(\cos\psi + i \operatorname{sen}\psi),$$

encontramos

$$zv = r\rho(\cos\varphi + i \operatorname{sen}\varphi)(\cos\psi + i \operatorname{sen}\psi) =$$
$$= r\rho(\cos(\varphi + \psi) + i \operatorname{sen}(\varphi + \psi)).$$

De aquí deducimos que

$$z^n = r^n(\cos n\varphi + i \operatorname{sen} n\varphi).$$

Esta igualdad lleva el nombre *de Moivre*. Permite hallar con facilidad las raíces de la ecuación (68.1). En efecto, supongamos que el número complejo a está representado en la forma triangular

$$a = \alpha(\cos\theta + i \operatorname{sen}\theta).$$

La ecuación

$$a - z^n = 0$$

respecto de z es equivalente a la ecuación

$$\alpha(\cos\theta + i \operatorname{sen}\theta) = r^n(\cos n\varphi + i \operatorname{sen} n\varphi)$$

respecto de r y φ. Mas, la última ecuación tiene, a ciencia cierta, tales soluciones para $k = 0, 1, 2, \ldots, n-1$:

$$r = +\sqrt[n]{\alpha}, \qquad \varphi = \frac{\theta + 2k\pi}{n}.$$

Por consiguiente, los números complejos

$$a_k = +\sqrt[n]{\alpha}\left(\cos\frac{\theta+2k\pi}{n} + i \operatorname{sen}\frac{\theta+2k\pi}{n}\right) \tag{68.2}$$

son las raíces de la ecuación (68.1). Llamaremos estos números raíces de *n-ésimo grado del número* a y las designaremos mediante el símbolo general

$$a_k = \sqrt[n]{a}\,.$$

Ahora, sea dado un polinomio arbitrario $f(z)$ con coeficientes complejos. Consideraremoslo como una función compleja del argumento complejo z. Para tales funciones, al igual que para las funciones reales de un argumento real, pueden introducirse las nociones de continuidad, derivada, etc. De éstas no todas las nociones las necesitaremos en una medida igual, pero todas ellas se basan en el empleo de la completitud del espacio de números complejos.

Una función compleja uniforme $f(z)$ del argumento complejo z se llama *continua en el punto* z_0, si para todo número $\varepsilon > 0$, tan pequeño como se quiera, existe un $\delta > 0$ tal que para cualquier número complejo z que satisface la desigualdad

$$|z - z_0| < \delta$$

tendremos

$$|f(z) - f(z_0)| < \varepsilon.$$

La función $f(z)$, continua en todo punto del campo de definición, se denomina *continua en todo punto* o, simplemente, *continua*.

LEMA 68.1. *El polinomio $f(z)$ con coeficientes complejos es una función continua del argumento complejo z.*

DEMOSTRACION. Sea

$$f(z) = a_0 + a_1 z + \ldots + a_n z^n \qquad (68.3)$$

y supongamos que z_0 es un número complejo arbitrario fijado. Designemos $h = z - z_0$. Mostremos que para cualquier número $\varepsilon > 0$, tan pequeño como se quiera, se puede hallar tal $\delta > 0$, que, siendo $|h| < \delta$, se cumpla la desigualdad $|f(z) - f(z_0)| < \varepsilon$.

Al desarrollar el polinomio dado $f(z)$ en potencias de $(z - z_0)$, obtendremos

$$f(z) = A_0 + A_1 (z - z_0) + \ldots + A_n (z - z_0)^n.$$

Puesto que $A_0 = f(z_0)$ y $(z - z_0)$ está designado con h, resulta

$$f(z_0 + h) - f(z_0) = A_1 h + \ldots + A_n h^n. \qquad (68.4)$$

De aquí proviene que

$$|f(z_0 + h) - f(z_0)| \leqslant |A_1|\,|h| + \ldots$$
$$\ldots + |A_n|\,|h|^n = A(|h|). \qquad (68.5)$$

La función real $A(|h|)$ es un polinomio con coeficientes reales $|A_i|$ respecto de la variable real $|h|$. Según se sabe del curso del análisis matemático, $A(|h|)$ es una función continua en todo punto y, en particular, cuando $|h| = 0$. Como $A(0) = 0$, por $\varepsilon > 0$

dado se puede hallar tal $\delta > 0$ que para

$$|h| < \delta \tag{68.6}$$

tendremos

$$A(|h|) < \varepsilon.$$

Tomando en consideración la desigualdad (68.5), concluimos que si se cumple (68.6), se cumplirá también la desigualdad

$$|f(z_0 + h) - f(z_0)| < \varepsilon.$$

COROLARIO. *El módulo de un polinomio es una función continua.*

Dicha afirmación se deduce inmediatamente de la siguiente correlación:

$$||f(z)| - |f(z_0)|| \leqslant |f(z) - f(z_0)|.$$

COROLARIO. *Si una sucesión de números complejos $\{z_k\}$ converge a z_0, para todo polinomio $f(z)$*

$$\lim_{k\to\infty} f(z_k) = f(z_0).$$

LEMA 68.2. *Si el polinomio $f(z)$ de grado $n \geqslant 1$ no se anula cuando $z = z_0$, entonces siempre existe un número complejo h tal, que*

$$|f(z_0 + h)| < |f(z_0)|.$$

DEMOSTRACION. Examinemos nuevamente el desarrollo (68.4). Supongamos que entre los coeficientes $A_1, A_2, \ldots, A_n$ el primer coeficiente distinto de cero es A_k. Tomemos

$$h = t\sqrt[k]{-\frac{f(z_0)}{A_k}}, \tag{68.7}$$

donde a título de raíz de la k-ésima potencia se toma cualquiera de los valores de este coeficiente y

$$0 \leqslant t \leqslant 1. \tag{68.8}$$

Designaremos

$$B_p = A_p\left(\sqrt[k]{-\frac{f(z_0)}{A_k}}\right)^p.$$

Ahora, teniendo en cuenta (68.7), (68.8), de (68.4) encontramos

$$\begin{aligned}
|f(z_0 + h)| &= |f(z_0) - t^k f(z_0) + t^{k+1}B_{k+1} + \ldots + t^n B_n| \leqslant \\
&\leqslant |(1 - t^k) f(z_0)| + t^{k+1}|B_{k+1}| + \ldots + t^n|B_n| = \\
&= (1 - t^k)|f(z_0)| + t^{k+1}|B_{k+1}| + \ldots + t^n|B_n| = \\
&= |f(z_0)| + t^k(-|f(z_0)| + t|B_{k+1}| + \ldots \\
&\qquad \ldots + t^{n-k}|B_n|) = |f(z_0)| + t^k B(t).
\end{aligned}$$

En definitiva tenemos

$$|f(z_0 + h)| \leqslant |f(z_0)| + t^k B(t).$$

La función $B(t)$ es un polinomio con coeficientes reales y el argumento real t. Es una función continua. Pero, $B(0) = -|f(z_0)| < 0$, razón por la cual, en virtud de la continuidad de $B(t)$, existe tal t_0 dentro de los límites $0 < t_0 \leqslant 1$, que $B(t_0)$ será también negativo. Para el número complejo h, que se determina mediante el número t_0, de acuerdo con (68.7) obtendremos

$$|f(z_0+h)| \leqslant |f(z_0)| + t_0^k B(t_0) < |f(z_0)|.$$

LEMA 68.3 *Para todo polinomio $f(z)$ de grado $n \geqslant 1$ y toda sucesión infinita de números complejos $\{z_k\}$ se verifica la correlación límite*

$$\lim_{k\to\infty} |f(z_k)| = +\infty. \qquad (68.9)$$

DEMOSTRACIÓN. Consideraremos el polinomio (68.3). Para cualquier $z \neq 0$ encontramos

$$|f(z)| \geqslant |a_n||z|^n \left(1 - \left|\frac{a_0}{a_n}\right| |z|^{-n} - \ldots - \left|\frac{a_{n-1}}{a_n}\right| |z|^{-1}\right). \qquad (68.10)$$

Puesto que la sucesión $\{z_k\}$ es infinitamente creciente, entonces

$$\lim_{k\to\infty} |z_k| = +\infty.$$

El segundo miembro de la correlación (68.10) es una función real, por lo cual calculamos

$$\lim_{|z_k|\to\infty} \left(1 - \left|\frac{a_0}{a_n}\right| |z_k|^{-n} - \ldots - \left|\frac{a_{n-1}}{a_n}\right| |z_k|^{-1}\right) = 1.$$

Pero, para el otro factor de (68.10) tenemos

$$\lim_{|z_k|\to\infty} |a_n||z_k|^n = +\infty.$$

Por consiguiente, la correlación (68.9) es válida.

TEOREMA 68.1 (*teorema fundamental del álgebra*). *Todo polinomio $f(z)$ de grado $n \geqslant 1$ con coeficientes complejos tiene por lo menos una raíz, que es, en el caso general, compleja.*

DEMOSTRACIÓN. Consideraremos un conjunto de toda clase de valores del módulo del polinomio $f(z)$. Puesto que $|f(z)| \geqslant 0$, este conjunto está acotado inferiormente. Del curso del análisis matemático sabemos que cualquier conjunto no vacío de números reales, acotado inferiormente, tiene la cota inferior exacta. Supongamos que para el conjunto de valores $|f(z)|$ esta cota es l. Esto significa que para cualquier número natural k se puede hallar tal número complejo z_k que

$$0 \leqslant |f(z_k)| - l \leqslant 2^{-k}.$$

De aquí se deduce que

$$\lim_{k\to\infty} |f(z_k)| = l. \qquad (68.11)$$

Si suponemos que la sucesión $\{z_k\}$ no está acotada, se podría elegir de ella una subsucesión infinitamente grande y, de acuerdo con el lema (68.3), la correlación (68.11) no podría verificarse. Por esta razón la sucesión $\{z_k\}$ es acotada. Elijamos de ésta una subsucesión convergente $\{z_{k_\nu}\}$ y supongamos que

$$\lim_{k_\nu \to \infty} z_{k_\nu} = z_0.$$

Conforme al corolario del lema 68.1, el módulo de un polinomio es una función continua. Por consiguiente,

$$|f(z_0)| = \lim_{k_\nu \to \infty} |f(z_{k_\nu})| = l.$$

Si $l \neq 0$, del lema 68.2 se desprende que existe tal número z_0, para el cual $|f(z_0')| < l$. Esto contradice a que l es la cota inferior exacta de los valores del módulo del polinomio, razón por la cual $l = 0$.

Así pues, hemos probado la existencia de un número complezo z_0 tal, que $|f(z_0)| = 0$, ó, lo que es igual,

$$f(z_0) = 0.$$

Esto significa que z_0 es la raíz del polinomio $f(z)$.

Ejercicios.

1. Demuéstrese que el conjunto de todas las raíces de n-ésimo grado del número complejo 1 forma un grupo conmutativo de multiplicación.

2. Demuéstrese que para que una sucesión de números complejos $\{z_k\}$ sea acotada, es necesario y suficiente que la sucesión $\{f(z_k)\}$ sea acotada por lo menos para un polinomio $f(z)$ de grado $n \geqslant 1$.

3. Demuéstrese que para cualquier polinomio $f(z)$ de grado $n \geqslant 1$ y todo número complejo z_0 existe un número complejo h tal que $|f(z_0+h)| > |f(z_0)|$.

4. Demuéstrese que todas las raíces del polinomio (68.3) se hallan dentro del anillo

$$\left(1+\max_{k>0}\left|\frac{a_k}{a_0}\right|\right)^{-1} \leqslant |z| \leqslant \left(1+\max_{k<n}\left|\frac{a_k}{a_n}\right|\right).$$

5. Rigiéndose por el mismo esquema que se ha usado para los números complejos, trátese de «demostrar» que el campo de números reales es algebraicamente cerrado. ¿En qué lugar la «demostración» no tiene analogía?

§ 69. Corolarios del teorema fundamental

Del teorema fundamental se deduce toda una serie de corolarios. Consideraremos los más importantes de ellos.

El polinomio $f(z)$ de grado $n \geqslant 1$ con coeficientes complejos tiene por lo menos una raíz z_1. Por esto, $f(z)$ puede ser descompuesto así:

$$f(z) = (z - z_1)\,\varphi(z),$$

donde $\varphi(z)$ es un polinomio de grado $n - 1$. Los coeficientes del polinomio $\varphi(z)$ son, como antes, números complejos. Por consiguiente, $\varphi(z)$ tiene la raíz z_2 (siempre que $n \geqslant 2$) y

$$\varphi(z) = (z - z_2)\,\psi(z),$$

de donde se deduce que

$$f(z) = (z - z_1)(z - z_2)\,\psi(z).$$

Continuando este proceso, obtendremos la descomposición del polinomio en un producto de factores lineales:

$$f(z) = b\,(z - z_1)(z - z_2)\ldots(z - z_n),$$

donde b es un cierto número. Abriendo los paréntesis en el segundo miembro de la descomposición obtenida y comparando los coeficientes de las potencias con los coeficientes a_i del polinomio $f(z)$, llegamos a la conclusión de que $b = a_n$.

Entre los números $z_1, z_2, \ldots, z_n$ pueden haber números iguales. Supongamos, para simplificar, que $z_1, \ldots, z_r$ son distintos dos a dos, mientras que cada uno de los números $z_{r+1}, \ldots, z_n$ es igual a uno de los primeros. En este caso el polinomio $f(z)$ puede escribirse en la forma:

$$f(z) = a_n (z - z_1)^{k_1} (z - z_2)^{k_2} \ldots (z - z_r)^{k_r}, \tag{69.1}$$

donde $z_i \neq z_j$ cuando $i \neq j$, y, además,

$$k_1 + k_2 + \ldots + k_r = n.$$

La descomposición (69.1) se denomina *descomposición canónica del polinomio $f(z)$ en factores.*

Para el polinomio $f(z)$ la descomposición canónica es única, salvo el orden en que se disponen los factores. En efecto, supongamos que a la par con la descomposición (69.1) existe otra descomposición canónica, por ejemplo

$$f(z) = a_n (z - v_1)^{l_1} (z - v_2)^{l_2} \ldots (z - v_m)^{l_m}.$$

En este caso será verídica la igualdad

$$(z - z_1)^{k_1} (z - z_2)^{k_2} \ldots (z - z_r)^{k_r} = (z - v_1)^{l_1} (z - v_2)^{l_2} \ldots (z - v_m)^{l_m}. \tag{69.2}$$

Observemos que la totalidad de los números $z_1, \ldots, z_r$ debe coincidir con la de los números $v_1, \ldots, v_m$. Si, por ejemplo, z_1 no es igual a ninguno de los números $v_1, \ldots, v_m$, entonces, al sustituir

$z = z_1$ en (69.2), obtendremos cero en el primer miembro de la igualdad, mientras que en el segundo miembro, un número diferente de cero. Así pues, si existen dos descomposiciones canónicas del polinomio $f(z)$, la igualdad (69.2) sólo puede tener esta forma:

$$(z-z_1)^{k_1}(z-z_2)^{k_2}\dots(z-z_r)^{k_r}=(z-z_1)^{l_1}(z-z_2)^{l_2}\dots(z-z_r)^{l_r}.$$

Supongamos, por ejemplo, que $k_1 \neq l_1$ y sea, para concretar $k_1 > l_1$. Al dividir ambos miembros de la última igualdad por el mismo divisor $(z-z_1)^{l_1}$, llegamos a que

$$(z-z_1)^{k_1-l_1}(z-z)^{k_2}\dots(z-z_r)^{k_r}=(z-z_2)^{l_2}\dots(z-z_r)^{l_r}.$$

Otra vez, al sustituir aquí $z = z_1$, vemos que en el primer miembro de la igualdad figura cero y en el segundo miembro, un número diferente de cero. De este modo, la unicidad de la descomposición canónica queda demostrada.

Si en la descomposición canónica (69.1) $k_i = 1$, entonces la raíz z_i se llama *simple*; en cambio si $k_i > 1$, entonces la raíz z_i se denomina *múltiple*. El número k_i lleva el nombre de *multiplicidad* de la raíz z_i. Ahora podemos enunciar una deducción muy importante:

Todo polinomio de grado $n \geqslant 1$ con coeficientes complejos tiene n raíces, si cada una de las raíces se cuenta tantas veces cual es su multiplicidad.

Un polinomio de grado nulo no tiene raíces. El único polinomio que tiene un número tan grande como se quiera de raíces distintas dos a dos es el polinomio nulo. Haciendo uso de lo citado, podemos enunciar la deducción siguiente:

Si dos polinomios $f(z)$ y $g(z)$ de grado no superior a n tienen valores iguales para más de n diferentes valores del argumento, todos los coeficientes correspondientes de dichos polinomios serán iguales entre sí.

En efecto, el polinomio $f(z) - g(z)$ tiene, por hipótesis, más de n raíces. Pero su grado no es superior a n, por lo cual $f(z) - g(z) = 0$.

Así pues, un polinomio $f(z)$, cuyo grado no es superior a n, se determina completamente mediante sus valores para cualesquiera $n + 1$ diferentes valores del argumento. Esto permite restablecer el polinomio por medio de sus valores. No es difícil señalar la forma explícita de este polinomio "restaurador". Si el polinomio $f(z)$ toma, para los valores del argumento $\alpha_1, \dots, \alpha_{n+1}$, los valores $f(\alpha_1), \dots, f(\alpha_{n+1})$, entonces

$$f(z)=\sum_{i=1}^{n+1} f(\alpha_i)\frac{(z-\alpha_1)\dots(z-\alpha_{i-1})(z-\alpha_{i+1})\dots(z-\alpha_{n+1})}{(\alpha_i-\alpha_1)\dots(\alpha_i-\alpha_{i-1})(\alpha_i-\alpha_{i+1})\dots(\alpha_i-\alpha_{n+1})}.$$

Está claro que el grado del polinomio en el segundo miembro no es superior a n, y en los puntos $z = \alpha_i$ el polinomio toma los valores $f(\alpha_i)$. Un polinomio construido de esta manera se denomina *polinomio de interpolación de Lagrange.*

Examinemos ahora un polinomio $f(z)$ de grado n y supongamos que $z_1, z_2, \ldots, z_n$ son sus raíces escritas tantas veces cual es la multiplicidad de cada una de ellas. En este caso

$$f(z) = a_n (z - z_1)(z - z_2) \ldots (z - z_n).$$

Multiplicando las expresiones encerradas entre paréntesis que figuran en el segundo miembro, reduciendo los términos semejantes y comparando los coeficientes obtenidos con los de la expresión (68.3), podemos deducir las siguientes igualdades:

$$\begin{aligned}
a_{n-1}/a_n &= -(z_1 + z_2 + \ldots + z_n),\\
a_{n-2}/a_n &= +(z_1z_2 + z_1z_3 + \ldots + z_1z_n + \ldots + z_{n-1}z_n),\\
a_{n-3}/a_n &= -(z_1z_2z_3 + z_1z_2z_4 + \ldots + z_{n-2}z_{n-1}z_n),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
a_1/a_n &= (-1)^{n-1}(z_1z_2 \ldots z_{n-1} + \ldots + z_2z_3 \ldots z_n)\\
a_0/a_n &= (-1)^n z_1z_2 \ldots z_n.
\end{aligned}$$

Estas igualdades se llaman *fórmulas de Viète* y expresan los coeficientes de un polinomio en términos de sus raíces.

En el segundo miembro de la k-ésima igualdad figura la suma de toda clase de productos de k raíces tomada con el signo más o el signo menos, en dependencia de si es par o impar el número k.

En lo sucesivo nos serán útiles algunos corolarios del teorema fundamental del álgebra referentes a los polinomios con coeficientes reales. Supongamos que el polinomio

$$f(z) = a_0 + a_1 z + \ldots + a_n z^n$$

con coeficientes reales tiene una raíz compleja v (pero no real), es decir,

$$a_0 + a_1 v + \ldots + a_n v^n = 0.$$

La última igualdad no se violará, si todos los números en ella se sustituyen por los números complejos conjugados. No obstante, los coeficientes $a_0, \ldots, a_n$ y el número 0, siendo números reales, quedarán en esta sustitución intactos. Por ello

$$a_0 + a_1\bar{v} + \ldots + a_n\bar{v}^n = 0,$$

es decir, $f(\bar{v}) = 0$.

De este modo, si un número complejo (pero no real) v es una raíz del polinomio $f(z)$ con coeficientes reales, entonces el número complejo conjugado $\bar{v}$ también será raíz de $f(z)$.

De aquí se deduce que el polinomio $f(z)$ se dividirá por un trinomio de segundo grado

$$\varphi(z) = (z - v)(z - \bar{v}) = z^2 - (v + \bar{v}) z + v\bar{v},$$

cuyos coeficientes son reales. Haciendo uso de este hecho, demostremos que las raíces v y $\bar{v}$ son de la misma multiplicidad.

Supongamos que la multiplicidad de ellas es k y l, respectivamente, y, por ejemplo, $k > l$. En este caso $f(z)$ se divide por el l-ésimo grado del polinomio $\varphi(z)$, es decir,

$$f(z) = \varphi_1^l(z) \cdot q(z).$$

El polinomio $q(z)$, representando en sí un cociente de dos polinomios con coeficientes reales, tiene también coeficientes reales. Por hipótesis, el número v debe ser la raíz de multiplicidad $(k - l)$ de este polinomio que no tiene raíz igual a $\bar{v}$. De acuerdo con lo demostrado anteriormente, esto no es posible, razón por la cual $k = l$. De este modo, todas las raíces complejas de cualquier polinomio de coeficientes reales son complejas conjugadas dos a dos. De la unicidad de la descomposición canónica proviene la siguiente deducción.

Todo polinomio con coeficientes reales puede ser representado, salvo el orden en que se disponen los factores, de un modo único en forma del producto de su coeficiente mayor y los polinomios con coeficientes reales. Los últimos polinomios tienen sus coeficientes mayores equivalentes a uno y son lineales, si corresponden a las raíces reales, y cuadráticos, cuando corresponden a un par de raíces complejas conjugadas.

Por fin, viene la deducción más importante, en aras de la cual, en esencia, se demostraba el teorema fundamental del álgebra. Supongamos que el operador lineal A actúa en un espacio complejo. Los valores propios de este operador, y sólo ellos, son las raíces del polinomio característico. Conforme al teorema fundamental, el operador A tiene por lo menos un valor propio λ. Por consiguiente,

Todo operador lineal que actúa en un espacio complejo lineal tiene por lo menos un vector propio.

Ha de ser observado que si el operador A actúa en un espacio real y racional dicha deducción ya *no es justa*.

En relación con los valores propios se empleará la misma terminología que se ha aplicado respecto a las raíces del polinomio. En particular, un valor propio se llamará simple, si es una raíz simple del polinomio característico y se denominará múltiple, en el caso contrario. Se llamará multiplicidad del valor propio λ la multiplicidad de λ, al intervenir éste en calidad de raíz del polinomio característico.

Ejercicios.

1. Demuéstrese que si el número complejo $a \neq 0$, para todo número natural n sólo existen n números complejos distintos, cuya n-ésima potencia es igual a a.

2. ¿De qué modo están ligadas entre sí las raíces de los polinomios $f(z)$ y $f(z - a)$, donde a es un número complejo?

3. Supongamos que el polinomio $f(z)$ con coeficientes reales, de grado no superior a n toma los valores iguales para $n+1$ distintos valores del argumento. Demuéstrese que $f(z)$ es un polinomio de grado nulo.

4. Demuéstrese que cualquier polinomio de grado impar, con coeficientes reales, tiene por lo menos una raíz real.

5. Demuéstrese que el polinomio $f(z)$ tiene por lo menos una raíz en cada uno de los dos dominios

$$|z| \leqslant \sqrt[n]{\left|\frac{a_0}{a_n}\right|}, \quad |z| \geqslant \sqrt[n]{\left|\frac{a_0}{a_n}\right|}.$$

6. Demuéstrese que el operador A es de estructura simple, cuando, y sólo cuando, a todo valor propio le corresponden tantos valores propios linealmente independientes cual es la multiplicidad de λ.

CAPITULO 9 ESTRUCTURA DEL OPERADOR LINEAL

§ 70. Subespacios invariantes

Las investigaciones que siguen se realizarán bajo el supuesto de que el operador lineal viene dado en un espacio complejo X. Como ya se ha indicado antes, dicha suposición asegura que para cualquier operador lineal existe al menos un vector propio.

El subespacio L de un espacio lineal X se denomina *invariante* respecto del operador A, si para todo vector x de L su imagen Ax pertenece también a L.

Todo operador lineal tiene al menos dos subespacios invariantes *triviales*: el subespacio nulo y todo el espacio X. Significación esencial sólo tienen los subespacios invariantes no triviales. A los subespacios semejantes pertenecen, por ejemplo, los subespacios propios. Como en un espacio lineal complejo cualquier operador tiene, a ciencia cierta, por lo menos un vector propio, cualquier operador en tal espacio tiene obligatoriamente al menos un subespacio invariante no trivial.

Es fácil comprobar que para todo operador A el campo de valores T_A y el núcleo N_A serán subespacios invariantes. Estos subespacios son triviales cuando, y sólo cuando, el operador A sea regular o nulo.

Si L es un subespacio invariante, existen varios métodos por medio de los cuales se puede construir un subespacio complementario M tal que $X = L \dot{+} M$. Sin embargo, entre estos subespacios complementarios *puede no haber ninguno* que sea invariante. En cambio, si existe aunque sea un solo subespacio complementario invariante, podemos hablar sobre la descomposición del espacio en una suma directa de subespacios invariantes.

El conocimiento de algún subespacio invariante y más aún, de la descomposición del espacio en una suma directa de subespacios invariantes permite construir una base en la cual la matriz del operador tenga la forma más simple. Supongamos que el operador A tiene en el espacio m-dimensional X un subespacio invariante L de dimensión n. Elijamos en X una base $e_1, e_2, \ldots, e_m$ de tal modo que sus primeros n vectores pertenezcan a L. En este caso las imágenes $Ae_1, \ldots, Ae_n$ de los vectores $e_1, \ldots, e_n$ pertenecerán a L y se podrá descomponerlas según los vectores $e_1, \ldots, e_n$, considerando los últimos como vectores de la base L. Por consiguiente,

$$\begin{aligned} Ae_1 &= a_{11}e_1 + a_{21}e_2 + \ldots + a_{n1}e_n, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Ae_n &= a_{1n}e_1 + a_{2n}e_2 + \ldots + a_{nn}e_n. \end{aligned}$$

Recordemos que los elementos de las columnas de la matriz de un operador coinciden con las coordenadas de las imágenes de los vectores de la base. Por ello la matriz A_e del operador A en la base $e_1, e_2, \ldots, e_n$ tendrá por expresión:

$$A_e = \left(\begin{array}{ccc|ccc} a_{11} & \ldots & a_{1n} & a_{1,\,n+1} & \ldots & a_{1m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \ldots & a_{nn} & a_{n,\,n+1} & \ldots & a_{nm} \\ \hline 0 & \ldots & 0 & a_{n+1,\,n+1} & \ldots & a_{n+1,\,m} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & 0 & a_{m,\,n+1} & \ldots & a_{mm} \end{array}\right).$$

Por regla general, las matrices de este tipo se escriben en la así llamada forma *celular*. A saber,

$$A_e = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}. \tag{70.1}$$

Aquí, A_{11} es una matriz cuadrática de orden n, A_{22} es una matriz cuadrática de orden $m - n$, 0 es una matriz nula de dimensiones $(m - n) \times n$, y A_{12}, la matriz de dimensiones $n \times (m - n)$.

Supongamos ahora que el espacio X está descompuesto en suma directa de los subespacios invariantes L y M. Elijamos en X una base $e_1, e_2, \ldots, e_m$ de modo tal que los primeros n vectores de la base pertenezcan a L y los demás $m - n$ vectores pertenezcan a M. En este caso las imágenes $Ae_1, \ldots, Ae_n$ pueden descomponerse sólo según los vectores $e_1, \ldots, e_n$, mientras que las imágenes $Ae_{n+1}, \ldots, Ae_m$, sólo según los vectores $e_{n+1}, \ldots, e_m$. La matriz A_{12} en (70.1) será, evidentemente, nula. Por esta razón la matriz A_e del operador A en la base considerada tendrá una forma aún más simple. A saber,

$$A_e = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}$$

Supongamos que la operación del operador A se investiga sólo en los vectores del subespacio invariante L. Si $x \in L$, se tiene $Ax \in L$. Por consiguiente, podemos considerar que el operador A genera en L

otro operador $A \mid L$ definido por la igualdad

$$(A \mid L)\, x = Ax$$

para todo $x \in L$. El operador $A \mid L$ se denomina *operador inducido, generado por el operador* A. Con relación al operador $A \mid L$, el operador A se llama *generador*. Por ser el operador A lineal, el operador inducido será también lineal. El operador inducido $A \mid L$ coincide con el operador generador A en el subespacio invariante L y no está definido fuera de L. De este modo, dichos operadores se diferencian, principalmente, por el campo de definición.

A pesar de que la introducción del operador inducido parece ser algo artificial, éste representa en sí un aparato auxiliar muy cómodo en la ejecución de las más diversas investigaciones. Por ejemplo, un operador inducido, al igual que cualquier otro operador lineal, tiene por lo menos un vector propio. Pero, como dicho operador coincide con el operador generador en su campo de definición, esto significa que:

Todo operador lineal tiene en cualquier subespacio invariante por lo menos un vector propio.

Si un espacio está descompuesto en una suma directa de r subespacios invariantes, el operador lineal tiene por lo menos r vectores propios linealmente independientes.

Está claro que todo valor propio y todo vector propio de un operador inducido son, respectivamente, el valor propio y el vector propio del operador generador. Resulta menos evidente el

TEOREMA 70.1. *El polinomio característico de un operador inducido engendrado en un subespacio no trivial es el divisor del polinomio característico del operador generador.*

DEMOSTRACION. Supongamos que el operador inducido $A \mid L$ está definido en el subespacio invariante L. Elijamos nuevamente en el espacio X una base $e_1, \ldots, e_m$ de tal modo que los vectores $e_1, \ldots, e_n$ compongan la base en L. Si la matriz del operador generador es A_e de (70.1), la matriz del operador inducido $A \mid L$ es A_{11} de (70.1). El polinomio característico para el operador A es igual a $\det(\lambda E - A_e)$; para el operador $A \mid L$ el mismo es igual a $\det(\lambda E - A_{11})$. Aplicando el teorema de Laplace para el desarrollo del determinante $\det(\lambda E - A_e)$ por las primeras n columnas, encontramos

$$\det(\lambda E - A_e) = \det\begin{pmatrix} \lambda E - A_{11} & -A_{12} \\ 0 & \lambda E - A_{22} \end{pmatrix} = \det(\lambda E - A_{11}) \times \det(\lambda E - A_{22}).$$

La igualdad obtenida significa precisamente la validez de la afirmación del teorema.

La determinación de todos los valores propios del operador A se reduce a la búsqueda de todas las raíces del polinomio característico. Si el operador A tiene un subespacio invariante no trivial, entonces, de acuerdo con el teorema 70.1, la tarea puede ser reducida a la búsqueda de todas las raíces de dos polinomios de grados inferiores. Si el operador inducido tiene por sí mismo un subespacio invariante no trivial, el proceso de descomposición del polinomio característico en factores puede ser continuado.

Ejercicios.

1. Demuéstrese que la suma y la intersección de los subespacios invariantes son unos subespacios invariantes.
2. Demuéstrese que si el operador A es regular, entonces cualquier operador inducido es también regular.
3. Demuéstrese que si el operador A es de estructura simple, cualquier operador inducido es también de estructura simple.
4. ¿En qué caso el subespacio invariante de un operador de estructura simple será la suma directa de los subespacios propios?
5. Demuéstrese que si aunque sea uno de los subespacios invariantes del operador A está privado de un subespacio invariante complementario, A no puede ser operador de estructura simple.
6. Demuéstrese que si el operador A es de estructura simple, el campo de valores y el núcleo están privados de vectores no nulos comunes.
7. Demuéstrese que si un subespacio es invariante respecto del operador A, será invariante también respecto del operador $a_0E + a_1A + \ldots + a_pA^p$.

§ 71. Polinomio operacional

Uno de los métodos más importantes para construir subespacios invariantes de un operador lineal consiste en el empleo de los polinomios con coeficientes complejos.

Supongamos que el operador lineal A actúa en el espacio complejo X. Elijamos un polinomio arbitrario

$$\varphi(z) = a_0\, a_1 z + \ldots + a_p z^p$$

con coeficientes complejos y consideremos un operador lineal

$$\varphi(A) = a_0E + a_1A + \ldots + a_pA^p.$$

Este operador actúa en el espacio X y se denomina *polinomio operacional* o *polinomio del operador* A.

Fijemos el operador A y construyamos el conjunto de todos los polinomios operacionales que dependen de A. Puesto que el conjunto de todos los polinomios es un anillo conmutativo, lo será también el conjunto de todos los polinomios operacionales. En particular, de aquí se deduce que

$$\varphi(A)\,A = A\,\varphi(A)$$

para cualquier polinomio $\varphi(z)$. *La conmutatividad de un anillo de polinomios operacionales desempeña un papel exclusivamente importante en todas las investigaciones a seguir.*

Es fácil mostrar que el campo de valores T_φ de todo polinomio operacional $\varphi(A)$ es un subespacio invariante para el operador A. En efecto, sea $x \in T_\varphi$. Esto significa que $x = \varphi(A)\,y$ para cierto $y \in X$. Por ser A y $\varphi(A)$ conmutables, tenemos

$$Ax = A\varphi(A)\,y = \varphi(A)\,(Ay).$$

Por consiguiente, el vector Ax es el resultado de aplicar el operador $\varphi(A)$ al vector $Ay \in X$, es decir, $Ax \in T_\varphi$.

El núcleo N_φ del polinomio operacional $\varphi(A)$ es también un subespacio invariante para el operador A. Si $x \in N_\varphi$, entonces $\varphi(A)\,x = 0$, pero en este caso

$$\varphi(A)\,(Ax) = A\,(\varphi(A)\,x) = A\,(0) = 0.$$

Ya se ha observado anteriormente que cualquier subespacio invariante contiene por lo menos un vector propio del operador. Ahora podemos enunciar una afirmación más exacta. A saber,

Si el valor propio del operador A es (no es) una raíz del polinomio $\varphi(z)$, todos los vectores propios del operador A, correspondientes a dicho valor propio, pertenecen al núcleo (campo de valores) del operador $\varphi(A)$.

Efectivamente, sea x un vector propio del operador A correspondiente al valor propio λ. En los ejercicios del § 65 se subrayaba que el vector x es también propio para el operador $\varphi(A)$, pero corresponde al valor propio $\varphi(\lambda)$. Por consiguiente, $\varphi(A)\,x = \varphi(\lambda)\,x$. Si λ es una raíz del polinomio $\varphi(z)$, entonces $\varphi(\lambda) = 0$ y el vector x pertenece al núcleo del operador $\varphi(A)$. En cambio si $\varphi(\lambda) \neq 0$, el vector $\varphi(\lambda)\,x$ será no nulo y el vector x pertenecerá al campo de valores del operador $\varphi(A)$.

No podemos demostrar que cualquier subespacio invariante del operador A es o bien un campo de valores o bien un núcleo de cierto polinomio operacional. Esta afirmación, por lo general, no es cierta, lo cual se afirma en el ejemplo del operador idéntico. Cualquiera que sea el polinomio $\varphi(z)$, se tiene $\varphi(E) = \varphi(1)\,E$, por lo cual el operador $\varphi(E)$ es o bien nulo o bien regular. Por consiguiente, el campo de valores y el núcleo para $\varphi(E)$ representan siempre unos subespacios triviales. En cuanto al subespacio invariante del operador E, lo constituirá cualquier subespacio. Con todo esto, cada subespacio invariante del operador A está ligado de modo determinado a los polinomios operacionales de A. Es válido el

TEOREMA 71.1. *Sea L un subespacio invariante arbitrario del operador A. Si todos los valores propios del operador, inducido sobre L, son las raíces del polinomio $\varphi(z)$, entonces L está contenido en los núcleos de los operadores $\varphi^k(A)$ para todas las potencias k, enteras positivas y suficientemente grandes.*

DEMOSTRACION. Designaremos mediante T'_1, T'_2, ... los campos de valores de los operadores inducidos sobre L con ayuda de los polinomios operacionales $\varphi^k(A)$ para $k = 1, 2, \ldots$. El operador

$\varphi(A)$ es degenerado en L, puesto que en su núcleo están contenidos por lo menos todos los vectores propios del operador A pertenecientes a L. Por ello, $T'_1 \subset L$, $\dim T'_1 < \dim L$. El subespacio T'_1 es invariante respecto de A. Si T'_1 es no nulo, entonces, de acuerdo con el teorema (70.1), el polinomio característico del operador inducido sobre T'_1 con ayuda de A es el divisor del polinomio característico del operador inducido sobre L con ayuda de A. Por consiguiente, todos los valores propios del operador, inducido sobre T'_1, son también las raíces del polinomio $\varphi(z)$. Pero de aquí se desprende nuevamente que $T'_2 \subset T'_1$, $\dim T'_2 < \dim T'_1$, etc. Las dimensiones de los subespacios T'_1, T'_2, ... no pueden decrecer ilimitadamente. Por esta razón, a partir de cierto número k, dichos subespacios irán quedando nulos lo que es testimonio de la validez de la afirmación del teorema.

Las investigaciones efectuadas permiten establecer un hecho muy importante concerniente a la existencia de los subespacios invariantes no triviales.

TEOREMA 71.2. *Todo operador lineal A, que actúa en el espacio complejo m-dimensional X, tiene por lo menos un subespacio invariante de dimensión $m - 1$.*

DEMOSTRACIÓN. El operador A tiene por lo menos un vector propio x. Supongamos que dicho vector corresponde al valor propio λ. De acuerdo con lo demostrado, el campo de valores T_λ del operador $A - \lambda E$ representa un subespacio invariante del operador A. Pero, puesto que el operador $A - \lambda E$ es degenerado, la dimensión del subespacio T_λ no es superior a $m - 1$.

Consideraremos ahora cualquier subespacio L de dimensión $m - 1$ en el que está íntegramente contenido el subespacio T_λ. Todo vector del espacio X se transforma por el operador $A - \lambda E$ en cierto vector de T_λ. Por eso todo vector de L pasa a ser, de nuevo, un vector de L. De este modo, L es el subespacio invariante respecto de $A - \lambda E$, y, por supuesto, invariante respecto de A. El teorema queda demostrado.

Ejercicios

1. Sea A un operador de diferenciación que actúa en un espacio real de polinomios, dimensión finita. ¿Qué representa en sí el operador $\varphi(A)$ para el polinomio $\varphi(z)$ con coeficientes reales?

2. Sea $\varphi(z)$ un polinomio característico del operador inducido generado por el operador A en el subespacio invariante N. Demuéstrese que N pertenece al núcleo del operador $\varphi^k(A)$ para cierto k entero y positivo.

3. Demuéstrese que si todos los valores propios del operador A son las raíces del polinomio $\varphi(z)$, entonces $\varphi^k(A) = 0$ para cierto k entero y positivo.

4. Demuéstrese que un anillo de polinomios operacionales generados por un operador cualquiera tiene divisores de cero.

5. Demuéstrese que si el operador A es de estructura simple, el operador $\varphi(A)$ será también de estructura simple. ¿Será cierta la afirmación inversa?

§ 72. Forma triangular

Ahora podemos resolver el problema de reducción de la matriz de un operador a una de las formas más sencillas, esto es, a la así llamada *forma triangular.*

TEOREMA 72.1. *Para cualquier operador lineal A que actúa en un espacio m-dimensional X existen tales subespacios invariantes L_p de dimensión p, $p = 0, 1, \ldots, m - 1$, que*

$$L_0 \subset L_1 \subset \ldots \subset L_{m-1} \subset L_m.$$

DEMOSTRACION. La existencia de los subespacios L_0 y L_m es evidente. Conforme a lo demostrado más arriba, el operador A tiene subespacio invariante L_{m-1} de dimensión $m - 1$.

Consideraremos en el subespacio L_{m-1} un operador inducido. Al igual que cualquier otro operador prefijado en L_{m-1}, el operador inducido tiene el subespacio invariante L_{m-2} de dimensión $m - 2$. Pero un subespacio que es invariante para el operador inducido será también invariante para el operador generador A. De este modo, la existencia del subespacio L_{m-2} está demostrada. Al considerar un operador inducido en el subespacio L_{m-2}, veremos que de modo análogo se demuestra la existencia del subespacio L_{m-3}, etc.

Este teorema es interesante, en primer lugar, por su interpretación matricial. Construyamos la base $e_1, e_2, \ldots, e_m$ del espacio X, haciendo uso de los subespacios invariantes L_p. A título de vector e_1 tomemos cualquier vector no nulo de L_1, a título de vector e_2 tomemos cualequier vector no nulo de L_2, no perteneciente a L_1, y, en general, a título de vector e_p tomemos cualquier vector no nulo de L_p, no perteneciente a L_{p-1}. Consideremos la matriz A_e del operador A en esta base. Puesto que e_j pertenece a L_j, mientras que L_j es invariante respecto de A, entonces el vector Ae_j debe representar una combinación lineal sólo de los vectores $e_1, e_2, \ldots, e_j$. Quiere decir que en la descomposición

$$Ae_j = a_{1j}e_1 + a_{2j}e_2 + \ldots + a_{mj}e_m$$

el coeficiente de e_i debe ser igual a cero para cualesquiera $i > j$. Por consiguiente, la matriz del operador A tiene la forma

$$A_e = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1,m-1} & a_{1m} \\ 0 & a_{22} & \ldots & a_{2,m-1} & a_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & a_{m-1,m-1} & a_{m-1,m} \\ 0 & 0 & \ldots & 0 & a_{mm} \end{pmatrix}$$

donde $a_{ij} = 0$ para $i > j$.

Una matriz, todos los elementos de la cual, dispuestos por debajo (por arriba) de la diagonal principal, son nulos, se denomina *matriz* *triangular derecha* (*izquierda*). En el lenguaje de las matrices el resultado obtenido significa que toda matriz cuadrada es semejante a una matriz triangular derecha.

La forma triangular de una matriz es de amplio uso en la demostración de los más diversos hechos concernientes a los operadores lineales. Esto se debe, principalmente, a la siguiente propiedad de la misma:

Si el operador A tiene en cierta base la matriz triangular A_e, los elementos diagonales de la matriz A_e coinciden con los valores propios del operador A incluso cuando se toma en consideración la multiplicidad de éstos.

En efecto, al aplicar el teorema de Laplace, encontramos que el polinomio característico de la matriz A_e es igual a

$$\det(\lambda E - A_e) = \prod_{i=1}^{m} (\lambda - a_{ii}),$$

de donde proviene que la afirmación enunciada es cierta.

Una parte considerable de la ulterior teoría de los operadores lineales se dedica al perfeccionamiento del resultado recién obtenido acerca de la reducción de la matriz de un operador a la forma triangular. La forma más simple que puede tener la matriz de un operador es diagonal. Como se sabe, a esta forma pueden reducirse sólo las matrices de los operadores de estructura simple. No obstante, para los operadores de estructura no simple la forma triangular tampoco es la más sencilla.

Ejercicios

1. Demuéstrese que cualquier matriz cuadrada es semejante a una matriz triangular izquierda.
2. Demuéstrese que un conjunto de matrices triangulares de un mismo orden y de una misma denominación forma un anillo.
3. Demuéstrese que un conjunto de matrices triangulares regulares de un mismo orden y de la misma denominación forma un grupo.
4. Sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ los valores propios del operador A escritos por orden tomando en consideración su multiplicidad. Demuéstrese que, tomando en consideración la multiplicidad, los números $\varphi(\lambda_1), \varphi(\lambda_2), \ldots, \varphi(\lambda_m)$ serán valores propios del operador $\varphi(A)$, cualquiera que sea el polinomio $\varphi(z)$.
5. Demuéstrese que si todos los elementos diagonales de una matriz triangular A de orden m son nulos, entonces $A^m = 0$.
6. Supongamos que una matriz triangular es semejante a la diagonal. Demuéstrese que la matriz de la transformación de semejanza puede ser elegida triangular de la misma denominación.

§ 73. Suma directa de los operadores

Un operador lineal cuyos valores propios son todos iguales es, en cierto sentido, una exclusión. No obstante, mostraremos que todo operador lineal puede ser compuesto precisamente de operadores de este tipo.

Supongamos que el espacio X está representado en forma de la suma directa de los subespacios L y M. Fijemos en el subespacio L cierto operador B y en el subespacio M, el operador C. Para todo vector $x \in X$ tiene lugar la descomposición única

$$x = x_L + x_M, \tag{73.1}$$

donde $x_L \in L$, $x_M \in M$.

Un operador A, definido mediante la igualdad

$$Ax = Bx_L + Cx_M,$$

se llama *suma directa* de los operadores B y C. Si uno de los subespacios L, M es trivial, la suma directa se denomina también trivial.

Es fácil comprobar que A es un operador lineal en X. Señalemos que dicho operador puede ser representado de un modo único en forma de la suma directa de los operadores definidos en los subespacios L y M. En efecto, para todo vector $x \in L$ se tiene $Ax = Bx$. Por analogía, $Ax = Cx$ para todo $x \in M$. Esto significa que el operador C coincide con el operador inducido $A \mid L$ y el operador C coincide con el $A \mid M$.

Consideremos ahora un operador arbitrario A en el espacio X. Si X queda descompuesto de tal o cual manera en la suma directa de los subespacios L y M, invariantes respecto del operador A, entonces el propio *operador A puede ser descompuesto en suma directa*. Efectivamente, construyamos los operadores inducidos $A \mid L$ y $A \mid M$. Al descomponer una vez más el vector arbitrario $x \in X$ en forma de la suma (73.1), obtendremos

$$Ax = (A \mid L)\, x_L + (A \mid M)\, x_M.$$

En este caso, en virtud del teorema 70.1, el polinomio característico del operador A es igual al producto de los polinomios característicos de los operadores inducidos $A \mid L$ y $A \mid M$.

El operador A puede descomponerse en una suma directa con ayuda de cualquier polinomio operacional $\varphi(A)$. Designemos con N_k el núcleo del operador $\varphi^k(A)$. Esto es un subespacio invariante respecto de A y, evidentemente, $N_1 \subset N_2 \subset \ldots$ Demostremos primeramente que si $N_k = N_{k+1}$ para cierto k, entonces $N_k = N_p$ para todo $p > k$. En efecto, elijamos cualquier vector $x \in N_p$, entonces $\varphi^p(A)\, x = 0$. Al escribir esta igualdad en la forma $\varphi^{k+1}(A) \times$ $\times\, (\varphi^{p-k-1}(A)\, x) = 0$, concluimos que el vector $\varphi^{p-k-1}(A)\, x \in N_{k+1}$. En virtud de la igualdad $N_k = N_{k+1}$, este mismo vector pertenece a N_k. Por consiguiente,

$$\varphi^k(A)\,(\varphi^{p-k-1}(A)\, x) = \varphi^{p-1}(A)\, x = 0,$$

es decir, el vector $x \in N_{p-1}$. La validez de la afirmación enunciada se establece ahora por inducción según p.

El espacio X, donde actúa el operador A, es de dimensión finita, por lo cual las dimensiones de los subespacios N_k no pueden crecer ilimitadamente. Sea q un número positivo entero mínimo, para el cual $N_q = N_{q+1}$. Designemos mediante T_k el campo de valores del operador $\varphi^k(A)$ y examinemos cualquier vector común x de los subespacios T_q y N_q. Tenemos $\varphi^q(A)\,x = 0$ y $x = \varphi^q(A)\,y$ para cierto vector $y \in X$. De aquí proviene que $\varphi^{2q}(A)\,y = 0$, es decir, $y \in N_{2q}$. Pero, de acuerdo con lo demostrado, se verifica la igualdad $N_q = N_{2q}$. Por esta razón $y \in N_q$, es decir, $x = \varphi^q(A)\,y = 0$.

Así pues, para los subespacios T_q y N_q sólo el vector nulo resulta ser común. En virtud de la fórmula (56.3), esto significa que $X = T_q \dot{+} N_q$. Puesto que los subespacios T_q y N_q son invariantes, la posibilidad de descomponer el operador queda establecida.

Según se ha observado anteriormente, todos los vectores propios del operador A deben encontrarse en los subespacios T_q y N_q. Además, en N_q se encuentran aquellos de los vectores que corresponden a los valores propios, coincidentes con las raíces cualesquiera del polinomio $\varphi(z)$, mientras que en T_q se hallan aquellos vectores propios, para los cuales los valores propios correspondientes no coinciden con ninguna de las raíces de $\varphi(z)$. Puesto que a todo valor propio le corresponde aunque sea un solo vector propio, de estos razonamientos se deduce que:

Cada una de las raíces (ninguna de las raíces) del polinomio característico de un operador, inducido sobre N_q (T_q), es (no es) la raíz del polinomio $\varphi(z)$.

La característica definitiva de las descomposiciones de un operador en una suma directa, obtenida con ayuda de los polinomios operacionales, la proporciona el

TEOREMA 73.1 *Supongamos que el polinomio característico $f(z)$ del operador A está descompuesto en el producto de los polinomios $\varphi(z)$ y $\psi(z)$ que no tienen raíces comunes. En este caso el operador A puede descomponerse del modo único en la suma directa de los operadores B y C con polinomios característicos $\varphi(z)$ y $\psi(z)$.*

DEMOSTRACION. Consideraremos la descomposición del operador A en una suma directa, obtenida con ayuda del polinomio $\varphi(z)$. Como el producto de los polinomios característicos de los operadores, que definen la suma directa, coincide con el polinomio característico $f(z)$, entonces la existencia de al menos una descomposición se deduce de las investigaciones realizadas más arriba.

Supongamos ahora que el espacio X está descompuesto de uno u otro modo en la suma directa de los espacios invariantes N y T. En este caso el operador inducido tiene en N un polinomio característico $\varphi(z)$ y el operador en T, el polinomio $\psi(z)$. Según el teorema 71.1, $N \subset N_k$ para cualesquiera k suficientemente grandes, por lo cual $N \subset N_q$. El operador $\varphi(A)$ es regular en T y, por consiguiente, el conjunto de imágenes de los vectores de T respecto de $\varphi(A)$ coincide con T. Pero esto precisamente significa que $T \subset T_k$

para todo k. Los subespacios N, T, como también N_q, T_q, forman en suma directa el espacio X. Por ello, las inclusiones $N \subset N_q$, $T \subset T_q$ pueden tener lugar sólo en el caso en que $N = N_q$, $T = T_q$. El teorema está demostrado.

Sea un operador A que actúa en el espacio m-dimensional X. Representemos el polinomio característico $f(z)$ del operador A en forma de la descomposición canónica

$$f(z) = (z-\lambda_1)^{k_1}(z-\lambda_2)^{k_2}\ldots(z-\lambda_r)^{k_r}, \qquad (73.2)$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_r$ son unos valores propios distintos dos a dos y $k_1 + k_2 + \ldots + k_r = m$. Examinemos los polinomios

$$(z-\lambda_1)^{k_1}, (z-\lambda_2)^{k_2}, \ldots, (z-\lambda_r)^{k_r}.$$

Estos son los divisores del polinomio característico $f(z)$ y ningún par de ellos tiene raíces comunes. De acuerdo con el teorema 73.1, existen los subespacios invariantes $R_1, R_2, \ldots, R_r$ tales que

$$X = R_1 + R_2 + \ldots + R_r.$$

En estas circunstancias la dimensión del subespacio R_i es igual a k_i y el operador inducido en R_i tiene el polinomio característico $(z-\lambda_i)^{k_i}$.

El subespacio R_i se llama subespacio *radical* del operador A, correspondiente al valor propio λ_i. Los vectores de un subespacio radical se denominan *radicales*. De lo dicho se deduce que todo operador puede ser descompuesto en la suma directa de los operadores inducidos en los subespacios radicales.

El subespacio radical R_i coincide con el núcleo del operador $((A-\lambda_i E)^{k_i})^q$ para cierto q entero y positivo. Probemos que en el caso dado siempre se puede poner $q = 1$. Examinemos los operadores $(A-\lambda_i E)^p$ para $p = 1, 2, \ldots$ Sea p_i un número mínimo para el cual el núcleo del operador $(A-\lambda_i E)^{p_i}$ coincide con el del operador $(A-\lambda_i E)^{p_i+1}$. En este caso el subespacio radical R_i coincidirá con el núcleo del operador $(A-\lambda_i E)^{p_i}$. Como las dimensiones de los núcleos de los operadores $(A-\lambda_i E)^p$ para $p = 1, 2, \ldots$ crecen de modo monótono y la dimensión del subespacio R_i es igual a k_i, resulta que $p_i \leqslant k_i$.

De este modo, R_i, correspondiente al valor propio λ_i de multiplicidad k_i, coincide, a ciencia cierta, con el núcleo del operador $(A-\lambda_i E)^{k_i}$.

TEOREMA 73.2. (*teorema de Cayley—Hamilton*). *Si $f(z)$ es un polinomio característico del operador A, entonces $f(A)$ es un operador nulo.*

DEMOSTRACIÓN. Representemos el polinomio característico en forma de la descomposición canónica (73.2). Puesto que el polinomio operacional $f(A)$ contiene el factor $(A-\lambda_i E)^{k_i}$ y cualesquiera polinomios de un mismo operador son permutables, $f(A)x_i = 0$ para todo vector x_i de R_i. Elijamos ahora un vector arbitrario x y representémoslo en forma de la descomposición $x = x_1 + x_2 + \ldots + x_r$, donde $x_i \in R_i$. Ahora está claro que $f(A)x = 0$, es decir, $f(A)$ es un operador nulo.

Esta vez también representa un interés considerable la interpretación matricial de los resultados obtenidos. Formemos la base de un espacio como reunión sucesiva de cualesquiera bases de los subespacios radicales $R_1, R_2, \ldots, R_r$. Los subespacios radicales son invariantes y la suma directa de ellos coincide con X. Por esta razón, la matriz A_e del operador A tendrá en la base dada la así llamada forma *casi diagonal*

$$\begin{pmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{rr} \end{pmatrix}. \tag{73.3}$$

Toda matriz A_{ii} tiene el orden k_i y representa en sí una matriz del operador inducido sobre el subespacio R_i.

Ejercicios.

1. ¿Se podrá descomponer en una suma directa no trivial un operador de diferenciación dado en un espacio de polinomios de dimensión finita?
2. Demuéstrese que un sistema de vectores radicales, que corresponden dos a dos a los diferentes valores propios, es linealmente independiente.
3. Demuéstrese que si el operador A es regular, entonces $A^{-1} = \varphi(A)$ para cierto polinomio $\varphi(z)$.
4. Un operador A se denomina *nilpotente*, si $A^p = 0$ para cierto p entero y positivo. Demuéstrese que un operador es nilpotente cuando, y sólo cuando, todos sus valores propios son iguales a cero.
5. Sea $\varphi(z)$ un polinomio de grado inferior, para el cual $\varphi(A) = 0$. Demuéstrese que $\varphi(z)$ es un divisor para el polinomio característico del operador A.

§ 74. Forma de Jordan

La simplificación ulterior de la matriz de un operador en comparación con la forma casi diagonal (73.3) puede llevarse a cabo sólo a cuenta de la construcción especial de las bases de cada uno de los subespacios radicales. Por supuesto, las bases radicales se pueden escoger de un modo tal que cualquier matriz A_{ii} en (73.3) sea triangular. No obstante, esta forma de la matriz de un operador tampoco es la más sencilla.

Volvamos al estudio más detallado de la estructura de subespacios radicales. Si es que $x \in R_i$, entonces $(A - \lambda_i E)^{k_i} x = 0$. Pero para cada vector concreto x es muy posible que se verifique la igualdad $(A - \lambda_i E)^m x = 0$ también con $m < k_i$. En particular, si x es un vector propio correspondiente al valor propio múltiple λ_i, entonces $(A-\lambda_i E)x=0$, aunque $k_i \geqslant 2$.

Se llama *altura* del vector radical x un número entero mínimo no negativo m, para el cual $(A-\lambda_i E)^m x=0$.

Todos los vectores radicales correspondientes al valor propio λ_i tienen alturas no superiores a la multiplicidad de λ_i. Sin embargo, recordemos que en el caso general las alturas de los vectores radicales y las multiplicidades de los valores propios son dos conceptos diferentes. Así, por ejemplo, para un operador de estructura simple no existen, en general, vectores radicales cuya altura sea más que uno, independientemente de las multiplicidades de los valores propios.

Sea R_i un subespacio radical correspondiente al valor propio λ_i de multiplicidad k_i. Designemos mediante t la altura máxima de los vectores radicales de R_i. Está claro que $t \leqslant k_i$. Si el vector x es de altura k, el vector $(A-\lambda_i E)x$ tendrá la altura $k-1$. Por ello, en el subespacio R_i hay vectores radicales de cualquier altura desde 0 hasta t.

Para todo $k \leqslant t$, designemos con H_k la totalidad de todos los vectores cuyas alturas no son superiores a k. Es fácil mostrar que H_k es un subespacio en R_i. Si $x, y \in H_k$, entonces $(A-\lambda_i E)^k x = (A-\lambda_i E)^k y=0$. Pero en este caso con cualesquiera números α, β tenemos $(A-\lambda_i E)^k(\alpha x+\beta y)=0$, es decir, $\alpha x+\beta y \in H_k$. Luego, es evidente que

$$0=H_0 \subset H_1 \subset \ldots \subset H_{t-1} \subset H_t = R_i.$$

Las dimensiones de estos subespacios se designarán mediante m_k, $0=m_0<m_1<\ldots<m_{t-1}<m_t=k_i$.

Sean $f_1, \ldots, f_{p_1}$ unos vectores arbitrarios linealmente independientes de H_t cuya cápsula lineal da H_t en la suma directa con H_{t-1}. Está claro que éstos serán los vectores radicales de altura t, $p_1=m_t-m_{t-1}$, y no existe ninguna combinación lineal no nula de los vectores $f_1, \ldots, f_{p_1}$, que pertenezca a H_{t-1}. Consideraremos una totalidad de los vectores

$$\begin{array}{c} f_1, \ldots, \qquad f_{p_1} \\ (A-\lambda_i E)f_1, \ldots, (A-\lambda_i E)f_{p_1}, \\ (A-\lambda_i E)^2 f_1, \ldots, (A-\lambda_i E)^2 f_{p_1}, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ (A-\lambda_i E)^{t-1} f_1, \ldots, (A-\lambda_i E)^{t-1} f_{p_1}. \end{array} \tag{74.1}$$

Probemos que estos vectores son linealmente independientes. En efecto, formemos su combinación lineal e igualemos ésta a cero. Aplicando a ambos miembros de la igualdad obtenida el operador $(A-\lambda_i E)^{t-1}$, llegamos a que la combinación lineal de los vectores $f_1, \ldots, f_{p_1}$ se transforma por el operador $(A-\lambda_i E)^{t-1}$ en un vector nulo, es decir, dicha combinación es un vector de H_{t-1}. Por consiguiente, los coeficientes de estos vectores deben ser nulos. Apliquemos ahora a la misma igualdad el operador $(A - \lambda_i E)^{t-2}$. De manera análoga nos convencemos de que los coeficientes de los vectores dispuestos en la segunda fila de (74.1) deben ser nulos, etc.

Hemos de notar que por ser elegidos los vectores $f_1, \ldots, f_{p_1}$, no existe ninguna combinación lineal no nula de los vectores dispuestos en la t-ésima fila de (74.1) que pertenezca a H_{t-t}.

Completemos los vectores $(A - \lambda_i E) f_1, \ldots, (A - \lambda_i E) f_{p_1}$ con tales vectores $f_{p_1+1}, \ldots, f_{p_2}$ de H_{t-1} que toda esta totalidad sea linealmente independiente y que su cápsula lineal dé H_{t-1} en suma directa con H_{t-2}. Está claro que éstos serán vectores radicales de altura $t - 1$, $p_2 = m_{t-1} - m_{t-2}$, y no existe ninguna combinación lineal no nula de los vectores dados que pertenezca a H_{t-2}. Construyamos nuevamente una totalidad de vectores

$$\begin{array}{c} f_{p_1+1}, \ \ldots, \qquad\qquad f_{p_2} \\ (A-\lambda_i E) f_{p_1+1}, \ \ldots, \ (A-\lambda_i E) f_{p_2}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (A-\lambda_i E)^{t-2} f_{p_1+1}, \ \ldots, \ (A-\lambda_i E)^{t-2} f_{p_2}. \end{array} \tag{74.2}$$

Con relación a la totalidad de los vectores $(A - \lambda_i E) f_1, \ldots$ $\ldots, (A - \lambda_i E) f_{p_1}, f_{p_1+1}, \ldots, f_{p_2}$ pueden demostrarse las mismas afirmaciones que se han demostrado respecto de la totalidad de los vectores $f_1, \ldots, f_{p_1}$, al sustituir, por supuesto, t por $t - 1$. Pasando del mismo modo a los subespacios $H_{t-2}, H_{t-3}, \ldots, H_1$, obtendremos un sistema linealmente independiente de k_i vectores pertenecientes al subespacio radical R_i. Las tablas del tipo (74.1), (74.2) terminan con una tabla de una línea

$$f_{p_{t-1}+1}, \ \ldots, \ f_{p_t}. \tag{74.3}$$

Estos vectores pertenecen a H_1, es decir, son propios, $p_t = m_1 - m_0$.

Dispondremos las tablas del tipo (74.1)—(74.3) sucesivamente de izquierda a derecha, nivelando sus últimas líneas e introduciendo las designaciones más compactas para todos los vectores. En este caso obtendremos la siguiente tabla:

$$\begin{array}{l} e_1^{(t)}, \ \ldots, \ e_{p_1}^{(t)}, \\ e_1^{(t-1)}, \ \ldots, \ e_{p_1}^{(t-1)}, \ e_{p_1+1}^{(t-1)}, \ \ldots, \ e_{p_2}^{(t-1)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ e_1^{(1)}, \ \ldots, \ e_{p_1}^{(1)}, \ e_{p_1+1}^{(1)}, \ \ldots, \ e_{p_2}^{(1)}, \ \ldots, e_{p_{t-1}+1}^{(1)}, \ \ldots, \ e_{p_t}^{(1)}. \end{array} \tag{74.4}$$

Los vectores que se hallan en la primera línea de esta tabla tienen la altura t, los vectores de la siguiente línea tienen la altura $t-1$, etc. Los vectores de la última línea tienen por altura 1, es decir, se transforman por el operador $A-\lambda_i E$ en un vector nulo. Cada columna de la tabla determina un subespacio invariante del operador $A-\lambda_i E$ y, por consiguiente, del operador A. Estos subespacios se llaman *cíclicos*. Los primeros p_1 subespacios cíclicos tienen dimensión t, los siguientes p_2-p_1 subespacios tienen una dimensión $t-1$, etc. Las últimas columnas determinan los subespacios cíclicos unidimensionales. *Todo el subespacio radical R_1 es la suma directa de los p_t subespacios cíclicos radicales.*

Escribamos la matriz de un operador inducido en un subespacio cíclico. Supongamos, por ejemplo, que como base se han tomado los vectores $e_1^{(1)}, e_1^{(2)}, \ldots, e_1^{(t-1)}, e_1^{(t)}$. Puesto que

$$A(-\lambda_i E)\, e_1^{(1)}=0, \quad (A-\lambda_i E)\, e_1^{(2)}=e_1^{(1)}, \ \ldots, \ (A-\lambda_i E)\, e_1^{(t)}=e_1^{(t-1)},$$

tenemos

$$Ae_1^{(1)}=\lambda_i e_1^{(1)}, \ Ae_1^{(2)}=\lambda_i e_1^{(2)}+e_1^{(1)}, \ \ldots, \ Ae_1^{(t)}=\lambda_i e_1^{(t)}+e_1^{(t-1)}.$$

Por consiguiente, la matriz del operador inducido tendrá la forma

$$\begin{pmatrix} \lambda_i & 1 & 0 \ldots 0 & 0 \\ 0 & \lambda_i & 1 \ldots 0 & 0 \\ \cdot & \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot \\ 0 & 0 & 0 \ldots \lambda_i & 1 \\ 0 & 0 & 0 \ldots 0 & \lambda_i \end{pmatrix}.$$

Las matrices de tipo semejante se denominan *cajas canónicas de Jordan*.

Construyamos ahora la base de un espacio como reunión sucesiva de las bases de los subespacios radicales $R_1, R_2, \ldots, R_r$. Como base de cada subespacio radical R_i tomemos los vectores del tipo (74.4) ordenados por turno de abajo a arriba y de izquierda a derecha. La base del subespacio construida de la manera indicada se denomina *base radical*.

En una base radical la matriz J del operador A adquiere la así llamada *forma canónica de Jordan*. Es una matriz casi diagonal compuesta de las cajas de Jordan. Primeramente van dispuestas las cajas de Jordan que corresponden al valor propio λ_1, con la particularidad de que sus dimensiones, según las cuales están ordenadas, no crecen. Luego, en el mismo orden se disponen las cajas de Jordan correspondientes a λ_2, etc. Así pues,

$$J = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_1 \end{matrix}} & & & & 0 \\ & \ddots & & & \\ & & \boxed{\lambda_1} & & \\ & & & \ddots & \\ & & & \boxed{\begin{matrix} \lambda_r & 1 & \dots & 0 \\ 0 & \lambda_r & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_r \end{matrix}} & \\ & & & \ddots & \\ 0 & & & & \boxed{\lambda_r} \end{pmatrix} \qquad (74.5)$$

Naturalmente, en el caso general algunas de las cajas de Jordan de órdenes inferiores pueden faltar.

La definición del operador en un espacio lineal determina la clase de matrices semejantes. El resultado obtenido significa que cualquier matriz cuadrada puede ser reducida, por medio de una transformación de semejanza, a la forma canónica de Jordan. Evidentemente, dos matrices cuadradas de un mismo orden son semejantes cuando, y sólo cuando, ambas tienen las formas de Jordan iguales. Por ello, con la base fijada:

Dos matrices cuadradas de un mismo orden determinan el mismo operador en un espacio complejo cuando, y sólo cuando, tienen iguales las formas de Jordan.

Ejercicios.

1. Sea x un vector radical de altura ν, correspondiente al valor propio λ_t del operador A. Demuéstrese que si λ_t es una raíz de multiplicidad p del polinomio $\varphi(z)$, el vector $v = \varphi(A)x$ será un vector radical de altura $r = \text{máx}\{0, \nu - p\}$ correspondiente al mismo valor propio λ_t. ¿Qué podrá decirse acerca del vector v, si λ_t no es la raíz del polinomio $\varphi(z)$?

2. Sea x un vector no nulo arbitrario y sea $\varphi(z)$ un polinomio de grado mínimo, para el cual $\varphi(A)x = 0$. Demuéstrese que $\varphi(z)$ es el divisor del polinomio característico del operador A.

3. Demuéstrese que toda matriz cuadrada puede ser reducida, salvo una permutación de las cajas de Jordan, a la forma canónica de Jordan de un tipo único.

4. Demuéstrese que si una matriz es semejante a la matriz J de (74.5), será también semejante a la matriz J'.

5. Demuéstrese que las matrices cuadradas A y A' son matrices de un mismo operador.
6. Sea J una matriz canónica de Jordan. ¿Qué forma tendrá la matriz J^p para los números enteros positivos p?

§ 75. Operador conjugado

Ahora pasamos a la investigación de los operadores lineales que actúan en un espacio unitario. Por supuesto, todos los resultados obtenidos anteriormente respecto de los operadores en un espacio complejo tienen lugar también en el caso dado. Por esta razón estudiaremos aquí sólo las propiedades adicionales de los operadores relacionadas con el concepto de ortogonalidad. En algunos casos consideraremos también unos operadores que actúan de un espacio unitario en otro espacio unitario. El papel principal en nuestras investigaciones lo desempeñará el así llamado operador conjugado.

Sean dados dos espacios unitarios X, Y. Un operador A^*, que actúa de Y en X, se llama *conjugado* del operador A, que actúa de X en Y, si para cualesquiera vectores $x \in X$, $y \in Y$ se verifica la igualdad

$$(Ax, y) = (x, A^*y). \tag{75.1}$$

TEOREMA 75.1. *Para todo operador lineal A existe un operador conjugado A^* que es, además, único.*

DEMOSTRACION. Elijamos en X una base ortonormalizada $e_1, e_2, \ldots, e_m$. Recordemos que para todo vector $x \in X$ tiene lugar la descomposición

$$x = \sum_{k=1}^{m} (x, e_k) e_k. \tag{75.2}$$

Si el operador A^* existe, entonces, de acuerdo con esta fórmula, para cualquier vector $y \in Y$ tenemos

$$A^*y = \sum_{k=1}^{m} (A^*y, e_k) e_k$$

o bien, tomando en consideración (75.1),

$$A^*y = \sum_{k=1}^{m} \overline{(e_k, A^*y)}\, e_k = \sum_{k=1}^{m} \overline{(Ae_k, y)}\, e_k = \sum_{k=1}^{m} (y, Ae_k)\, e_k. \tag{75.3}$$

Esto precisamente significa que si el operador A^* existe, es el único.

Ahora, convengamos en considerar la igualdad (75.3) como definición del operador A^*. Es fácil comprobar que el operador A^*, construido de tal modo, es lineal. El operador satisface también la igualdad (75.1). En efecto, teniendo en cuenta el carácter ortonormalizado del sistema $e_1, e_2, \ldots, e_m$ y tomando en consideración

(75.2), (75.3), obtenemos para cualesquiera vectores $x \in X$, $y \in Y$

$$(Ax, y) = \left(A \sum_{k=1}^{m} (x, e_k)\, e_k, y\right) = \sum_{k=1}^{m} (x, e_k)\, (Ae_k, y),$$

$$(x, A^*y) = \left(\sum_{k=1}^{m} (x, e_k)\, e_k, \sum_{k=1}^{m} (y, Ae_k)\, e_k\right) =$$

$$= \sum_{k=1}^{m} (x, e_k)\overline{(y, Ae_k)} = \sum_{k=1}^{m} (x, e_k)\, (Ae_k, y).$$

El teorema queda demostrado.

El operador conjugado A^* está relacionado con el operador A por ciertas correlaciones. He aquí algunas de éstas:

$$\begin{aligned} (A^*)^* &= A, \\ (A+A)^* &= A^* + B^*, \\ (\alpha A)^* &= \bar{\alpha} A^*, \\ (AB)^* &= B^* A^*, \\ (A^*)^{-1} &= (A^{-1})^*. \end{aligned} \tag{75.4}$$

La raya por encima de α significa aquí la conjugación compleja. Todas las correlaciones se demuestran siguiendo el mismo esquema. Por ello investiguemos detalladamente sólo las propiedades primera y segunda.

Consideraremos un operador arbitrario A y un operador A^*, conjugado de A. A su vez, para el operador A^*, el operador conjugado será $(A^*)^*$. Ahora, para cualesquiera $x \in X$, $y \in Y$ tenemos

$$(y, (A^*)^* x) = (A^*y, x) = \overline{(x, A^*y)} = \overline{(Ax, y)} = (y, Ax).$$

El primer miembro es igual al segundo para cualquier vector y Por consiguiente, $(A^*)^* x = Ax$. Pero, como la igualdad dada es verídica para todo vector x esto es testimonio de que $(A^*)^* = A$.

Supongamos ahora que el operador A actúa en el espacio X y es regular. Demostremos, al principio, que el operador A^* también es regular. Sea $A^*y = 0$. Conforme a (75.3), proviene que

$$\sum_{k=1}^{m} (y, Ae_k)\, e_k = 0.$$

El sistema de vectores $e_1, \ldots, e_m$ constituye una base, razón por la cual

$$(y, Ae_k) = 0 \tag{75.5}$$

para cualesquiera $k = 1, 2, \ldots, m$. El operador A es regular y, por lo tanto, toda base se transforma por él en otra base. Pero, en este caso el sistema de los vectores $Ae_1, \ldots, Ae_m$ será también una base y de (75.5) se deduce que $y = 0$. De este modo, el núcleo del operador A^* sólo contiene un vector nulo, es decir, este operador es regular.

Tomemos unos vectores arbitrarios $x, y \in X$. Existen los únicos vectores u, v, para los cuales

$$Au = x, \qquad A^*v = y.$$

Encontramos, luego,

$$(x, (A^{-1})^* y) = (A^{-1}x, y) = (u, A^*v) = (Au, v) = (x, (A^*)^{-1} y).$$

El primer miembro es igual al segundo para todo vector x. Por consiguiente, $(A^{-1})^* y = (A^*)^{-1} y$. Por ser y arbitrario, esto significa que $(A^{-1})^* = (A^*)^{-1}$.

Muchas propiedades conjuntas de los operadores A y A^* pueden establecerse en el proceso de investigación de las matrices de dichos operadores. Supongamos que en el espacio X se ha elegido una base ortonormalizada $e_1, e_2, \ldots, e_m$ y en el espacio Y, una base ortonormalizada $q_1, q_2, \ldots, q_n$. Si X coincide con Y, consideraremos coincidentes también sus bases. Supongamos que la matriz A_{qe} con los elementos a_{ij} corresponde al operador A. En este caso

$$Ae_j = \sum_{i=1}^{n} a_{ij} q_i.$$

Por esto, de acuerdo con (75.2) concluimos que

$$a_{ij} = (Ae_j, q_i). \tag{75.6}$$

Supongamos, luego, que al operador A^* en las mismas bases le corresponde la matriz A^*_{eq} con los elementos a^*_{ij}. Conforme a la fórmula (75.6),

$$a^*_{ij} = (A^*q_j, e_i)$$

Comparando los elementos a_{ij}, a^*_{ij} y tomando en consideración (75.1), encontramos

$$a^*_{ij} = (A^*q_j, e_i) = \overline{(e_i, A^*q_j)} = \overline{(Ae_i, q_j)} = \bar{a}_{ji}.$$

Esta fórmula nos permite enunciar la siguiente definición.

La matriz A^* de dimensiones $m \times n$ con los elementos a^*_{ij} se denomina *conjugada* de la matriz A de dimensiones $n \times m$ con los elementos a_{ij}, siempre que $a^*_{ij} = \bar{a}_{ji}$ para cualesquiera i, j.

De este modo, en cualesquiera bases ortonormalizadas, a los operadores conjugados les corresponden unas matrices conjugadas. Las matrices conjugadas satisfacen, evidentemente, todas las correlaciones (75.4). La matriz conjugada A^* está ligada con la matriz A por las operaciones de transposición y conjugación compleja. A saber,

$$A^* = (\overline{A'}) = (\bar{A})'. \tag{75.7}$$

Aquí la raya sobre la A y A' significa que todos los elementos de la matriz se sustituyen por los conjugados complejos.

El rango del operador coincide con el rango de su matriz. Por ello, de la fórmula (75.7) proviene que los operadores A y A^* tienen rangos iguales.

Designemos mediante $N \subset X$, $N^* \subset Y$ y $T \subset Y$, $T^* \subset X$ los núcleos y los campos de valores de los operadores A y A^*, respectivamente. Si $x \in N$, entonces $Ax = 0$ y $(x, A^*y) = 0$. Esto significa que el campo de valores del operador A^* es un subespacio ortogonal al núcleo del operador A. Naturalmente, el campo de valores del operador A es también ortogonal al núcleo del operador A^*. Por ser iguales las dimensiones de los subespacios T, T^* y las correlaciones (56.4), concluimos que

$$X = N \oplus T^*, \qquad Y = N^* \oplus T. \tag{75.8}$$

La base $y_1, y_2, \ldots, y_m$ del espacio unitario X se denomina *dual* con relación a la base $x_1, x_2, \ldots, x_m$ del mismo espacio, si se verifica

$$(x_i, y_j) = \begin{cases} 0 & \text{para } i \neq j, \\ 1 & \text{para } i = j. \end{cases}$$

La base dual se usa con frecuencia para la investigación de las propiedades conjuntas de los operadores A y A^* que actúan en un mismo espacio. Demostremos, al principio, que toda base dispone de una base dual que es, además, única. Sea $x_1, x_2, \ldots, x_m$ una base arbitraria. Para cualquier j el vector y_j debe ser ortogonal a los vectores $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_m$, y, por lo tanto, ortogonal a la cápsula lineal L_j construida sobre estos vectores. De aquí se deduce que el vector y_j se halla en un subespacio unidimensional $L_j^\perp$. La condición de normalización $(x_j, y_j) = 1$ lo determina de un modo único.

Es obvio que una base será dual con relación a sí misma cuando, y sólo cuando, sea ortonormalizada. La relación de dualidad de las bases es simétrica, por lo cual resulta razonable hablar de un par de bases mutuamente duales. Las bases mutuamente duales se llaman *biortonormalizadas*.

TEOREMA 75.2. *Si el operador A tiene en una base la matriz J, entonces, en la base dual con relación a la dada, el operador conjugado A^* tiene la matriz J^*.*

DEMOSTRACION. Supongamos que en la base ortonormalizada $e_1, e_2, \ldots, e_m$ los operadores A y A^* tienen las bases respectivas A_e y A_e^*, mientras que en la base $x_1, x_2, \ldots, x_m$ el operador A tiene la matriz J. Designemos con P la matriz de la transformación de coordenadas al pasar de la base $e_1, e_2, \ldots, e_m$ a la base $x_1, x_2, \ldots, x_m$. Entonces, de acuerdo con la fórmula (64.5), tenemos

$$J = P^{-1}A_eP.$$

Tomando una conjugación matricial de los miembros primero y segundo de esta igualdad, encontramos

$$J^* = P^* A_e^* (p^{-1})^*$$

o bien, lo que es igual,

$$J^* = ((P^{-1})^*)^{-1} A_e^* ((P^{-1})^*).$$

La correlación obtenida muestra que el operador conjugado A^* tiene la matriz J^* en la base $y_1, y_2, \ldots, y_m$, para la cual la matriz de la transformación de coordenadas, al pasar de la base $e_1, e_2, \ldots, e_m$, es $(P^{-1})^*$. Según la fórmula (63.3), las coordenadas de los vectores $x_1, x_2, \ldots, x_m$ en la base $e_1, e_2, \ldots, e_m$ son, en realidad, los elementos de las columnas de la matriz P; las coordenadas de los vectores $y_1, y_2, \ldots, y_m$ en la base $e_1, e_2, \ldots, e_m$ no son otra cosa que los elementos de las columnas de matriz $(P^{-1})^*$. El cálculo de los productos escalares de dos en dos de los vectores, pertenecientes a la base $x_1, x_2, \ldots, x_m$, y de los vectores, pertenecientes a la base $y_1, y_2, \ldots, y_m$, es equivalente al cálculo de los elementos de la matriz $P' \overline{(P^{-1})^*}$. Pero

$$P' \overline{(P^{-1})^*} = P' \overline{\overline{(P^{-1})'}} = P' (P^{-1})' = (P^{-1}P)' = E.$$

Por consiguiente, la base $y_1, y_2, \ldots, y_m$ es dual respecto a la base $x_1, x_2, \ldots, x_m$.

El teorema demostrado permite enunciar muchos corolarios. Si, por ejemplo, J es una matriz canónica de Jordan, entonces en su diagonal se disponen los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_m$. Mas, los valores propios de la matriz J^* son $\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m$. Por ello todos los valores propios del operador A^* son complejos conjugados respecto de los valores propios del operador A. Si el operador A es de estructura simple el teorema 75.2 permite afirmar que el operador conjugado A^* es también de estructura simple. En este caso los sistemas básicos de los vectores propios de los operadores A y A^* pueden elegirse de tal modo que sean biortonormalizados, etc.

Ejercicios.

1. Supongamos que las coordenadas de los vectores de cierta base de un espacio euclídeo, dadas en la base ortonormalizada $e_1, e_2, \ldots, e_m$, forman columnas de la matriz A. Demuéstrese que las coordenadas de los vectores de la base dual, dadas en la misma base $e_1, e_2, \ldots, e_m$, forman columnas de la matriz A^{-1}.
2. ¿De qué modo están ligados entre sí los polinomios característicos de los operadores A y A^*?
3. Demuéstrese que si un subespacio es invariante respecto del operador A, su complemento ortogonal es invariante respecto de A^*.
4. Demuéstrese que todo vector propio del operador A, correspondiente al valor propio λ, es ortogonal a cualquier vector propio del operador A^* correspondiente al valor propio $\mu \neq \bar{\lambda}$.

5. Demuéstrese que cualquier vector radical del operador A correspondiente al valor propio λ es ortogonal a todo vector radical del operador A^* correspondiente al valor propio $\mu \neq \lambda$.

§ 76. Operador normal

El hecho de que en un espacio existen una base ortonormalizada y una base formada por los vectores propios del operador lineal, es de mucha importancia para realizar las más diversas investigaciones. Por esto, nuestra tarea próxima consistirá en estudiar una clase de operadores que tienen en un espacio unitario sistemas básicos ortonormalizados compuestos de vectores propios. Operadores de tal índole existen a ciencia cierta. Por ejemplo, entre ellos figuran todos los operadores escalares.

TEOREMA 76.1. (*teorema de Schur*). *Para todo operador lineal en un espacio unitario existe una base ortonormalizada en la cual la matriz del operador es triangular.*

DEMOSTRACIÓN. Examinemos, por ejemplo, el caso de una matriz triangular derecha. De acuerdo con el teorema 72.1, para todo operador A existen los subespacios invariantes L_p, $p = 1, 2, \ldots, m$, tales que la dimensión de L_p es igual a p y cualquier subespacio de índice menor figura en todos los subespacios de índices mayores. La base buscada se construirá de la manera siguiente. A título de vector e_1 se tomará cualquier vector normalizado de L_1. A título de e_2 se tomará un vector normalizado de L_2, ortogonal al subespacio L_1, etc. A título de vector e_m se tomará un vector normalizado de L_m que sea ortogonal al subespacio L_{m-1}. La base $e_1, e_2, \ldots, e_m$ es ortonormalizada y, como se ha observado en el § 72, la matriz del operador en tal base es triangular derecha.

El operador lineal A se llama *normal*, si es permutable con su operador conjugado, es decir, si

$$AA^* = A^*A.$$

Probemos que los operadores normales, y sólo ellos, tienen en un espacio unitario sistemas básicos de vectores propios ortonormalizados.

Para el estudio de estos operadores resulta útil la siguiente observación. Si una matriz triangular es permutable con su matriz conjugada ésta es diagonal. En efecto, sea, por ejemplo, la matriz B de orden m triangular derecha y supongamos que $B^*B = BB^*$. Designemos mediante b_{ij} los elementos de la matriz B. La condición de que los elementos diagonales de la matriz $B^*B - BB^*$ son iguales a cero nos proporciona el siguiente sistema de ecuaciones respecto de los elementos no diagonales de la matriz B:

$$
\begin{aligned}
-|b_{12}|^2-|b_{13}|^2-|b_{14}|^2-\ldots-|b_{1m}|^2 &=0,\\
|b_{12}|^2-|b_{23}|^2-|b_{24}|^2-\ldots-|b_{2m}|^2 &=0,\\
|b_{13}|^2+|b_{23}|^2-|b_{34}|^2-\ldots-|b_{3m}|^2 &=0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\cdots\\
|b_{1m}|^2+|b_{2m}|^2+|b_{3m}|^2+\ldots+|b_{m-1,m}|^2 &=0.
\end{aligned}
$$

La única solución de este sistema es una solución nula, por lo que se demuestra precisamente la validez de la observación enunciada.

TEOREMA 76.2. *Para que un operador en un espacio unitario sea normal, es necesario y suficiente que tenga un sistema básico de vectores propios ortonormalizados.*

DEMOSTRACIÓN. Sea A un operador normal. Elijamos, de acuerdo con el teorema 76.1, una base ortonormalizada en la cual la matriz del operador sea triangular. En la misma base al operador A^* le corresponde una matriz triangular conjugada. Por hipótesis, el operador A es normal, por lo cual las matrices de los operadores A y A^* en la base elegida deben ser conmutables. Conforme a la observación enunciada anteriormente, estas matrices son diagonales. Así pues, hemos construido una base ortonormalizada en la cual la matriz del operador tiene una forma diagonal. Esto significa que la base dada está compuesta íntegramente de los vectores propios del operador.

Supongamos ahora que el operador A tiene un sistema básico de vectores propios ortonormalizados. En este caso la matriz del operador A en la base compuesta de los vectores citados será diagonal. Mas, en la misma base, al operador A^* le corresponde una matriz conjugada, la cual, evidentemente, será también diagonal. Las matrices diagonales son siempre conmutables y, por lo tanto, son conmutables también los operadores A y A^*.

En el transcurso de la demostración del teorema hemos probado que si el operador A es normal, entonces, en una base compuesta de vectores propios ortonormalizados no sólo la matriz del operador A, sino también la matriz del operador A^* será diagonal. Esto nos lleva al

COROLARIO. *Si el operador A es normal, todo sistema ortonormalizado de vectores propios del operador A es un sistema ortonormalizado de vectores propios del operador A^*, y viceversa.*

COROLARIO. *Si el operador A es normal, los valores propios de los operadores A y A^*, correspondientes al vector propio común, son complejos conjugados.*

Efectivamente, si $Ax=\lambda x$ y $A^*x=\mu x$, entonces, de acuerdo con (75.1), para todo vector propio normalizado x tendremos

$$\lambda=(\lambda x,\ x)=(Ax,\ x)=(x,\ A^*x)=(x,\ \mu x)=\bar{\mu}.$$

Este hecho es justo, por supuesto, para cualquier operador A que tiene unos vectores propios comunes con el operador A^*. El carácter normal del operador A garantiza *la presencia* de los vectores comunes.

La significación de los operadores normales en la teoría general se dicta por dos circunstancias. En primer lugar, esta clase de operadores es más simple en un espacio unitario. En segundo lugar, la investigación de un operador arbitrario normal se reduce frecuentemente a la investigación de los operadores normales.

Ejercicios.

1. Sea A un operador lineal arbitrario y α, β unos números complejos iguales en módulo. Demuéstrese que el operador $\alpha A + \beta A^*$ es normal.
2. Sea A un operador normal. Demuéstrese que para todo polinomio $\varphi(z)$ el operador $\varphi(A)$ será normal.
3. Demuéstrese que para un operador normal cualquier operador inducido será normal.
4. Demuéstrese que el operador A es normal cuando, y sólo cuando, para todo subespacio invariante L el complemento ortogonal $L^\perp$ es también invariante.
5. Sea A un operador de estructura simple en un espacio complejo. Demuéstrese que al prefijar de modo adecuado un producto escalar en un espacio, el operador A puede siempre hacerse normal.

§ 77. Operadores unitario y hermitiano

Entre los operadores normales son de mayor empleo los operadores de dos tipos: unitarios y hermitianos.

Un operador lineal U se llama *unitario*, si el operador conjugado U^* coincide con el inverso U^{-1}, es decir,

$$UU^* = U^*U = E.$$

TEOREMA 77.1. *Un operador normal U es unitario cuando, y sólo cuando, todos sus valores propios son iguales en módulo a la unidad.*

DEMOSTRACION. Sea U un operador unitario. Elijamos cualquiera de sus valores propios λ y un vector propio normalizado x que corresponde a λ. Tenemos

$$1 = (x, x) = (x, U^*Ux) = (Ux, Ux) = (\lambda x, \lambda x) = \lambda \cdot \bar{\lambda} (x, x) = |\lambda|^2.$$

Supongamos ahora que todos los valores propios del operador normal U son iguales en módulo a la unidad. Designemos mediante $x_1, \ldots, x_m$ los vectores propios ortonormalizados del operador U y mediante $\lambda_1, \ldots, \lambda_m$, sus valores propios. Por hipótesis, $|\lambda_i| = 1$ para todo i. Recordemos que para el operador conjugado U^* los vectores $x_1, \ldots, x_m$ siguen siendo propios, pero corresponden a los valores propios $\bar{\lambda}_1, \ldots, \bar{\lambda}_m$. Tomemos un vector arbitrario x y descompongámoslo según los vectores propios del operador U

$$x = \alpha_1 x_1 + \ldots + \alpha_m x_m.$$

Ahora calculamos

$$U^*Ux = U^*(Ux) = U^*(\alpha_1\lambda_1 x_1 + \ldots + \alpha_m\lambda_m x_m) =$$
$$= \alpha_1\lambda_1\bar{\lambda}_1 x_1 + \ldots + \alpha_m\lambda_m\bar{\lambda}_m x_m = \alpha_1 x_1 + \ldots + \alpha_m x_m = x.$$

Puesto que x es un vector arbitrario, esto significa que $U^*U = E$. De modo análogo se demuestra que $UU^* = E$.

TEOREMA 77.2. *El operador U es unitario cuando, y sólo cuando, para cualesquiera dos vectores el producto escalar de éstos es igual al producto escalar de sus imágenes.*

DEMOSTRACION. Sea U un operador unitario, entonces para cualesquiera dos vectores x, y tenemos

$$(x, y) = (x, U^*Uy) = (Ux, Uy). \tag{77.1}$$

Supongamos ahora que para cierto operador U se verifican las igualdades (77.1), cualesquiera que sean los vectores x, y. De aquí se deduce que

$$(x, (U^*U - E)\ y) = 0.$$

Puesto que los vectores x, y son arbitrarios, esto significa que $U^*U = E$. El operador U es regular, dado que en el caso contrario la igualdad $U^*U = E$ sería imposible. Por consiguiente, el operador U^{-1} existe. Al multiplicar la igualdad $U^*U = E$ a la izquierda por el operador U y a la derecha, por el operador U^{-1}, obtenemos otra igualdad: $UU^* = E$. De este modo, el operador U es unitario.

COROLARIO. *El operador U es unitario cuando, y sólo cuando, o bien $UU^* = E$ o bien $U^*U = E$.*

COROLARIO. *Todo operador unitario transforma cualquier sistema ortonormalizado de vectores en otro sistema, también ortonormalizado.*

COROLARIO. *Si el operador lineal U transforma una base ortonormalizada en otra base ortonormalizada, entonces U es un operador unitario.*

En efecto, sea $x_1, \ldots, x_m$ una base ortonormalizada, $Ux_i = y_i$ e $y_1, \ldots, y_m$, también una base ortonormalizada. Tomemos dos vectores arbitrarios x, y. Si

$$x = \sum_{i=1}^{m} \alpha_i x_i, \qquad y = \sum_{i=1}^{m} \beta_i x_i,$$

entonces

$$(x, y) = \sum_{i=1}^{m} \alpha_i \bar{\beta}_i.$$

Por ser lineal el operador U, tenemos

$$Ux = \sum_{i=1}^{m} \alpha_i y_i, \qquad Uy = \sum_{i=1}^{m} \beta_i y_i.$$

Por eso, tenemos nuevamente

$$(Ux, Uy) = \sum_{i=1}^{m} \alpha_i \beta_i.$$

Así pues, las igualdades (77.1) son válidas para cualesquiera vectores x, y.

Hemos de notar que podríamos definir el operador unitario como operador *isométrico*, es decir, un operador que conserva invariables las longitudes de todos los vectores. Esto proviene del teorema 77.2 y de la siguiente correlación, fácilmente comprobada:

$$(x, y) = \frac{|x+y|^2 - |x-y|^2 + i|x+iy|^2 - i|x-iy|^2}{4}.$$

Un operador lineal H se denomina *hermitiano* o *autoconjugado*, si coincide con su operador conjugado, es decir, si

$$H = H^*.$$

TEOREMA 77.3. *Un operador normal H es hermitiano cuando, y sólo cuando, todos sus valores propios son números reales.*

DEMOSTRACION. Sea H un operador hermitiano. Tomemos cualquier valor propio λ de este operador y un vector propio normalizado x que corresponde a λ. Tenemos

$$\lambda = (\lambda x, x) = (Hx, x) = (x, H^*x) = (x, Hx) = (x, \lambda x) = \bar{\lambda}.$$

es decir, λ es un número real. Supongamos ahora que el operador normal H tiene valores propios reales. Entonces, en la base compuesta de los vectores propios ortonormalizados del operador H las matrices de los operadores H y H^* coincidirán. Por consiguiente, son también coincidentes los operadores, es decir, H es un operador hermitiano.

El operador hermitiano H se llama *no negativo* (*definido positivo*) si para todo vector (no nulo) x se verifica la desigualdad

$$(Hx, x) \geqslant 0 \ (> 0).$$

TEOREMA 77.4. *El operador hermitiano H es no negativo (definido positivo) cuando, y sólo cuando, todos sus valores propios son no negativos (positivos).*

DEMOSTRACION. Elijamos una base ortonormalizada compuesta de los vectores propios $x_1, \ldots, x_m$ del operador hertmitiano H. En este caso, de la descomposición

$$x = \xi_1 x_1 + \ldots + \xi_m x_m$$

se desprende, para un vector arbitrario x, que

$$(Hx, x) = \lambda_1 |\xi_1|^2 + \ldots + \lambda_m |\xi_m|^2.$$

De aquí se deduce que si todos los valores propios del operador hermitiano son no negativos (positivos), entonces el mismo operador también es no negativo (definido positivo). Al hacer $x = x_i$, obtenemos

$$(Hx_i, x_i) = \lambda_i$$

para todo i. Por ello, todos los valores propios de un operador no negativo (definido positivo) son no negativos (positivos).

De lo dicho se infiere que un operador definido positivo es regular no negativo. Entre todos los operadores hermitianos los operadores no negativos y definidos positivos desempeñan un papel de especial importancia. He aquí algunas de sus propiedades.

Si H y S son unos operadores definidos positivos, el operador $\alpha H + \beta S$ será definido positivo para cualesquiera números no negativos α, β, no iguales a cero simultáneamente.

Efectivamente, el operador $\alpha H + \beta S$ es hermitiano, cualesquiera que sean los números reales α, β. Si estos números son negativos y no iguales a cero a la vez, entonces

$$((\alpha H + \beta S) x, x) = \alpha (Hx, x) + \beta (Sx, x) > 0$$

cuando $x \neq 0$.

Si el operador H es definido positivo, el operador H^{-1} será también definido positivo.

En efecto, como $H = H^*$, entonces $H^{-1} = (H^*)^{-1} = (H^{-1})^*$, es decir, el operador H^{-1} es hermitiano. Los valores propios del operador H^{-1} son magnitudes inversas respecto a los valores propios del operador H. Por ello son positivos y el operador H^{-1} es definido positivo.

Si H es definido positivo y A es un operador regular arbitrario, entonces los operadores A^*HA y AHA^* son definidos positivos.

Es fácil comprobar que estos operadores son hermitianos. En virtud de que el operador A es regular para cualquier $x \neq 0$, tendremos $Ax \neq 0$ y $A^*x \neq 0$. Por esta razón

$$(A^*HAx, x) = (HAx, Ax) > 0, \quad (AHA^*x, x) = (HA^*x, A^*x) > 0$$

para $x \neq 0$. De aquí se deduce, en particular, que para todo operador regular A los operadores A^*A y AA^* son definidos positivos. Si A es un operador degenerado, los operadores A^*A y AA^* son no negativos

Para todo operador no negativo H existe un operador no negativo S tal que $S^2 = H$.

En efecto, sean $\lambda_1, \ldots, \lambda_m$ los valores propios del operador H y $x_1, \ldots, x_m$, los vectores propios ortonormalizados correspondientes. En este caso $Hx_i = \lambda_i x_i$ para todo i. Definamos el operador S mediante las igualdades $Sx_i = \sqrt{\lambda_i} x_i$. El operador S es no negativo, puesto que tiene un sistema básico de vectores propios ortonormalizados $x_1, \ldots, x_m$ que corresponden a los valores propios no negativos $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}$. Además, $S^2x_i = Hx_i = \lambda_i x_i$. De este modo, los operadores S^2 y H coinciden sobre los vectores de la base $x_1, \ldots$

$\ldots, x_m$, a consecuencia de lo cual ellos coinciden también sobre todos los vectores, es decir, $S^2 = H$.

El operador no negativo S se denomina *raíz cuadrada aritmética* de un operador no negativo H, si $S^2 = H$.

Merece subrayar que todos los vectores propios de los operadores S y H coinciden. Efectivamente, supongamos que $\lambda_1, \ldots, \lambda_r$ y $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}$ son todos valores propios distintos de los operadores H y S, respectivamente. Designemos mediante X_i (Y_i), $i = 1, 2, \ldots, r$, el subespacio propio del operador H (S) que contiene todos los vectores propios correspondientes al valor propio λ_i $(\sqrt{\lambda_i})$. Las sumas directas de los subespacios propios $X_1, \ldots, X_r$ e $Y_1, \ldots, Y_r$ coinciden con todo el espacio. Por esto

$$\dim X_1 + \ldots + \dim X_r = \dim Y_1 + \ldots + \dim Y_r. \qquad (77.2)$$

Está claro que para todo i se tiene $Y_i \subset X_i$, es decir, $\dim Y_i \leqslant \dim X_i$. Por consiguiente, la igualdad (77.2) puede verificarse sólo cuando para cualquier i se cumpla la igualdad $\dim Y_i = \dim X_i$, es decir, cuando $Y_i = X_i$.

Así pues, los valores propios y los vectores propios del operador S se determinan unívocamente por el operador H. Puesto que S es un operador hermitiano, la raíz aritmética del operador H puede ser sólo *única*.

Ejercicios.

1. Demuéstrese que el conjunto de todos los operadores unitarios en un espacio unitario dado forma un grupo de multiplicación.
2. Demuéstrese que el conjunto de todos los operadores hermitianos en un espacio unitario dado forma un grupo de adición.
3. Supongamos que el operador A es hermitiano y el operador B es definido positivo. Demuéstrese que los valores propios de los operadores BA y $B^{-1}A$ son reales.
4. Demuéstrese que si A, B son unos operadores definidos positivos, todos los valores propios del operador BA son positivos.
5. Demuéstrese que si A, B son unos operadores definidos positivos conmutables, el operador BA será también definido positivo.
6. Demuéstrese que si A es un operador definido positivo en un espacio unitario, la función $(x, y)_A = (Ax, y)$ satisface todos los axiomas del producto escalar.

§ 78. Operadores A^*A y AA^*

Si el operador A actúa de un espacio unitario X en otro espacio unitario Y, entonces en X queda definido el operador A^*A y en Y, el operador AA^*. En lo que sigue estos operadores desempeñarán un papel considerable, por lo cual nos dedicaremos, ahora, a su estudio.

De las propiedades primera y cuarta (75.4) se deduce que los operadores A^*A y AA^* son hermitianos. Más aún, son no negativos, porque para cualesquiera vectores $x \in X$, $y \in Y$ tenemos

$$(A^*Ax,\ x) = (Ax,\ Ax) \geqslant 0$$

$$(AA^*y,\ y) = (A^*y,\ A^*y) \geqslant 0.$$

Por eso en el espacio X existe un operador no negativo G y en el espacio Y, un operador no negativo F tales que

$$A^*A = G^2, \qquad AA^* = F^2.$$

Los operadores G y F que satisfacen estas correlaciones son únicos.

Cualquiera que sea el operador A, el operador A^*A tiene un sistema *ortonormalizado* de los vectores propios $x_1, x_2, \ldots, x_m$. Este sistema se transforma siempre por el operador A en un sistema *ortogonal*. Efectivamente, sea

$$X^*Ax_k = \rho_k^2 x_k, \qquad \rho_k \geqslant 0 \tag{78.1}$$

para todo $k = 1, 2, \ldots, m$. Entonces

$$(Ax_k,\ Ax_l) = (A^*Ax_k,\ x_l) = \rho_k^2 (x_k,\ x_l) = 0$$

para $k \neq l$. Además, para todo k

$$|\,Ax_k\,| = \rho_k,$$

por lo cual el vector Ax_k es distinto del vector nulo cuando, y sólo cuando, el valor propio ρ_k^2 del operador A^*A no es nulo.

El vector *no nulo* Ax_k es un vector propio del operador AA^* y corresponde al valor propio ρ_k^2. En efecto, de acuerdo con (78.1)

$$AA^*(Ax_k) = A(A^*Ax_k) = A(\rho_k^2 x_k) = \rho_k^2 Ax_k.$$

De este modo, todos los valores propios no nulos del operador A^*A son valores propios del operador AA^*. Será cierta también, por supuesto, la afirmación inversa. Por esta razón los valores propios *no nulos* de los operadores A^*A y AA^* siempre *coinciden*.

Los valores propios de los operadores A^*A y AA^* se designarán mediante $\rho_1^2, \rho_2^2 \ldots$ En este caso puede considerarse, sin limitar la generalidad del razonamiento

$$\rho_1^2 \geqslant \rho_2^2 \geqslant \ldots \geqslant \rho_t^2 > 0,$$

mientras que los demás valores ρ_k^2 son iguales a cero. Es evidente que los valores propios de los operadores A^*A y AA^* se diferencian sólo en la multiplicidad del valor propio nulo. La multiplicidad del operador A^*A es $(m - t)$ y la del operador AA^*, $(n - t)$.

Se llaman *números singulares* (*principales*) *del operador* A los valores aritméticos de las raíces cuadradas de los valores propios comunes de los operadores A^*A y AA^*.

Haciendo uso de los vectores propios de los operadores A^*A y AA^*, se pueden construir en los espacios X e Y unas bases ortonormalizadas con ayuda de las cuales se describe y se investiga con facilidad la acción de los operadores A y A^*. Tomemos por base en el espacio X un sistema ortonormalizado $x_1, \ldots, x_m$ de vectores propios del operador A^*A. Según se deduce de (75.8), los vectores $x_1, \ldots, x_t$ forman una base en T^*, mientras que los vectores $x_{t+1}, \ldots, x_m$ forman una base en N. La base ortonormalizada $y_1, \ldots, y_n$ en el espacio Y se construirá de la manera siguiente. A título de $y_1, \ldots, y_t$ se tomarán los vectores obtenidos después de normalizar $Ax_1, \ldots, Ax_t$. Estos vectores forman una base en T. Por $y_{t+1}, \ldots, y_n$ se tomará cualquier base ortonormalizada en N^*. Está claro que los vectores $y_1, \ldots, y_n$ son propios para el operador AA^* y forman una base en Y. Tomando en consideración que $|Ax_k| = \rho_k$, deducimos que

$$Ax_k = \begin{cases} \rho_k y_k, & k \leqslant t, \\ 0, & k > t. \end{cases} \tag{78.2}$$

Al multiplicar estas igualdades por el operador A^* y teniendo en cuenta (78.1), obtenemos

$$A^* y_k = \begin{cases} \rho_k x_k, & k \leqslant t, \\ 0, & k > t. \end{cases} \tag{78.3}$$

Las bases ortonormalizadas en los espacios X, Y, ligadas con los operadores A, A^* mediante las correlaciones (78.2), (78.3), llevan el nombre de *bases singulares*.

Si los espacios X, Y son diferentes, en las bases singulares puede escribirse la matriz del operador A. Designémosla con Λ. Conforme a (78.2) esta matriz tiene la forma

$$\Lambda = \begin{pmatrix} \rho_1 & & & & & \\ & \rho_2 & & 0 & & \\ & & \ddots & & & \\ & & & \rho_t & & \\ & 0 & & & 0 & \\ & & & & & \ddots \end{pmatrix} \tag{78.4}$$

Si los espacios X, Y coinciden, las bases singulares, por regla general, no se usan, para la notación de la matriz del operador. Sin embargo, las correlaciones (78.2), (78.3) quedan en vigor.

Ejercicios.

1. Demuéstrese que los núcleos de los operadores A, A^*A (A^*, AA^*) coinciden, como también son coincidentes los campos de valores de los operadores A, AA^* (A^*, A^*A).
2. Demuéstrese que si $\dim X > \dim Y$ ($\dim X < \dim Y$), el operador A^*A (AA^*) es degenerado.
3. Demuéstrese que los números singulares no varían cuando el operador A se multiplica por cualesquiera operadores unitarios.

4. Supongamos que el operador A actúa en el espacio X y que todos sus números singulares son distintos dos a dos. Demuéstrese que las bases singulares se determinan unívocamente, salvo la multiplicación de cada uno de los vectores por un número que en módulo es igual a la unidad.

5. Demuéstrese que los números singulares de un operador normal coinciden con los módulos de los valores propios.

6. Demuéstrese que los números singulares del operador A^{-1} son inversos a los números singulares del operador A, mientras que las bases singulares de ambos operadores coinciden.

7. Supongamos que el operador A actúa en un espacio unitario m-dimensional X. Designemos mediante $\lambda_1, \ldots, \lambda_m$ sus valores propios y mediante $\rho_1, \ldots, \rho_m$, los números singulares. Demuéstrese que

$$\sum_{k=1}^{m} |\lambda_k|^2 \leqslant \sum_{k=1}^{m} \rho_k^2, \qquad \prod_{k=1}^{m} |\lambda_k| = \prod_{k=1}^{m} \rho_k.$$

8. Demuéstrese que si $|\lambda_k| = \rho_k$ para todo $k = 1, 2, \ldots, m$, entonces el operador es normal.

§ 79. Descomposiciones de un operador arbitrario

Una de las circunstancias que determina la significación de los operadores unitario y hermitiano consiste en la posibilidad de representar, sirviéndose de estos operadores, un operador lineal arbitrario.

Supongamos que un operador lineal arbitrario A actúa en el espacio unitario X. Mostremos que dicho operador siempre puede ser representado en la forma

$$A = H_1 + iH_2, \tag{79.1}$$

donde H_1 y H_2 son unos operadores hermitianos. Efectivamente, si esta descomposición existe, entonces

$$A^* = H_1 - iH_2.$$

Pero, en este caso

$$H_1 = \frac{1}{2}(A + A^*), \qquad H_2 = \frac{1}{2i}(A - A^*).$$

Las fórmulas obtenidas determinan precisamente la descomposición (79.1). Puesto que

$$H_1H_2 - H_2H_1 = \frac{1}{2i}(A^*A - AA^*),$$

entonces del hecho de que el operador A es normal proviene la conmutatividad de los operadores H_1, H_2, y viceversa.

Sea $x_1, \ldots, x_m$ un sistema ortonormalizado de vectores propios del operador A^*A. De acuerdo con (78.2), existe un sistema ortonormalizado $y_1, \ldots, y_m$ de vectores propios del operador AA^* tal que para todo k se verifica

$$Ax_k = \rho_k y_k. \tag{79.2}$$

Definamos ahora los operadores lineales F y U en el espacio X mediante las siguientes igualdades en los sistemas básicos de vectores:

$$Ux_k = y_k, \qquad Fy_k = \rho_k y_k. \tag{79.3}$$

Las correlaciones (79.2), (79.3) significan que se ha obtenido la descomposición

$$A = FU. \tag{79.4}$$

Aquí F es un operador hermitiano no negativo, puesto que tiene el sistema básico ortonormalizado de vectores propios $y_1, y_2, \ldots, y_m$ y los valores propios no negativos $\rho_1, \rho_2, \ldots, \rho_m$. El operador U es unitario, puesto que transforma el sistema ortonormalizado de vectores $x_1, x_2, \ldots, x_m$ en otro sistema ortonormalizado $y_1, y_2, \ldots, y_m$. Ha de ser notado que de (79.4) se infiere

$$AA^* = F^2, \tag{79.5}$$

es decir, F es la raíz cuadrada aritmética del operador AA^*.

La descomposición (79.4) se llama *descomposición polar* del operador A. Como que la raíz aritmética es única, el operador F en la descomposición polar será siempre único. El operador U será único sólo en el caso en que el operador A sea regular. En este caso $U = F^{-1}A$.

Otra vez observamos la relación directa existente entre el carácter normal del operador A y la conmutatividad de los componentes de una descomposición polar. Efectivamente, sea $UF = FU$ para cierto operador A, entonces

$$A^*A = U^*F^*FU = F^*U^*UF = F^2,$$

lo que, junto con (79.5), es testimonio de que el operador A es normal.

Supongamos ahora que el operador A es normal, es decir, $A^*A = AA^*$. De acuerdo con (79.4) tenemos $A = FU$. Por consiguiente, $A^* = U^*F$. La condición de que el operador es normal conduce a la igualdad $U^*F^2U = F^2$ o bien

$$F^2U = UF^2.$$

Tomando en consideración la segunda de las correlaciones (79.3), obtenemos

$$F^2(Uy_k) = \rho_k^2(Uy_k)$$

para todo $k = 1, 2, \ldots, m$, es decir, los vectores Uy_k son propios para el operador F^2. Como se ha observado anteriormente, los operadores F^2 y F tienen los mismos vectores propios, por lo cual

$$(FU)\, y_k = F(Uy_k) = \rho_k (Uy_k)$$

para todo $k = 1, 2, \ldots, m$. Por otra parte, conforme a la segunda de las correlaciones (79.3),

$$(UF)\, y_k = U\,(Fy_k) = U\,(\rho_k y_k) = \rho_k\,(Uy_k).$$

Las igualdades obtenidas muestran que los operadores FU y UF coinciden en el sistema básico de vectores $y_1, y_2, \ldots, y_m$. Por consiguiente, $UF = FU$.

Ejercicios.

1. Demuéstrese que si un operador es normal, los valores propios del operador H_1 (H_2) de (79.1) son las partes reales (imaginarias) de los valores propios del operador A.
2. Demuéstrese que si el operador A es normal, los valores propios del operador F (argumentos de los valores propios del operador U) de (79.4) son los módulos de los valores propios (argumentos de los valores propios no nulos) del operador A.
3. Demuéstrese que si el operador A es normal, entonces ambos operadores en la descomposición (79.1) tienen los mismos vectores propios que el operador A. ¿Qué puede decirse acerca de los vectores propios de los componentes de la descomposición (79.4)?

§ 80. Operadores en el espacio real

Al investigar los operadores lineales que actúan en un espacio real nos encontramos con algunas dificultades adicionales. Están relacionadas principalmente con el hecho de que no todo operador lineal tiene en el espacio real siquiera un solo vector propio.

Naturalmente, si el polinomio característico de un operador en un espacio real tiene sólo raíces reales, subsiste la analogía completa en la teoría. Varía, de hecho, sólo la terminología. A saber, las palabras «complejo, unitario, hermitiano» se sustituyen respectivamente por las palabras «real, ortogonal, simétrico». Si, en cambio, el polinomio característico tiene, además, unas raíces complejas, la investigación de tal operador se hace más complicada.

Sea dado un espacio real $\mathbf{R}$. Consideraremos un conjunto de toda clase de pares $(x;\, y)$ de vectores x, y, de $\mathbf{R}$. Definamos las operaciones sobre dichos pares. Convengamos en considerar que

$$(x;\, y) + (u;\, v) = (x + u;\, y + v)$$

para cualesquiera dos pares, y para el número complejo $\xi + i\eta$ y el par $(x;\, y)$

$$(\xi + i\eta)\,(x;\, y) = (\xi x - \eta y;\, \eta x + \xi y).$$

Es fácil comprobar que el conjunto de todos los pares de vectores pertenecientes a $\mathbf{R}$, después de realizadas las operaciones de la manera indicada, representa en sí un espacio complejo $\mathbf{C}$.

El espacio construido **C** es de la misma dimensión que el espacio **R**. En efecto, sea $e_1, e_2, \ldots, e_m$ una base en **R**. Para todo par de vectores u, v de **R** tenemos

$$\left.\begin{aligned} u &= \alpha_1 e_1 + \ldots + \alpha_m e_m, \\ v &= \beta_1 e_1 + \ldots + \beta_m e_m \end{aligned}\right\} \tag{80.1}$$

donde α_i, β_i son los números reales. Pero de aquí proviene que

$$(u;\ v) = \sum_{k=1}^{m} (\alpha_k + i\beta_k)(e_k;\ 0). \tag{80.2}$$

El sistema $(e_1;\ 0), \ldots, (e_m;\ 0)$ es linealmente independiente, por lo cual la dimensión del espacio **C** es igual a m.

Para toda base $e_1, \ldots, e_m$ en el espacio **R** y cualesquiera números reales $\alpha_1, \ldots, \alpha_m$ se cumple la igualdad

$$\sum_{k=1}^{m} (\alpha_k + i0)(e_k;\ 0) = \left(\sum_{k=1}^{m} \alpha_k e_k;\ 0\right).$$

Por consiguiente, entre todos los vectores u de **R** y todos los pares del tipo $(u;\ 0)$ de **C** existe una correspondencia biunívoca. Más aún, esta correspondencia es un isomorfismo, si nos limitamos a las operaciones con números reales.

Si todos los pares del tipo $(u;\ 0)$ se identifican con los propios vectores u de **R**, entonces de (80.1), (80.2) se desprende que el espacio **C** puede considererse como un conjunto de elementos

$$w = u + iv,$$

donde $u, v \in$ **R**. En este caso se debe recordar, por supuesto, que en realidad los elementos u, v son los pares $(u;\ 0)$, $(v;\ 0)$ y la multiplicación por el número i y la adición se realizan conforme a las definiciones introducidas anteriormente. Cuando $v = 0$, obtenemos los elementos del espacio **R**. Es natural considerar este espacio como cierto conjunto de **C**. Los elementos del tipo $u + i0$ se llamarán *reales*, mientras que los elementos del tipo $u + iv$ y $u - iv$ se denominan *complejos conjugados*.

El espacio **C** se llama *extensión compleja del espacio real* **R**.

En la resolución de los diversos problemas en un espacio euclídeo podemos extender dicho espacio, de modo análogo, hasta obtener un espacio unitario. Examinemos la extensión compleja **C** del espacio euclídeo **R**. Para cualesquiera dos vectores

$$z = x + iy, \qquad w = u + iv$$

de **C**, convengamos en considerar, por definición, que

$$(z,\ w) = ((x,\ u) + (y,\ v)) + i\,((y,\ u) - (x,\ v)).$$

No es difícil establecer que el espacio **C** dotado de tal producto escalar se hace unitario. Además, el producto escalar para cualesquiera dos vectores de **R** se conserva.

Supongamos que el operador A actúa en el espacio **R**. Construyamos un operador nuevo $\tilde{A}$ que actúa en el espacio **C** y coincide con el operador A sobre los vectores de **R**. Para esto hagamos

$$\tilde{A}(u+iv)=Au+iAv.$$

Está claro que el operador $\tilde{A}$ es lineal y $\tilde{A}u = Au$ para todos los vectores $u \in \mathbf{R}$.

El operador $\tilde{A}$ lleva el nombre de *extensión del operador A en el espacio complejo* **C**.

Ahora, en vez de estudiar el operador A en el espacio real **R** podemos considerar el operador $\tilde{A}$ en un espacio complejo **C** e investigar su operación en **R**, considerando este último como conjunto del espacio **C**. Esto es un procedimiento más usado, si para alguna situación en el espacio complejo no existe analogía correspondiente en el espacio real.

Supongamos que en el espacio complejo **C** se ha elegido una base real. En esta base la matriz del operador dilatado $\tilde{A}$ será real y coincidirá con la matriz del operador A en la misma base. De aquí se infiere que el polinomio característico del operador $\tilde{A}$ coincide con el polinomio característico del operador A y, por consiguiente, tiene coeficientes reales. Es evidente que

Si el polinomio característico del operador A que actúa en un espacio real **R** *tiene raíz real, esta última es un valor propio del operador $\tilde{A}$ y le corresponde al menos un vector propio real.*

Consideraremos ahora una raíz compleja λ del polinomio característico del operador A. Es un valor propio del operador $\tilde{A}$ y le corresponde cierto vector propio w. Dado que los coeficientes del polinomio característico del operador $\tilde{A}$ son reales, el operador citado tendrá también un valor propio conjugado complejo $\bar{\lambda}$. El operador $\tilde{A}$ transforma vectores conjugados complejos en vectores conjugados complejos, por lo cual de $\tilde{A}w = \lambda w$ proviene $\tilde{A}\bar{w} = \bar{\lambda}\bar{w}$. Por consiguiente, a los valores propios conjugados complejos del operador $\tilde{A}$ les corresponden unos vectores conjugados complejos.

Si $\lambda \neq \bar{\lambda}$, entonces los vectores w, $\bar{w}$ serán linealmente independientes como vectores propios correspondientes a los diferentes valores propios.

Consideraremos los vectores x, y que se determinan del modo siguiente en términos de w, $\bar{w}$:

$$x=\frac{1}{2}(w+\bar{w}), \qquad y=\frac{1}{2i}(w-\bar{w}). \tag{80.3}$$

Es fácil comprobar que dichos vectores son reales. Además, no es difícil establecer que si $\lambda = \mu + i\nu$, entonces

$$Ax = \mu x - \nu y, \qquad Ay = \nu x + \mu y.$$

Por ello, una cápsula lineal construida sobre los vectores (80.3) en el espacio R es un subespacio invariante del operador A. La matriz del operador inducido sobre este subespacio en la base (80.3) es la siguiente:

$$\begin{pmatrix} \mu & \nu \\ -\nu & \mu \end{pmatrix}.$$

Por consiguiente, el polinomio característico del operador inducido es igual a $(z - \mu)^2 + \nu^2$ o bien, lo que es igual, $z^2 - (\lambda + \bar{\lambda})\, z + \lambda\bar{\lambda}$. Observemos que en el subespacio invariante construido el operador A no tiene ningún vector propio cuando $\nu \neq 0$. De este modo, hemos obtenido una deducción importante. A saber:

Si un polinomio característico del operador A, *que actúa en el espacio real* R, *tiene una raíz compleja* (*¡no real!*), *a esta raíz le corresponde en el espacio* R *un subespacio invariante bidimensional del operador* A *que no contiene vectores propios.*

Esta deducción juega el mismo papel para investigar los operadores en un espacio real que desempeña la existencia por lo menos de un único vector propio para la investigación de los operadores en un espacio complejo. Eligiendo de un modo adecuado las bases en el espacio R, se puede reducir la matriz del operador a una forma semejante, en cierto sentido, o bien a la diagonal, o bien a la triangular, o bien a la forma canónica de Jordan. Tal procedimiento de investigación del operador es de uso relativamente raro, puesto que las formas canónicas reales están privadas de muchas ventajas que poseen las formas canónicas complejas. Es mucho más fácil y fructífera la investigación de la dilatación de un operador en un espacio complejo.

Ejercicios.

1. Demuéstrese que el campo de valores (el núcleo) del operador $\hat{A}$ es la dilatación compleja del campo de valores (del núcleo) del operador A.

2. Supongamos que el operador dilatado $\hat{A}$ es de estructura simple. Demuéstrese que en el espacio R se puede elegir tal base en la que la matriz del operador A sea de forma casi diagonal con las matrices en la diagonal de primero y segundo órdenes.

3. Demuéstrese que en el espacio real R de dimensión m todo operador tiene un subespacio invariante de dimensión $m - 1$ ó $m - 2$.

4. ¿Que análogo tiene en un espacio real el teorema 72.1?

5. Demuéstrese que todo operador lineal que actúa en un espacio real de dimensión impar tiene por lo menos un vector propio.

§ 81. Matrices de tipo especial

Hemos considerado algunos operadores de tipo especial. Será natural suponer que las matrices de dichos operadores deben poseer ciertos rasgos específicos.

Una matriz compleja cuadrada U se llama *unitaria*, si la matriz conjugada U^* coincide con la inversa U^{-1}, es decir, si

$$UU^* = U^*U = E.$$

Recordemos que en una base ortonormalizada al operador conjugado le corresponde una matriz conjugada. Por consiguiente, la matriz de un operador unitario en la base ortonormalizada es unitaria.

Sean dadas en un espacio unitario dos bases ortonormalizadas cualesquiera. Construyamos la matriz de la transformación de coordenadas al pasar de una de estas bases a la otra. De acuerdo con la fórmula (63.3), las columnas de la matriz están compuestas por las coordenadas que tienen los vectores de la segunda base respecto de la primera. Mas la misma forma tiene también la matriz del operador lineal que transforma los vectores de la primera base en los de la segunda. Conforme al segundo corolario del teorema 77.2, este operador es unitario. Por esto

La matriz de la transformación de coordenadas al pasar de una base ortonormalizada a otra base ortonormalizada es unitaria.

Llamaremos dos matrices *semejantes unitarias*, si son semejantes y la matriz de la transformación de semejanza es unitaria. De las propiedades del operador unitario se deduce que toda matriz unitaria es semejante unitaria respecto de una matriz diagonal de elementos diagonales que, en módulo, son iguales a la unidad.

Se escriben con facilidad las correlaciones que definen los elementos de la matriz unitaria. Supongamos que la matriz U es de orden m. Designemos mediante u_{ij} sus elementos. Entonces, de la igualdad $UU^* = E$ se infiere que

$$\sum_{k=1}^{m} u_{ik}\bar{u}_{jk} = \begin{cases} 0, & \text{si} \quad i \neq j, \\ 1, & \text{si} \quad i = j. \end{cases}$$

Análogamente, de la igualdad $U^*U = E$ obtenemos:

$$\sum_{k=1}^{m} u_{ki}\bar{u}_{kj} = \begin{cases} 0, & \text{si} \quad i \neq j, \\ 1, & \text{si} \quad i = j, \end{cases}$$

De este modo, los sistemas de vectores columna y vectores fila de cualquier matriz unitaria representan en sí sistemas ortonormalizados.

La matriz unitaria real U se denomina *ortogonal*. Esta matriz se determina por las siguientes correlaciones:

$$UU' = U'U = E.$$

Todas las propiedades de las matrices ortogonales se deducen de las propiedades de las matrices unitarias.

Una matriz compleja cuadrada H se llama *hermitiana* o *autoconjugada*, si coincide con su inversa, es decir, si

$$H = H^*.$$

De este modo, la matriz de un operador hermitiano en la base ortonormalizada es hermitiana. De las propiedades del operador hermitiano se desprende que cualquier matriz hermitiana es semejante unitaria respecto a la matriz diagonal real. Si h_{ij} son los elementos de una matriz hermitiana H, entonces

$$h_{ij} = \overline{h}_{ji}$$

para cualesquiera i, j. De aquí obtenemos, en particular, que los elementos diagonales de cualquier matriz hermitiana son reales.

Una matriz hermitiana real H se denomina *simétrica*. Esta matriz se determina por la correlación siguiente:

$$H = H'.$$

Observemos que toda matriz simétrica es semejante ortogonal respecto de la matriz diagonal real. Una matriz cuadrada se llama *normal*, si es conmutable con su inversa.

De acuerdo con esta definición, una matriz de un operador normal en la base ortonormalizada es normal. Teniendo presente las propiedades del operador normal, es fácil comprender que toda matriz normal compleja es semejante unitaria respecto de la matriz diagonal.

Las matrices de tipo especial son de mucha importancia en la construcción de los más diversos algoritmos de cálculo. No obstante, no serán objeto de estudio detallado. Todas las propiedades de estas matrices son, de hecho, la reflexión de las propiedades análogas de los operadores correspondientes.

Ejercicios.

1. Demuéstrese que toda matriz compleja es semejante unitaria respecto de la matriz triangular.

2. Sean $\lambda_1, \lambda_2, \ldots, \lambda_m$ los valores propios de la matriz A, con la particularidad de que cada valor propio se ha escrito tantas veces cual es su multiplicidad. Demuéstrese que

$$\sum_{i=1}^{m} |\lambda_i|^2 \leqslant tr\,(A^*A). \tag{81.1}$$

3. Demuéstrese que la igualdad en la correlación (81.1) tiene lugar cuando y sólo cuando, la matriz A es normal.

4. Haciendo uso de la fórmula de Binet—Cauchy, demuéstrese que para cualquier matriz A los menores principales de la matriz A^*A son no negativos.

5. Demuéstrese que la suma de cuadrados de los módulos de todos los menores de una matriz unitaria dispuestos en cualesquiera filas o columnas fijadas es igual a la unidad.

6. Demuéstrese que toda matriz rectangular A puede ser representada en la forma $A = Q \Lambda S$, donde Q, S son unas matrices unitarias y Λ es una matriz diagonal de elementos no negativos.

CAPÍTULO 10 PROPIEDADES MÉTRICAS DEL OPERADOR

§ 82. Continuidad y acotación del operador

Hemos introducido el concepto de operador lineal como cierta generalización del concepto de función. Si suponemos que en los espacios se ha definido cierta métrica, podemos observar la analogía con la acotación de una función, la continuidad de una función, etc. Al estudiar estos problemas, partiremos siempre de que un operador actúa del espacio normalizado m-dimensional X en el espacio normalizado n-dimensional Y. Si X no coincide con Y, las normas en ambos espacios pueden ser introducidas independientemente la una de la otra.

Un operador A que actúa de X en Y se denomina *continuo en el punto* $x_0 \in X$, si de la condición $x_k \to x_0$ se desprende que $Ax_k \to Ax_0$ para cualquier sucesión $\{x_k\}$ de X. Si el operador es continuo en todo punto del espacio X, se llama *continuo en todo punto* o, simplemente, *continuo*.

TEOREMA 82.1. *Un operador lineal que actúa en unos espacios normalizados arbitrarios de dimensión finita es continuo.*

DEMOSTRACIÓN. Tomemos un vector arbitrario $x_0 \in X$ y elijamos en X una base cualquiera $e_1, e_2, \ldots, e_m$. Tenemos

$$x_0 = \xi_1^{(0)} e_1 + \ldots + \xi_m^{(0)} e_m.$$

Supongamos que $x_k \to x_0$ y

$$x_k = \xi_1^{(k)} e_1 + \ldots + \xi_m^{(k)} e_m.$$

En conformidad con el teorema 53.1, de la convergencia en norma proviene la convergencia de coordenadas. Por ello, $\xi_s^{(k)} \to \xi_s^{(0)}$ para todo s. Pero

$$Ax_0 = \xi_1^{(0)} Ae_1 + \ldots + \xi_m^{(0)} Ae_m$$

y, además,

$$Ax_k = \xi_1^{(k)} Ae_1 + \ldots + \xi_m^{(k)} Ae_m.$$

Ahora, de la convergencia $\xi_s^{(k)} \to \xi_s^{(0)}$ para todos los s se deducirá la convergencia $Ax_k \to Ax_0$ en norma del espacio Y.

El operador A se llama *acotado*, si existe una constante M tal que $\| Ax \| \leqslant M \| x \|$ para todo vector $x \in X$.

TEOREMA 82.2 *Un operador lineal que actúa en unos espacios normalizados arbitrarios de dimensión finita es acotado.*

DEMOSTRACIÓN. Supongamos que en cierto caso el operador A no es acotado. Entonces, existe una sucesión de vectores no nulos $\{x_k\}$ tal que

$$\| Ax_k \| \geqslant k \| x_k \|.$$

Examinemos una sucesión de vectores

$$y_k = \frac{1}{k \| x_k \|} x_k.$$

Esta sucesión converge a cero, puesto que

$$\| y_k \| = \frac{1}{k \| x_k \|} \| x_k \| = \frac{1}{k} \to 0.$$

Por otra parte,

$$\| Ay_k \| = \frac{1}{k \| x_k \|} \| Ax_k \| \geqslant 1.$$

Esto es testimonio de que la sucesión $\{Ay_k\}$ no converge a cero, es decir, el operador A no es continuo en cero. La contradicción obtenida con el teorema 82.1 da por terminada la demostración.

Resulta natural plantear la cuestión acerca de la constante *mínima* de todas las constantes M que satisfacen la condición $\| Ax \| \leqslant M \| x \|$ para todos los vectores x. Puesto que el conjunto de estas constantes está acotado inferiormente por cero, la constante mínima existe a ciencia cierta. Se llama *norma del operador A* y se designa con el símbolo $\| A \|$. Por definición, la norma de un operador posee las siguientes dos propiedades:

1) para todo vector x del espacio X es válida la desigualdad

$$\| Ax \| \leqslant \| A \| \cdot \| x \|, \tag{82.1}$$

2) para todo número $\varepsilon > 0$ existe tal vector $x_\varepsilon \in X$ que

$$\| Ax_\varepsilon \| \geqslant (\| A \| - \varepsilon) \| x_\varepsilon \| \tag{82.2}$$

Demostremos que

$$\| A \| = \sup_{\| x \| \leqslant 1} \| Ax \| \tag{82.3}$$

o bien, que es lo mismo,

$$\| A \| = \sup_{x \neq 0} \frac{\| Ax \|}{\| x \|}; \tag{82.4}$$

si, desde luego, $\dim X > 0$.

Tomemos un vector arbitrario x que satisface la desigualdad $\| x \| \leqslant 1$. Entonces, de (82.1) se deduce que

$$\| Ax \| \leqslant \| A \| \| x \| \leqslant \| A \|.$$

Por consiguiente,

$$\sup_{\|x\|\leqslant 1} \|Ax\| \leqslant \|A\|. \tag{82.5}$$

Luego, tomemos, de acuerdo con (82.2), un vector cualquiera x_ε y construyamos el vector

$$y_\varepsilon = \frac{1}{\|x_\varepsilon\|} x_\varepsilon.$$

En este caso

$$\|Ay_\varepsilon\| = \frac{1}{\|x_\varepsilon\|} \|Ax_\varepsilon\| \geqslant \frac{1}{\|x_\varepsilon\|} (\|A\| - \varepsilon) \|x_\varepsilon\| = \|A\| - \varepsilon.$$

Como $\|y_\varepsilon\| = 1$, resulta

$$\sup_{\|x\|\leqslant 1} \|Ax\| \geqslant \|Ay_\varepsilon\| \geqslant \|A\| - \varepsilon.$$

Por ser ε arbitrario, se obtiene que

$$\sup_{\|x\|\leqslant 1} \|Ax\| \geqslant \|A\|. \tag{82.6}$$

Ahora, de (82.5), (82.6) se desprende la correlación (82.3) que se trataba de establecer.

Mostraremos a continuación que *la norma de un operador desempeña un papel excepcionalmente importante al introducir una métrica en los espacios de operadores lineales. En ese caso será esencial que la norma del operador tenga una forma explícita* (82.3).

Ejercicios.

1. Demuéstrese que en un conjunto cerrado acotado de vectores se alcanzan las cotas superior e inferior de las normas de los valores del operador lineal.
2. Demuéstrese que un operador lineal transforma todo conjunto cerrado acotado en otro conjunto cerrado acotado.
3. ¿Será cierta la afirmación del ejercicio anterior, si no se requiere la acotación del conjunto?
4. Demuéstrese que en la fórmula (82.3) la cota superior se alcanza en un conjunto de vectores que satisfacen la condición $\|x\| = 1$, siempre que $\dim X > 0$.
5. Supongamos que un operador A actúa en el espacio X. Demuéstrese que A es regular, si, y sólo si, existe tal número $m > 0$ que $\|Ax\| \geqslant m \|x\|$ para todo $x \in X$.

§ 83. Norma del operador

Un conjunto ω_{XY} de operadores lineales que actúan de X en Y es un espacio lineal de dimensión finita. Si este espacio es lineal o complejo, se lo puede transformar en un espacio métrico completo, introduciendo en el primero, de tal o cual manera, una norma.

La introducción de la norma en un espacio de operadores lineales se efectúa mediante los mismos procedimientos que se usan en cualquier otro espacio lineal. No obstante, en el caso dado el mayor interés lo representan sólo aquellas normas en ω_{XY} que están relacionadas, de una manera suficientemente estrecha, con las normas en los espacios X, Y. Una de las clases más importantes de normas de este género la constituyen las llamadas normas concordadas.

Si para todo operador de ω_{XY} se verifica la desigualdad

$$\| Ax \| \leqslant \| A \| \cdot \| x \|,$$

cualquiera que sea $x \in X$, la norma de los operadores se denomina *concordada* con las normas vectoriales en los espacios X, Y.

La ventaja de las normas concordadas se ve claramente en el siguiente ejemplo. Supongamos que λ es el valor propio del operador A que actúa en el espacio X, y x, un vector propio correspondiente a λ. En este caso $Ax = \lambda x$, y, por lo tanto,

$$| \lambda | \cdot \| x \| = \| \lambda x \| = \| Ax \| \leqslant \| A \| \cdot \| x \|.$$

Por consiguiente, $| \lambda | \leqslant \| A \|$. Así pues, hemos obtenido una deducción muy importante:

Los módulos de los valores propios de un operador lineal no son superiores a cualquiera de sus normas concordadas.

El ejemplo citado muestra que para obtener las mejores estimaciones, es deseable emplear la menor de las normas concordadas. Está claro que todas las normas concordadas están acotadas inferiormente por la expresión (82.3). Si probamos que está expresión satisface los axiomas de la norma, ella será precisamente la menor de las normas concordadas. De este modo *justificaremos* tanto la denominación de la expresión (82.3), como su designación.

Evidentemente, para cualquier operador A la expresión $\| A \|$ es no negativa. Si $\| A \| = 0$, es decir, si

$$\sup_{\| x \| \leqslant 1} \| Ax \| = 0,$$

entonces $\| Ax \| = 0$ para todos los vectores x cuya norma no es superior a la unidad. Pero, en este caso, en virtud de la linealidad del operador, $Ax = 0$ para cualquier x. Por consiguiente $A = 0$. Para todo operador A y un número λ se tiene

$$\| \lambda A \| = \sup_{\| x \| \leqslant 1} \| \lambda Ax \| = | \lambda | \sup_{\| x \| \leqslant 1} \| Ax \| = | \lambda | \cdot \| A \|.$$

Y, por fin, para cualesquiera dos operadores A, B, de ω_{XY}

$$\| A + B \| = \sup_{\| x \| \leqslant 1} \| Ax + Bx \| \leqslant \sup_{\| x \| \leqslant 1} (\| Ax \| + \| Bx \|) \leqslant$$

$$\leqslant \sup_{\| x \| \leqslant 1} \| Ax \| + \sup_{\| x \| \leqslant 1} \| Bx \| = \| A \| + \| B \|.$$

Todas estas correlaciones significan que la expresión (82.3) representa en sí una norma en el espacio de operadores lineales. La norma (82.3) se denomina norma del operador *subordinada* a las normas vectoriales en los espacios X, Y.

La norma subordinada posee también una propiedad muy importante con relación a la operación de multiplicación de los operadores. Supongamos que el operador A actúa de X en Y y el operador B, de Y en Z. Entonces, como se sabe, queda definido el operador BA. Teniendo presente la concordancia de las normas subordinadas, encontramos

$$\|BA\| = \sup_{\|x\|\leqslant 1} \|(BA)x\| = \sup_{\|x\|\leqslant 1} \|B(Ax)\| \leqslant$$
$$\leqslant \sup_{\|x\|\leqslant 1} (\|B\|\cdot\|Ax\|) = \|B\| \sup_{\|x\|\leqslant 1} \|Ax\| = \|B\|\cdot\|A\|.$$

De este modo, cualquier norma subordinada del operador posee las siguientes cuatro propiedades principales. Para cualesquiera operadores A, B y todo número λ

$$\begin{aligned}
&1)\ \|A\| > 0, \text{ si } A \neq 0;\ \|0\| = 0,\\
&2)\ \|\lambda A\| = |\lambda|\,\|A\|,\\
&3)\ \|A + B\| \leqslant \|A\| + \|B\|,\\
&4)\ \|BA\| \leqslant \|B\|\cdot\|A\|.
\end{aligned} \tag{83.1}$$

Como propiedad adicional indiquemos que para un operador idéntico E es válida la igualdad

5) $\|E\| = 1$.

Esta se desprende de (82.3), puesto que $Ex = x$ para todo vector x.

En el caso general, la norma subordinada de un operador depende tanto de la norma en el espacio X, como también de la norma en el espacio Y. Si los dos espacios son unitarios, a título de norma en ellos puede servir la longitud de los vectores. La correspondiente norma subordinada del operador se denomina norma *espectral* y se designa con el símbolo $\|\cdot\|_2$. Así pues, para todo operador A que actúa de X en Y se tiene

$$\|A\|_2^2 = \sup_{(x,\,x)\leqslant 1} (Ax,\ Ax). \tag{83.2}$$

Investiguemos algunas propiedades de la norma espectral.

La norma espectral no varía, cuando el operador se multiplica por cualesquiera operadores unitarios.

Sean V, U los operadores unitarios arbitrarios que actúan en los espacios X, Y, respectivamente. Consideraremos el operador $B = UAV$. Tenemos

$$\|B\|_2^2 = \sup_{(x,\,x)\leqslant 1} (Bx,\ Bx) = \sup_{(x,\,x)\leqslant 1} (UAVx,\ UAVx) =$$
$$= \sup_{(x,\,x)\leqslant 1} (AVx,\ U^*UAVx) = \sup_{(x,\,x)\leqslant 1} (AVx,\ AVx) =$$
$$= \sup_{(Vx,\,Vx)\leqslant 1} (AVx,\ AVx) = \sup_{(v,\,v)\leqslant 1} (Av,\ Av) = \|A\|_2^2,$$

La definición de una norma espectral en la forma de (83.2) permite establecer su relación con los números singulares del operador A. Sean $x_1, x_2, \ldots, x_m$ un sistema ortonormalizado de los vectores propios del operador A^*A y sean $\rho_1^2, \rho_2^2, \ldots, \rho_m^2$ los valores propios de dicho operador. Sin limitar la generalidad de los razonamientos supogamos que

$$\rho_1 \geqslant \rho_2 \geqslant \ldots \geqslant \rho_m \geqslant 0. \tag{83.3}$$

Representemos el vector $x \in X$ en forma de la descomposición

$$x = \alpha_1 x_1 + \ldots + \alpha_m x_m, \tag{83.4}$$

entonces

$$(x, x) = \sum_{i=1}^{m} |\alpha_i|^2.$$

Según se ha observado en el § 78, el sistema $x_1, x_2, \ldots, x_m$ se transforma por el operador A en un sistema ortogonal, siendo en este caso

$$(Ax_i, Ax_i) = \rho_i^2$$

para todo i. Por consiguiente,

$$(Ax, Ax) = \sum_{i=1}^{m} |\alpha_i|^2 \rho_i^2,$$

que da

$$\|A\|_2^2 = \sup_{\sum_{i=1}^{m} |\alpha_i|^2 \leqslant 1} \sum_{i=1}^{m} |\alpha_i|^2 \rho_i^2. \tag{83.5}$$

Está claro que bajo las condiciones (83.3)

$$\|A\|_2^2 \leqslant \rho_1^2.$$

Pero, para el vector x_1 el segundo miembro de (83.5) toma el valor ρ_1^2. Por ello

$$\|A\|_2^2 = \rho_1^2.$$

De este modo,

La norma espectral del operador A es igual al número singular máximo.

Recordemos que para un operador normal A los números singulares coinciden con los módulos de los valores propios. Por consiguiente, la norma espectral del operador unitario es igual a la unidad, la norma espectral de un operador no negativo es igual al valor propio máximo.

Ejercicios.

1. Demuéstrese que para cualquier valor propio λ del operador A se verifica la desigualdad

$$|\lambda| \leqslant \inf_k \|A^k\|^{1/k}.$$

2. Sea $\varphi(z)$ cualquier polinomio con coeficientes no negativos. Demuéstrese que

$$\| \varphi(A) \| \leqslant \varphi(\| A \|).$$

3. Demuéstrese que $\| A \| \geqslant \| A^{-1} \|^{-1}$ para todo operador regular A. ¿Cuándo tendrá lugar la igualdad para el caso de una norma espectral?

§ 84. Normas matriciales del operador

La norma espectral es, en esencia, la única norma subordinada del operador el cálculo de la cual no está relacionado explícitamente con las bases. En cambio, si en los espacios con operadores dados se han fijado algunas bases, entonces la posibilidad para introducir las normas operacionales se hace mucho más amplia.

Así pues, consideraremos una vez más los operadores lineales que actúan del espacio X en el espacio Y. Supongamos que en X está fijada la base $e_1, e_2, \ldots, e_m$ y en Y, la base $q_1, q_2, \ldots, q_n$. Descompongamos un vector arbitrario $x \in X$ según la base y obtendremos

$$x = x_1 e_1 + \ldots + x_m e_m. \tag{84.1}$$

Ahora, la norma en el espacio X puede introducirse, por ejemplo, de acuerdo con la fórmula (52.3) o por cualquier otro método, sirviéndose de los coeficientes de la descomposición. De modo análogo puede ser introducida también una norma en el espacio Y.

Las más usadas son las normas del tipo (52.4). Por ello, investigaremos las normas de los operadores subordinadas y concordadas precisamente con las del tipo (52.4). Más aún, convengamos en considerar que en ambos espacios X e Y se han introducido las normas de un mismo tipo. Es evidente que las normas correspondientes del operador A han de ser ligadas de tal o cual manera con los elementos a_{ij} de la matriz del operador en las bases elegidas.

Demos a conocer, al principio, las expresiones para las normas de los operadores subordinadas a las 1-normas y ∞-normas de (52.4). Tenemos

$$\| A \|_1 = \sup_{\|x\|_1 \leqslant 1} \| Ax \|_1 = \sup_{\|x\|_1 \leqslant 1} \Big(\sum_{i=1}^{n} \Big| \sum_{j=1}^{m} a_{ij} x_j \Big| \Big) \leqslant$$

$$\leqslant \sup_{\|x\|_1 \leqslant 1} \Big(\sum_{i=1}^{n} \sum_{j=1}^{m} | a_{ij} | \, | x_j | \Big) < \sup_{\|x\|_1 \leqslant 1} \Big(\sum_{j=1}^{m} | x_j | \sum_{i=1}^{n} | a_{ij} | \Big) \leqslant$$

$$\leqslant \Big(\max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} | a_{ij} | \Big) \Big(\sup_{\|x\| \leqslant 1} \| x \|_1 \Big) = \max_{1 \leqslant j \leqslant m} \sum_{i=1}^{n} | a_{ij} |.$$

Mostremos ahora que para cierto vector x que satisface la condición $\| x_1 \| \leqslant 1$, $\| Ax \|_1$ coincide con el segundo miembro de la correlación obtenida.

Supongamos que el valor máximo en el segundo miembro se alcanza cuando $j = l$. En este caso todas las desigualdades se convierten en igualdades, por ejemplo, para $x = e_l$. Así pues,

$$\| A \|_1 = \max_{1 \leq j \leq m} | a_{ij} |.$$

Análogamente se investiga también la otra norma:

$$\| A \|_\infty = \sup_{\|x\|_\infty \leq 1} \| Ax \|_\infty = \sup_{\|x\|_\infty \leq 1} \left(\max_{1 \leq i \leq n} \left| \sum_{j=1}^{m} a_{ij} x_j \right| \right) \leq$$

$$\leq \sup_{\|x\|_\infty \leq 1} \left(\max_{1 \leq i \leq n} \sum_{j=1}^{m} | a_{ij} | \, | x_j | \right) \leq$$

$$\leq \sup_{\|x\|_\infty \leq 1} \left(\left(\max_{1 \leq i \leq n} \sum_{j=1}^{m} | a_{ij} | \right) \left(\max_{1 \leq j \leq m} | x_j | \right) \right) =$$

$$= \left(\max_{1 \leq i \leq n} \sum_{j=1}^{m} | a_{ij} | \right) \left(\sup_{\|x\|_\infty \leq 1} \| x \|_\infty \right) = \max_{1 \leq i \leq n} \sum_{j=1}^{m} | a_{ij} |.$$

Supongamos que el valor máximo en el segundo miembro se alcanza para $i = l$. Tomemos un vector x cuyas coordenadas son $x_j = | a_{lj} | / a_{lj}$, si $a_{lj} \neq 0$ y $x_j = 1$, si $a_{lj} = 0$. No es difícil comprobar que para dicho vector todas las desigualdades se convierten en igualdades. Por consiguiente,

$$\| A \|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{m} | a_{ij} |.$$

Con el objeto de hallar la norma del operador subordinada a las 2-normas de (52.4) procedamos de la manera siguiente. Por analogía con (32.1), introduzcamos un producto escalar en los espacios X, Y. Entonces la 2-norma de (52.4) coincidirá con la longitud del vector. Por ello, la norma subordinada no es otra cosa que la norma espectral del operador, correspondiente al producto escalar dado. Elegidos los productos escalares, las bases se convierten en ortonormalizadas, razón por la cual en estas bases al operador conjugado le corresponderá una matriz conjugada. Al designar mediante A_{qe} la matriz del operador A, obtendremos de lo dicho la siguiente deducción.

La norma de un operador subordinada a las 2-normas es igual al número singular máximo de la matriz A_{qe}.

Las normas examinadas son ciertas funciones de la matriz del operador. Mediante un procedimiento semejante pueden construirse no sólo las normas subordinadas, sino también las concordadas. Una de las más importantes normas concordadas es la llamada norma *euclidiana*. Designarémosla con el símbolo $\|\cdot\|_E$. Si el operador A tiene en las bases elegidas la matriz A_{qe} con elementos a_{ij}, entonces, por definición,

$$\| A \|_E = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} | a_{ij} |^2\right)^{1/2}.$$

El segundo miembro de esta expresión es una norma en el espacio $n \times m$-dimensional de operadores lineales. Por ello, el cumplimiento de las primeras tres propiedades de (83.1) no causa duda alguna. Es de mucha importancia el hecho de que para la norma euclidiana se cumple también la cuarta propiedad de (83.1). Para demostrar esto haremos uso de la desigualdad de Cauchy — Buniakovski del tipo (27.5).

Consideraremos los espacios lineales X, Y, Z de dimensiones respectivas m, n, p. Supongamos que el operador A actúa de X en Y, y el operador B, de Y en Z. Designemos mediante a_{ij}, b_{ij} los elementos de las matrices de estos operadores en las bases elegidas. Tenemos

$$\| BA \|_E = \left(\sum_{i=1}^{p} \sum_{j=1}^{m} \left| \sum_{k=1}^{m} b_{ik} a_{kj} \right|^2\right)^{1/2} \leqslant \left(\sum_{i=1}^{p} \sum_{j=1}^{m} \left(\sum_{k=1}^{n} | b_{ik} | \, | a_{kj} | \right)^2\right)^{1/2} \leqslant$$

$$\leqslant \left(\sum_{i=1}^{p} \sum_{j=1}^{m} \left(\sum_{k=1}^{n} | b_{ik} |^2 \right) \left(\sum_{i=1}^{n} | a_{ij} |^2 \right)\right)^{1/2} =$$

$$= \left(\left(\sum_{i=1}^{p} \sum_{k=1}^{n} | b_{ik} |^2\right) \left(\sum_{i=1}^{n} \sum_{j=1}^{m} | a_{ij} |^2 \right)\right)^{1/2} = \| B \|_E \cdot \| A \|_E.$$

En el caso general la norma euclidiana no es subordinada. El hecho de que ella esté compatible con las 2-normas se demuestra del mismo modo que el utilizado para demostrar la propiedad que acabamos de considerar.

La comprobación directa permite establecer una fórmula de importancia para la norma euclidiana. A saber,

$$\| A \|_E = \mathrm{tr}\,(A_{qe}^* A_{qe}) = \mathrm{tr}\,(A_{qe} A_{qe}^*). \tag{84.2}$$

Ahora podemos enunciar las siguientes deducciones.

A una matriz conjugada en las bases ortonormalizadas le corresponde un operador conjugado. Al introducir en los espacios X, Y los productos escalares por analogía con (32.1), convertiremos las bases elegidas en las ortonormalizadas. Dado que la traza de una matriz es igual a la suma de sus valores propios de (84.2) se deduce que:

El cuadrado de la norma euclidiana de un operador es igual a la suma de los cuadrados de sus números singulares.

Al introducir productos escalares en X, Y podemos hablar de los operadores unitarios. Con referencia a estos operadores unitarios resulta fácil mostrar que:

La norma euclidiana no varía al multiplicar el operador por cualesquiera operadores unitarios.

En efecto, como ya se ha indicado en los ejercicios del § 78, los números singulares no cambian cuando se multiplican por los operadores unitarios, mientras que la norma euclidiana se expresa solamente en términos de los números singulares.

En la mayoría de las aplicaciones relacionadas con las normas resulta importante no tanto la definición explícita de la norma del operador, como el cumplimiento de las propiedades (83.1). Por esta razón la norma de un operador puede definirse *axiomáticamente* por intermedio de su matriz. Elijamos en los espacios, donde están fijados unos operadores, algunas bases; entonces a todo operador corresponderá cierta matriz. A toda matriz pondremos en correspondencia un número denotado por el símbolo $\|\cdot\|$ y supongamos que en este caso quedan cumplidas como axiomas las condiciones (83.1). El número $\|\cdot\|$ se llamará *norma de la matriz*. Si ahora a todo operador se le pone en correspondencia la norma de su matriz, será obvio que de este modo en el espacio de operadores se introduce una norma. Es evidente que las condiciones (83.1) se cumplen también para los operadores. Lo recíproco es también cierto. Toda norma del operador engendra, con bases fijadas, una norma de la matriz. Estas normas de las matrices se designarán mediante símbolos análogos $\|\cdot\|_2$, $\|\cdot\|_\infty$, etc. Por lo visto, la concordancia de la norma también puede exigirse axiomáticamente.

Los ejemplos examinados muestran que la realización práctica de la definición axiomática de una norma del operador en términos de la norma de la matriz es posible. En lo que sigue, al hablar de las normas de las matrices y de los operadores, *siempre* supondremos su concordancia y el cumplimiento de las condiciones (83.1).

Ejercicios.

1. Demuéstrese que con toda norma para la matriz unidad se verifica la desigualdad

$$\| E \| \geqslant 1. \tag{84.3}$$

2. Sean $\lambda_1, \ldots, \lambda_m$ los valores propios de la matriz A. Demuéstrese que

$$\inf_B \| B^{-1}AB \|_E = \sum_{k=1}^{m} | \lambda_k |^2.$$

Compárese esta igualdad con (81.1).

§ 85. Ecuaciones operacionales

Uno de los más importantes problemas del álgebra consiste en la resolución de los sistemas de ecuaciones algebraicas lineales. Ya nos hemos encontrado más de una vez con este problema en el transcurso de nuestra narración. Ahora lo consideraremos desde el punto de vista de la teoría de operadores lineales.

Supongamos que está dado un sistema (60.2) con elementos del campo P de números reales o complejos. Tomemos un espacio m-dimensional X y un espacio n-dimensional Y sobre un mismo campo P y fijemos en ellos unas bases cualesquiera. En este caso las correlaciones (60.2) serán equivalentes a una igualdad matricial del tipo (61.2) y la última es, a su vez, equivalente a la igualdad operacional

$$Ax = y. \tag{85.1}$$

Aquí, A es un operador que actúa de X en Y y tiene en las bases elegidas la misma matriz de que se dispone el sistema (60.2). Las coordenadas de los vectores $x \in X$ e $y \in Y$ en las bases elegidas son, respectivamente, $(\xi_1, \ldots, \xi_m)$ y $(\eta_1, \ldots, \eta_n)$.

De este modo, en lugar de los sistemas de ecuaciones algebraicas lineales podemos considerar las ecuaciones (85.1). El problema consiste en hallar todos los vectores $x \in X$ que para el operador A y el vector $y \in Y$ dados satisfacen la igualdad (85.1). Una ecuación del tipo (85.1) se denomina *operacional*, el vector y lleva el nombre de *segundo miembro* y *el vector x es la solución*. Desde luego, todas las propiedades de los sistemas de ecuaciones se extienden automáticamente a las ecuaciones operacionales, y viceversa.

El teorema de Kronecker—Capelli enuncia la condición necesaria y suficiente para que el sistema sea resoluble en términos del rango de la matriz. Esto no es muy cómodo, porque no permite observar una relación profunda que existe entre los sistemas y las ecuaciones de otros tipos.

Sean X, Y unos espacios unitarios, entonces queda definido el operador A^*. La ecuación (85.1) se llamará *ecuación no homogénea* fundamental y la ecuación

$$A^*u = v,$$

ecuación no homogénea conjugada. Si los segundos miembros son nulos, las ecuaciones correspondientes se denominarán *homogéneas*. Resulta lícita la siguiente afirmación:

O bien la ecuación no homogénea fundamental tiene solución, cualquiera que sea el segundo miembro, o bien la ecuación homogénea conjugada tiene por lo menos una solución no nula.

En efecto, designemos con r el rango del operador A. El mismo rango tendrá también el operador A^*. Pueden observarse dos casos: o bien $r = n$, o bien $r < n$. En el primer caso el campo de valores del operador A tiene dimensión n y, por lo tanto, coincide con el espacio Y. Por eso la ecuación no homogénea fundamental debe tener solución, cualquiera que sea el segundo miembro. En el mismo caso el defecto del operador conjugado es igual a cero, razón por la cual el núcleo no tiene vectores no nulos, es decir, la ecuación homogénea conjugada no tiene soluciones no nulas. Si $r < n$, el campo de valores del operador A no coincide con Y y la ecuación no homogénea fundamental no puede tener solución, cualquiera que sea el segundo miembro. En esta circunstancia el núcleo del operador conjugado se compone no sólo del vector nulo, a consecuencia de lo cual la ecuación conjugada homogénea tiene soluciones no nulas.

La afirmación demostrada es de significación especial cuando los espacios X, Y coinciden. En este caso la existencia de una solución de la ecuación no homogénea fundamental, cualquiera que sea el segundo miembro, significa que el operador A es regular. Por ello, en el caso dado es válida la así llamada

ALTERNATIVA DE FREDHOLM. *O bien la ecuación no homogénea fundamental tiene siempre solución y, además, única, cualquiera que sea el segundo miembro o bien la ecuación homogénea conjugada tiene por lo menos una solución no nula.*

TEOREMA DE FREDHOLM. *Para que la ecuación no homogénea fundamental sea resoluble, es necesario y suficiente que su segundo miembro sea ortogonal a todas las soluciones de la ecuación homogénea conjugada.*

DEMOSTRACION. Designamos con N^* el núcleo del operador A^* y mediante T, el campo de valores del operador A. Si la ecuación no homogénea fundamental es resoluble, entonces el segundo miembro $y \in T$. De acuerdo con (75.8), se infiere que $y \perp N^*$, es decir, $(y, u) = 0$ para todos los vectores u que satisfacen la ecuación $A^*u = 0$. Sea ahora $(y, u) = 0$ para los mismos vectores u, entonces $y \perp N^*$ y, conforme a (75.8), $y \in T$. Mas, esto significa que existe un vector $x \in X$ tal que $Ax = y$, es decir, la ecuación no homogénea fundamental es resoluble.

Ejercicios.

1. Demuéstrese que la ecuación $A^*Ax = A^*y$ es resoluble.
2. Demuéstrese que la ecuación $(A^*A)^p x = (A^*A)^q y$ es resoluble para cualesquiera p, q enteros y positivos.
3. Establézcase el sentido geométrico de la alternativa y del teorema de Fredholm.

§ 86. Seudosoluciones y un operador seudoinverso

La definición arbitraria del operador A y del miembro segundo y puede conducir a que la ecuación (85.1) no tenga ninguna solución. Evidentemente, esto se debe sólo a *qué se entiende exactamente por solución de una ecuación.*

Tomemos un vector arbitrario $x \in X$ y consideremos el vector $r = Ax - y$, llamado *residuo* del vector x. Para que x sea la solución de la ecuación (85.1), es necesario y suficiente que su residuo sea nulo. A su vez, para que un residuo sea nulo, es necesario y suficiente que sea nula su longitud. De este modo, todas las soluciones de la ecuación (85.1), y sólo ellas, satisfacen la igualdad

$$|Ax - y|^2 = 0.$$

Dado que el valor nulo de la longitud del residuo es el mínimo, la determinación de las soluciones de la ecuación (85.1) puede considerarse como el problema de la búsqueda de los vectores x, para los cuales la expresión

$$\Phi_0(x) = |Ax - y|^2 \qquad (86.1)$$

alcanza su valor mínimo. El segundo miembro de esta expresión se denomina *funcional del residuo*. La búsqueda de los vectores que minimizan la funcional del residuo tiene también sentido en el caso cuando las soluciones de la ecuación (85.1) no existan. Esto sirve de base para la siguiente definición.

Se llama *seudosolución* (o *solución generalizada*) de la ecuación (85.1) todo vector $x \in X$, para el cual la funcional del residuo alcanza su valor mínimo. La seudosolución de longitud mínima recibe el nombre de seudosolución *normal*.

Mostremos que la seudosolución normal existe siempre y es, además, única. En los espacios X, Y fijemos las bases singulares $x_1, \ldots, x_m$ e $y_1, \ldots, y_n$. Sea

$$x = \sum_{k=1}^{m} \alpha_k x_k, \qquad y = \sum_{p=1}^{n} \beta_p y_p. \qquad (86.2)$$

Tomando en consideración las correlaciones (78.2), obtenemos

$$Ax - y = \sum_{k=1}^{m} \rho_k \alpha_k y_k - \sum_{p=1}^{n} \beta_p y_p.$$

Convengamos en considerar, como hasta ahora, que los números singulares $\rho_1, \ldots, \rho_t$ son distintos de cero, mientras que los números restantes son nulos. Puesto que las bases singulares son ortonormalizadas, tenemos

$$\Phi_0(x) = \sum_{k=1}^{t} |\rho_k \alpha_k - \beta_k|^2 + \sum_{p=t+1}^{n} |\beta_p|^2.$$

Es evidente que el valor mínimo de la funcional del residuo se conseguirá para aquellos vectores x, en los cuales las últimas $m - t$ coordenadas α_k son arbitrarias, y las primeras t coordenadas se determinan por la fórmula

$$\alpha_k = \beta_k / \rho_k. \qquad (86.3)$$

La seudosolución normal será

$$x_0 = \sum_{k=1}^{t} \frac{\beta_k}{\rho_k} x_k. \qquad (86.4)$$

Recordemos que los vectores $x_{t+1}, \ldots, x_m$ forman la base del núcleo N del operador A. Por eso el conjunto de todas las seudosoluciones representa en sí un plano en el espacio X cuyo subespacio director coincide con el núcleo N y el vector de desplazamiento coincide con cualquier seudosolución. *La solución seudonormal es el único vector, ortogonal a N, de este plano.*

Haciendo uso de las correlaciones (78.2), (78.3), es fácil mostrar que las seudosoluciones, y sólo ellas, satisfacen la ecuación

$$A^*Ax = A^*y. \tag{86.5}$$

En efecto, escribamos los vectores x, y en forma de las descomposiciones (86.2). Tenemos

$$A^*Ax = \sum_{k=1}^{t} \rho_k^2 \alpha_k x_k, \quad A^*y = \sum_{p=1}^{t} \rho_p \beta_p x_p.$$

De aquí se deduce que las soluciones de la ecuación (86.5) serán los vectores x, y sólo ellos, para los cuales las primeras t coordenadas α_k se calculan según (86.3) y las últimas $m - t$ coordenadas son arbitrarias.

De este modo, *si la resolubilidad de la ecuación* (85.1) *no es garantizada, siempre podemos sustituir la resolución de dicha ecuación por la resolución de la ecuación* (86.5). *En este caso se asegura la minimización de la funcional del residuo para la ecuación* (85.1).

El operador inverso juega un papel importante en muchas investigaciones. No obstante, fue definido sólo para un operador regular y por ahora no disponemos del análogo correspondiente para un operador degenerado y un operador que actúa de un espacio en otro. Dicho análogo puede ser construido sobre la base de seudosoluciones.

Supongamos que el operador A actúa del espacio X en el espacio Y. En este caso, a todo vector $y \in Y$ podemos ponerle en correspondencia un único vector $x_0 \in X$ que representa la seudosolución normal de la ecuación (85.1). Dicha correspondencia determina cierto operador A^+, el cual actúa de Y en X y lleva el nombre de operador *seudoinverso* (o *inverso generalizado*) del operador A. Por consiguiente, según la definición,

$$x_0 = A^+y \tag{86.6}$$

para cualquier $y \in Y$. Está claro que si el operador A es regular, el operador *seudoinverso coincide con el inverso*. Investiguemos las propiedades del operador seudoinverso.

Supongamos que a la par con (86.6) disponemos de $u_0 = A^+v$ para cierto vector $v \in Y$. Consideremos un vector $\alpha y + \beta v$ para cualesquiera números α, β. Al tomarlo en calidad de segundo miembro de la ecuación (85.1), el vector $\alpha x_0 + \beta u_0$ satisfará, a ciencia cierta, una ecuación correspondiente del tipo (86.5) y por esta razón será una seudosolución. Puesto que x_0, u_0 son ortogonales al núcleo del operador A, será también ortogonal al núcleo del vector $\alpha x_0 + \beta u_0$. Por lo tanto, dicho vector es la seudosolución normal. De este modo, *la linealidad* del operador seudoinverso queda establecida.

Las propiedades del operador seudoinverso pueden ser fácilmente establecidas, si consideramos su acción sobre los vectores de las bases singulares. De acuerdo con (86.4), tenemos

$$A^+y_k = \begin{cases} \rho_k^{-1}x_k, & k \leqslant t, \\ 0, & k > t. \end{cases} \tag{86.7}$$

De aquí se desprende que:

El campo de definición, el núcleo y el campo de valores de los operadores seudoinverso y conjugado coinciden.

Con ayuda de las fórmulas (78.2), (78.3), (86.7) se pueden obtener diferentes correlaciones que relacionan los operadores A, A^*, A^+. He aquí algunas de ellas:

1) $(A^*)^+ = (A^+)^*$,

2) $(A^+)^+ = A$,

3) $(AA^+)^* = AA^+$, $\quad (AA^+)^2 = AA^+$,

4) $(A^+A)^* = A^+A$, $\quad (A^+A)^2 = A^+A$,

5) $AA^+A = A$.

Estas correlaciones se demuestran según el mismo esquema, razón por la cual a título de ejemplo consideraremos detalladamente sólo las correlaciones primera y tercera.

Comparando (78.2) y (86.7) elijamos, en calidad del operador A, el operador conjugado A^*. Puesto que para este último se verifica (78.3), resulta

$$(A^*)^+ x_k = \begin{cases} \rho_k^{-1}y_k, & k \leqslant t, \\ 0, & k > t. \end{cases}$$

Ahora, partiendo de (86.7), apliquemos una correlación, análoga a (78.3), al operador $(A^+)^*$. En este caso

$$(A^+)^* x_k = \begin{cases} \rho_k^{-1}y_k, & k \leqslant t, \\ 0, & k > t. \end{cases}$$

De este modo, los operadores $(A^*)^+$ y $(A^+)^*$ coinciden sobre la base $x_1, \ldots, x_m$, y, por tanto, son iguales.

Al tomar en consideración (78.2), (86.7), concluimos que para el operador AA^+ son válidas las correlaciones

$$AA^+y_k = \begin{cases} 1 \cdot y_k, & k \leqslant t, \\ 0, & k > t. \end{cases} \tag{86.8}$$

Esto significa que el operador AA^+ tiene un sistema ortonormalizado de vectores propios $y_1, \ldots, y_n$, así como también los valores propios reales-1 y 0, es decir, es hermitiano. De este modo queda establecida la primera de las correlaciones del tercer grupo. La segunda igualdad se desprende, evidentemente, de (86.8).

Ejercicios.

1. ¿Qué representa en sí un operador que es seudoinverso respecto al operador nulo?

2. Supongamos que los espacios X, Y son distintos. Escríbase la matriz de un operador seudoinverso en las bases singulares y compárese ésta con (78.4).

3. Sean U, V unos operadores unitarios que actúan en los espacios X e Y, respectivamente. Demuéstrese que

$$(VAU)^+ = U^*A^+V^*.$$

4. Demuéstrese que existen tales operadores K en X y L en Y que se verifica

$$A^+ = KA^* = A^*L.$$

Descríbase la acción de los operadores K, L.

5. Demuéstrese que un operador seudoinverso se define unívocamente por las condiciones

$$AA^+A = A,$$
$$A^+ = KA^* = A^*L.$$

6. Demuéstrese que todas las seudosoluciones, y sólo ellas, sirven de soluciones para la ecuación

$$Ax = AA^+y.$$

7. Establézcase el sentido geométrico de las seudosoluciones.

§ 87. Perturbación y regularidad del operador

Hemos subrayado más de una vez que una variación pequeña de una base, las coordenadas de un vector o de los elementos de una matriz, etc. puede tener por resultado el cambio de varias propiedades relacionadas con la noción de dependencia lineal. Esta noción desempeña un papel decisivo en toda la teoría de operadores lineales, a consecuencia de lo cual resulta muy importante investigar *la influencia de la pequeña variación* de los operadores en las propiedades de los mismos.

En la resolución de los más diversos problemas hemos de utilizar, como un medio auxiliar, un operador *próximo al operador idéntico*. Por este término se entenderá un operador que actúa en el espacio X y tiene la forma $E + A$, donde $\| A \| < 1$ para una de las normas.

Si λ es un valor propio cualquiera del operador A, entonces $1 + \lambda$ será un valor propio del operador $E + A$. Dado que $|\lambda| \leqslant \leqslant \| A \|$, entonces, en virtud de la condición $\| A \| < 1$, todos los valores propios del operador A son inferiores en módulo a la unidad Por consiguiente, todos los valores propios del operador $E + A$ son distintos de cero y este operador será regular.

De este modo, *al cumplirse la condición* $\|A\| < 1$, *existe el operador* $(E + A)^{-1}$. *En cambio, si el operador* $E + A$ *es degenerado, entonces* $\|A\| \geqslant 1$ *para cualquier norma.*

Para cualquier número α, cuyo módulo es inferior a la unidad, es válida la correlación límite

$$(1+\alpha)^{-1} = \lim_{p\to\infty} \alpha_p$$

donde

$$\alpha_p = \sum_{k=0}^{p} (-\alpha)^k.$$

Mostremos que una correlación análoga tiene lugar también para el operador $(E + A)^{-1}$, si $\|A\| < 1$. Examinemos una sucesión de operadores $\{A_p\}$

$$A_p = \sum_{k=0}^{p} (-A)^k.$$

Es fácil comprobar que

$$(E + A)\, A_p = E - (-A)^{p+1},$$

por lo cual

$$\|(E + A)\, A_p - E\| = \|A^{p+1}\|. \tag{87.1}$$

Formalmente esta igualdad es verídica también para $p = -1$, siempre que se considere $A_{-1} = 0$.

Luego, se tiene

$$\|(E + A)\, A_p - E\| = \|(A_p - (E + A)^{-1} + A(A_p-(E+A)^{-1})\| \geqslant$$
$$\geqslant \big|\, \|A_p - (E + A)^{-1}\| - \|A\|\cdot\|A_p - (E+A)^{-1}\|\,\big| =$$
$$= (1 - \|A\|)\, \|A_p - (E + A)^{-1}\|.$$

Ahora, teniendo presente (87.1), obtenemos, para $p = -1$, la estimación de la norma del operador $(E + A)^{-1}$, es decir

$$\|(E + A)^{-1}\| \leqslant \frac{\|E\|}{1-\|A\|}.$$

Para cualquier norma subordinada se verifica la igualdad $\|E\| = 1$, por consiguiente, en este caso

$$\|(E+A)^{-1}\| \leqslant \frac{1}{1-\|A\|}. \tag{87.2}$$

Cuando $p \geqslant 0$, obtenemos la estimación para la desviación del operador A_p del operador $(E + A)^{-1}$. A saber,

$$\|A_p - (E+A)^{-1}\| \leqslant \frac{\|A\|^{p+1}}{1-\|A\|}. \tag{87.3}$$

En virtud de la condición $\|A\| < 1$, esto significa que la sucesión $\{A_p\}$ convergerá al operador $(E + A)^{-1}$. Si el operador A_p se considera como *aproximación* hacia el operador $(E+A)^{-1}$, la fórmula (87.3) nos ofrece *la estimación de la exactitud con que se realiza la aproximación.*

Sea A un operador regular cualquiera. Consideraremos el operador $A+\varepsilon_A$, donde ε_A es un operador arbitrario. Llamaremos ε_A *perturbación del operador* A, y $A+\varepsilon_A$, operador perturbado. Aclaremos en qué condiciones, impuestas sobre la magnitud de la norma de perturbación, el operador perturbado será regular. Nos serán de interés en este caso sólo los valores *pequeños* de la norma de perturbación.

El operador A es regular y por eso existe el operador A^{-1}. Por consiguiente, se verifica la igualdad

$$A+\varepsilon_A = A\,(E+A^{-1}\varepsilon_A).$$

De aquí se desprende que el operador $A+\varepsilon_A$ será regular, si, y sólo si, es regular el operador $E+A^{-1}\varepsilon_A$. Esta condición se cumple a ciencia cierta, siempre que

$$\| A^{-1}\varepsilon_A \| < 1$$

para una norma cualquiera. Se cumple con mayor razón, si $\| A^{-1} \| \, \| \varepsilon_A \| < 1$.

De suerte que *un operador perturbado será regular con todas perturbaciones que satisfacen la desigualdad*

$$\| \varepsilon_A \| < \| A^{-1} \|^{-1}. \tag{87.4}$$

Cuando el operador A es perturbado en ε_A, el operador inverso A^{-1} recibirá una perturbación igual a $(A+\varepsilon_A)^{-1}-A^{-1}$. Designemos mediante

$$\delta A = \frac{\| \varepsilon_A \|}{\| A \|}, \qquad \delta A^{-1} = \frac{\| (A+\varepsilon_A)^{-1}-A^{-1} \|}{\| A^{-1} \|} \tag{87.5}$$

las magnitudes de *las perturbaciones relativas* de los operadores A y A^{-1}. Al cumplirse las condiciones (87.4), el operador $E+A^{-1}\varepsilon_A$ será regular y por eso

$$(A+\varepsilon_A)^{-1}-A^{-1} = ((A+\varepsilon_A)^{-1}A-E)\,A^{-1} =$$
$$= ((A^{-1}(A+\varepsilon_A))^{-1}-E)\,A^{-1} = ((E+A^{-1}\varepsilon_A)^{-1}-E)\,A^{-1}.$$

De acuerdo con la fórmula (87.3), para $p=0$, encontramos que

$$\| (A+\varepsilon_A)^{-1}-A^{-1} \| \leqslant \frac{\| A^{-1}\varepsilon_A \| \cdot \| A^{-1} \|}{1-\| A^{-1}\varepsilon_A \|} \leqslant \frac{\| A^{-1} \|^2 \, \| \varepsilon_A \|}{1-\| A^{-1} \| \cdot \| \varepsilon_A \|}.$$

Ahora, teniendo presentes las designaciones (87.5), obtenemos la siguiente estimación:

$$\delta A^{-1} \leqslant \frac{\nu_A \delta A}{1-\nu_A \delta A}, \tag{87.6}$$

donde

$$\nu_A = \| A^{-1} \| \cdot \| A \|. \tag{87.7}$$

El número ν_A se denomina *número de condicionalidad* del operador A. Aunque este número depende de la norma elegida, nunca puede ser muy pequeño. De la igualdad

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

concluímos, tomando en consideración (84.3), que

$$1 \leqslant \| E \| \leqslant \| A^{-1} \| \cdot \| A \| = \nu_A.$$

La fórmula (87.6) muestra que la pequeña perturbación relativa del operador A conduce a una pequeña perturbación relativa de A^{-1} sólo en el caso en que el número de condicionalidad del operador A *no es demasiado grande* en comparación con la unidad. Con este número nos encontraremos también resolviendo otros problemas.

Supongamos que con un operador regular A se resuelve la ecuación operacional

$$Ax = y. \tag{87.8}$$

Examinemos una ecuación perturbada

$$(A+\varepsilon_A)\tilde{x} = y+\varepsilon_y. \tag{87.9}$$

Si se cumple la condición (87.4), la ecuación perturbada (87.9) y la exacta (87.8) tendrán las únicas soluciones $\tilde{x}$ y x. Evaluemos su diferencia.

A la par con (87.5), (87.7) introduzcamos las designaciones correspondientes para las perturbaciones *relativas* en x, y, es decir,

$$\delta x = \frac{\| \tilde{x}-x \|}{\| x \|}, \qquad \delta_y = \frac{\| \varepsilon_y \|}{\| y \|}.$$

Tenemos

$$x = A^{-1}y, \qquad \tilde{x} = (A+\varepsilon_A)^{-1}(y+\varepsilon_y).$$

De aquí encontramos

$$\tilde{x}-x = ((E+A^{-1}\varepsilon_A)^{-1}-E)\, A^{-1}y + (E+A^{-1}\varepsilon_A)^{-1}A^{-1}\varepsilon_y$$

y luego

$$\| \tilde{x}-x \| \leqslant \| (E+A^{-1}\varepsilon_A)^{-1}-E \| \cdot \| x \| + \| (E+A^{-1}\varepsilon_A)^{-1} \| \times \\ \times \| A^{-1} \| \cdot \| \varepsilon_y \|.$$

Convengamos en considerar que se emplea la norma subordinada. Tomando en consideración las estimaciones (87.2), (87.3), como también la desigualdad $\|y\| \leqslant \|A\| \cdot \|x\|$, obtendremos

$$\|\tilde{x}-x\| \leqslant \frac{\|A^{-1}\| \cdot \|\varepsilon_A\| \cdot \|x\|}{1-\|A^{-1}\| \cdot \|\varepsilon_A\|} + \frac{\|A^{-1}\| \cdot \|\varepsilon_y\|}{1-\|A^{-1}\| \cdot \|\varepsilon_A\|} \leqslant$$

$$\leqslant \frac{\|A^{-1}\| \cdot \|A\| \dfrac{\|\varepsilon_A\|}{\|A\|}}{1-\|A^{-1}\| \cdot \|A\| \dfrac{\|\varepsilon_A\|}{\|A\|}} \|x\| + \frac{\|A^{-1}\| \cdot \|A\| \dfrac{\|\varepsilon_y\|}{\|y\|}}{1-\|A^{-1}\| \cdot \|A\| \dfrac{\|\varepsilon_A\|}{\|A\|}} \|x\|.$$

De acuerdo con las designaciones aceptadas esto significa que

$$\delta x \leqslant \frac{\nu_A}{1-\nu_A \delta_A} (\delta A + \delta y). \tag{87.10}$$

La fórmula obtenida muestra nuevamente la significación del número de condicionalidad y esta vez también es importante, desde el punto de vista de estabilidad, que este número sea *no demasiado grande*.

Ejercicios.

1. Demuéstrese que un número de condicionalidad expresado en forma espectral es igual a la razón entre el número singular máximo y el mínimo.
2. Existen los operadores cuyo número de condicionalidad es mínimo. ¿Qué representan en sí estos operadores, si se emplea la norma espectral?
3. Demuéstrese que si un operador se multiplica por operadores unitarios, su número de condicionalidad, expresado en la norma espectral o euclidiana, no varía.
4. Demuéstrese que para cualesquiera operadores regulares A, B se verifica la desigualdad

$$\frac{\|B^{-1}-A^{-1}\|}{\|B^{-1}\|} \leqslant \nu_A \frac{\|A-B\|}{\|A\|}.$$

5. ¿En qué radica la razón por la que el sistema de vectores descrito en el § 22 es muy inestable? Evalúese el número de condicionalidad de un operador en el que las columnas de la matriz coinciden con las coordenadas de los vectores (22.7).

§ 88. Solución estable de las ecuaciones

La fórmula (87.10) muestra que para un operador, próximo a un operador degenerado, *pueden observarse grandes perturbaciones de la solución, incluso cuando las perturbaciones en el operador y en el segundo miembro son pequeñas*. Puede parecer que este hecho sólo se debe a que la propia solución no siempre existe. Sin embargo, la situación es análoga en el caso en que se definen las seudosoluciones.

Efectivamente, supongamos que un operador actúa en un espacio bidimensional. Supongamos, además, que en cierta base ortonormalizada, a la ecuación (85.1) le corresponde un sistema de ecuaciones algebraicas lineales del tipo siguiente:

$$1\cdot x_1 + 0\cdot x_2 = 1$$
$$0\cdot x_1 + 0\cdot x_2 = 1.$$

Es fácil determinar que la seudosolución normal u_0 tendrá las siguientes coordenadas:

$$u_0 = (1,\ 0).$$

Es muy posible que la ecuación perturbada conducirá en la misma base al sistema

$$1\cdot x_1 + 0\cdot x_2 = 1,$$
$$0\cdot x_1 + \varepsilon\cdot x_2 = 1,$$

donde el número ε, aunque pequeño, será, sin embargo distinto de cero. Ahora la seudosolución normal $u_0^{(\varepsilon)}$ de la ecuación perturbada tiene las siguientes coordenadas

$$u_0^{(\varepsilon)} = (1,\ \varepsilon^{-1}).$$

Cuando ε son pequeños, los vectores u_0 y $u_0^{(\varepsilon)}$ no sólo son muy diferentes, sino que son *casi ortogonales.*

Si una ecuación tiene más de una seudosolución, en el caso general, las perturbaciones pequeñas en el operador y en el segundo miembro siempre causarán grandes perturbaciones en la seudosolución normal. No obstante, mostraremos que *a pesar de que muchas nociones relacionadas con las ecuaciones operacionales son inestables, la seudosolución normal puede ser definida de una manera estable.*

Supongamos que el operador A actúa del espacio X en el espacio Y y que, además, se resuelve la ecuación (85.1). Por analogía con la funcional del residuo, consideraremos la así llamada *funcional de regularización*

$$\Phi_\alpha(x) = \alpha \mid x \mid^2 + \mid Ax - y \mid^2, \qquad (88.1)$$

donde $\alpha \geqslant 0$. Está claro que para $\alpha = 0$ esta funcional coincide con la del residuo y alcanza su mínimo en las seudosoluciones de la ecuación (85.1). Aclaremos, en qué vectores alcanza el mínimo la funcional de regularización para $\alpha > 0$. Haciendo uso de la descomposición (86.2), encontramos

$$\Phi_\alpha(x) = \sum_{k=1}^{t} (\alpha \mid \alpha_k \mid^2 + \mid \rho_k\alpha_k - \beta_k \mid^2) + \alpha \sum_{k=t+1}^{m} \mid \alpha_k \mid^2 + \sum_{p=t+1}^{n} \mid \beta_p \mid^2.$$

De aquí se deduce que para alcanzar el mínimo, es necesario tomar los valores nulos de las últimas coordenadas $\alpha_{t+1}, \ldots, \alpha_m$ y minimizar, para todo $k \leqslant t$, la expresión

$$\alpha \mid \alpha_k \mid^2 + \mid \rho_k\alpha_k - \beta_k \mid^2.$$

Esto nos da, para $k \leqslant t$,

$$\alpha_k = \frac{\rho_k \beta_k}{\alpha + \rho_k^2}.$$

De suerte que, el valor mínimo de la funcional de regularización (88.1) se alcanza para todo $\alpha > 0$ en *el único* vector

$$x_\alpha = \sum_{k=1}^{t} \frac{\rho_k \beta_k}{\alpha + \rho_k^2} x_k. \tag{88.2}$$

La comparación de las fórmulas (86.4), (88.2) permite establecer ciertas correlaciones que ligan x_α y x_0. Para ρ, $\alpha > 0$ tenemos

$$0 \leqslant \frac{|\beta|^2}{\rho^2} - \frac{\rho^2 |\beta|^2}{(\alpha+\rho^2)^2} = \frac{|\beta|^2 \alpha^2 + 2|\beta|^2 \alpha\rho^2}{\rho^2 (\alpha+\rho^2)^2} =$$

$$= \frac{2\alpha |\beta|^2 \left(\frac{\alpha}{2} + \rho^2\right)}{\rho^2 (\alpha+\rho^2)^2} \leqslant \frac{2\alpha |\beta|^2}{\rho^4},$$

por lo cual se desprende que

$$|x_\alpha|^2 \leqslant |x_0|^2 \leqslant |x_\alpha|^2 + 2\alpha\eta^2, \tag{88.3}$$

donde

$$\eta^2 = \sum_{k=1}^{t} \frac{|\beta_k|^2}{\rho_k^4}.$$

A continuación encontramos

$$x_0 - x_\alpha = \alpha \sum_{k=1}^{t} \frac{\beta_k}{\rho_k (\alpha + \rho_k^2)} x_k,$$

de donde concluimos que

$$|x_0 - x_\alpha| \leqslant \alpha\gamma, \tag{88.4}$$

donde

$$\gamma^2 = \sum_{k=1}^{t} \frac{|\beta_k|^2}{\rho_k^6}.$$

Por consiguiente,

$$\lim_{\alpha \to +0} x_\alpha = x_0.$$

De este modo, cuando los valores de α son pequeños, el vector x_α puede servir de *aproximación* a la seudosolución normal x_0.

Descompongamos los vectores x_α y x_0 según las bases singulares, por analogía con (86.2). Por comprobación directa es fácil convencerse de que x_α satisface la ecuación

$$(A^*A + \alpha E)\, x_\alpha = A^* y. \tag{88.5}$$

Para $\alpha > 0$ el operador $A^*A + \alpha E$ es definido positivo, por lo cual existe el operador $(A^*A + \alpha E)^{-1}$, es decir,

$$x_\alpha = (A^*A + \alpha E)^{-1}A^*y. \tag{88.6}$$

En el vector x_α se alcanza el valor mínimo de la funcional (88.1), por consiguiente, $\Phi_\alpha\ (x_\alpha) \leqslant \Phi_\alpha\ (x_0)$. Teniendo presentes (88.3), (88.4), obtenemos

$$\begin{aligned} |Ax_\alpha - y|^2 &\leqslant |Ax_0 - y|^2 + \alpha(|x_0|^2 - |x_\alpha|^2) \leqslant \\ &\leqslant |Ax_0 - y|^2 + 2\alpha^2\eta^2. \end{aligned} \tag{88.7}$$

Además, $\Phi_\alpha\ (x_\alpha) \leqslant \Phi_\alpha\ (0)$, de donde se infiere

$$|x_\alpha| \leqslant \frac{|y|}{\alpha^{1/2}}.$$

Junto con (88.6) esto es testimonio de que para cualquier operador A y todo vector y se tiene, para $\alpha > 0$,

$$|(A^*A + \alpha E)^{-1} A^*y| \leqslant \frac{|y|}{\alpha^{1/2}}. \tag{88.8}$$

En la resolución *práctica* de la ecuación (85.1) el operador A y el segundo miembro y se fijan, corrientemente, de manera inexacta y nos vemos obligados a considerar, en lugar de ellos, el operador perturbado $\widetilde{A}$ y el segundo miembro, también perturbado, $\widetilde{y}$. Si en los espacios X, Y se emplea a título de norma la longitud de los vectores, entonces a ésta última le queda subordinada la norma espectral de los operadores. Por esta razón supondremos que

$$\|A - \widetilde{A}\|_2 \leqslant \bar{\varepsilon}_A, \qquad \|y - \widetilde{y}\| \leqslant \bar{\varepsilon}_y. \tag{88.9}$$

La determinación de la solución aproximada $\widetilde{x}_\alpha$, partiendo de $\widetilde{A}$ e $\widetilde{y}$ perturbados, conduce a una ecuación de tal índole:

$$(\widetilde{A}^*\widetilde{A} + \alpha E)\, \widetilde{x}_\alpha = \widetilde{A}^*\widetilde{y}. \tag{88.10}$$

De (88.5), (88.10) encontramos

$$\begin{aligned} (\widetilde{A}^*\widetilde{A} + \alpha E)(\widetilde{x}_\alpha - x_\alpha) &= A^*(Ax_\alpha - y) - \widetilde{A}^*(\widetilde{A}x_\alpha - \widetilde{y}) = \\ &= (A - \widetilde{A})^*(Ax_\alpha - y) - \widetilde{A}^*((\widetilde{A} - A)\, x_\alpha - (\widetilde{y} - y)). \end{aligned}$$

Esto significa que la diferencia $\widetilde{x}_\alpha - x$ es la solución de la ecuación con el operador $(\widetilde{A}^*\widetilde{A} + \alpha \mathrm{E})$ y el segundo miembro del tipo $z = u + \widetilde{A}^*v$, donde

$$u = (A - \widetilde{A})^*(Ax_\alpha - y),$$
$$v = -((\widetilde{A} - A)\, x_\alpha - (\widetilde{y} - y)).$$

Por eso

$$\tilde{x}_\alpha - x_\alpha = (\tilde{A}^*\tilde{A} + \alpha E)^{-1} u + (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \tilde{A}^* v.$$

Evaluemos ahora las normas de ambos sumandos en esta igualdad.

Los valores propios del operador $\tilde{A}^*\tilde{A} + \alpha E$ no son inferiores a α. Por consiguiente, los valores propios del operador $(\tilde{A}^*\tilde{A} + \alpha E)^{-1}$ no son superiores a α^{-1}. Para un operador definido positivo la norma espectral coincide con el valor propio máximo, es decir,

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \|_2 \leqslant \alpha^{-1}.$$

Teniendo presentes (88.7), (88.9), tendremos

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} u \| \leqslant \| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \|_2 \| u \| \leqslant$$
$$\leqslant \frac{\bar{\varepsilon}_A}{\alpha} \| Ax_\alpha - y \| \leqslant \frac{\bar{\varepsilon}_A}{\alpha} (\| Ax_0 - y \|^2 + 2\alpha^2\eta^2)^{1/2}.$$

Para estimar el segundo sumando, haremos uso de las fórmulas (88.3), (88.8), (88.9). Encontramos

$$\| (\tilde{A}^*\tilde{A} + \alpha E)^{-1} \tilde{A}^* v \| \leqslant \frac{\| v \|}{\alpha^{1/2}} \leqslant \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y).$$

Así pues,

$$\| \tilde{x}_\alpha - x \| \leqslant \frac{\bar{\varepsilon}_A}{\alpha} (\| Ax_0 - y \|^2 + 2\alpha^2\eta^2)^{1/2} + \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y).$$

El error total de la seudosolución calculada $\tilde{x}_\alpha$ es

$$\| x_0 - \tilde{x}_\alpha \| \leqslant \| x_0 - x_\alpha \| + \| \tilde{x}_\alpha - x_\alpha \| \leqslant$$
$$\leqslant \alpha\gamma + \frac{\bar{\varepsilon}_A}{\alpha} (\| Ax_0 - y \|^2 + 2\alpha^2\eta^2)^{1/2} +$$
$$+ \frac{1}{\alpha^{1/2}} (\bar{\varepsilon}_A \| x_0 \| + \bar{\varepsilon}_y). \quad (88.11)$$

El segundo miembro de esta desigualdad no contiene ninguna información referente a los perturbados $\tilde{A}$, $\tilde{y}$ dados. Por ello existe tal α, para el cual el miembro citado alcanza su máximo. Este valor de α asegurará la aproximación casi mejor de $\tilde{x}_\alpha$ a la seudosolución normal exacta x_0.

Supongamos que $\bar{\varepsilon}_A$ y $\bar{\varepsilon}_y$ son unas magnitudes de orden ε y el propio ε es suficientemente pequeño. Si la ecuación exacta (85.1) tiene solución, entonces $Ax_0 - y = 0$. En este caso el segundo miembro de (88.11) es, a juzgar por el carácter de dependencia de α

y ε, una función del tipo

$$\alpha+\varepsilon+\frac{\varepsilon}{\alpha^{1/2}}.$$

Cuando $\alpha=\varepsilon^{2/3}$, esta función toma un valor de orden $\varepsilon^{2/3}$. En cambio, si la ecuación exacta no tiene ninguna solución, entonces $Ax_0 - y \neq 0$. Ahora el segundo miembro de (88.11) es una función del tipo

$$\alpha+\frac{\varepsilon}{\alpha}+\frac{\varepsilon}{\alpha^{1/2}}.$$

Cuando $\alpha=\varepsilon^{1/2}$, ella toma un valor de orden $\varepsilon^{1/2}$.

De este modo, si los datos de entrada de la ecuación (85.1) *están prefijados con una exactitud del orden* ε, *la seudosolución normal puede hallarse con la exactitud del orden* $\varepsilon^{2/3}$, *si la ecuación exacta es soluble, y con una exactitud del orden* $\varepsilon^{1/2}$, *en el caso contrario.*

El parámetro α, que asegura la aproximación necesaria de $\tilde{x}_\alpha$, no puede ser determinado, partiendo sólo de $\tilde{A}$ e $\tilde{y}$ perturbados. Esto se debe, principalmente, a que las condiciones (88.9) no garantizan *la continuidad* de la seudosolución normal en el campo dado de variación del operador y del segundo miembro. Con el fin de determinar el parámetro α se usa, habitualmente, una información *adicional* acerca de la solución. En algunos problemas no se requiere una proximidad garantizada a la seudosolución normal, sino que se considera suficiente la definición estable del mínimo de la funcional del residuo. En los problemas de este tipo la determinación del parámetro resulta algo más simple. Aunque todas estas cuestiones son muy importantes, no nos detendremos ante ellas, puesto que salen de los márgenes de este curso.

Ejercicios.

1. Demuéstrese que η en la estimación (88.3) es la longitud de la solución normal de la ecuación

$$A^*A\,(A^*A)^{1/2}x=A^*y.$$

2. Demuéstrese que γ en la estimación (88.4) es la longitud de la solución normal de la ecuación

$$(A^*A)^2\,x=A^*y.$$

3. Demuéstrese que la diferencia $x_\alpha - x_\beta$ satisface la ecuación

$$(A^*A+\alpha E)\,(A^*A+\beta E)\,(x_\alpha - x_\beta)=(\beta-\alpha)\,A^*y$$

4. Compárense (88.11) y (87.10). ¿Qué puede decirse sobre la estimación (88.11) en el caso de un operador regular A?

5. ¿Con qué exactitud puede calcularse la seudosolución normal, si $A=0$?

§ 89. La perturbación y los valores propios

La perturbación de un operador conduce, generalmente, a la variación de todos sus valores propios y vectores propios. Siendo muy compleja la investigación de dicha dependencia, nos limitaremos a ilustrarla con unos ejemplos. Resulta *más cómodo* describir el problema dado en términos de las matrices de los operadores en vez de considerarlo en términos de los propios operadores.

Sea B una matriz arbitraria de estructura simple y sea H una matriz tal que

$$H^{-1}BH = \Lambda, \tag{89.1}$$

donde Λ es la matriz diagonal de los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_m$. Consideraremos una matriz perturbada $B + \varepsilon_B$ y alguno de sus valores propios λ. La matriz $B + \varepsilon_B - \lambda E$ es degenerada, por lo cual será degenerada también la matriz

$$H^{-1}(B + \varepsilon_B - \lambda E)\,H = (\Lambda - \lambda E) + H^{-1}\varepsilon_B H.$$

Se presentan dos posibles casos:

1) $\lambda = \lambda_i$ para cierto i,

2) $\lambda \neq \lambda_i$ para cualquier valor de i.

En el segundo caso la matriz $\Lambda - \lambda E$ es regular, por consiguiente,

$$(\Lambda - \lambda E) + H^{-1}\varepsilon_B H = (\Lambda - \lambda E)\,(E + (\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H).$$

La matriz, que interviene como segundo factor, es degenerada. Esto significa que toda norma de la matriz $(\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H$ debe ser no inferior a la unidad. En particular,

$$\|(\Lambda - \lambda E)^{-1}H^{-1}\varepsilon_B H\|_2 \geqslant 1.$$

De aquí se deduce que

$$\max_{1 \leqslant i \leqslant n} |(\lambda_i - \lambda)^{-1}|\,\|H^{-1}\|_2\,\|\varepsilon_B\|_2\,\|H\|_2 \geqslant 1$$

o bien

$$\min_{1 \leqslant i \leqslant n} |\lambda_i - \lambda| \leqslant \|H^{-1}\|_2\,\|\varepsilon_B\|_2\,\|H\|_2.$$

En el primer caso esta desigualdad se verifica también, por lo cual

$$|\lambda_i - \lambda| \leqslant \nu_H\,\|\varepsilon_B\|_2 \tag{89.2}$$

por lo menos para un valor de i. Aquí

$$\nu_H = \|H^{-1}\|_2\,\|H\|_2$$

es el número de condicionalidad de la matriz H expresado por la norma espectral.

La correlación obtenida quiere decir que cualquiera que sea la perturbación ε_B de la matriz B, para todo valor propio λ de la matriz perturbada $B + \varepsilon_B$ existe tal valor propio λ_i de la matriz B que se verifique la desigualdad (89.2). Cabe notar que en nuestros razonamientos *nunca hemos requerido que las perturbaciones ε_B fueran pequeñas*. La correlación (89.2) puede ser interpretada de una manera algo diferente. A saber

Los valores propios de una matriz perturbada se disponen en un dominio que representa la unión de todos los círculos de radio $\nu_H \| \varepsilon_B \|_2$ y centros en λ_i.

Las columnas de la matriz H representan en sí los vectores propios de la matriz B. Por ello, de (89.2) se infiere que como medida general de la sensibilidad de los valores propios a la perturbación de una matriz puede, evidentemente, servir el número de condicionalidad de la matriz H compuesta de los vectores propios (¡no de la matriz B!). La matriz H, que satisface (89.1), no es única, puesto que los valores propios están definidos, salvo unos factores arbitrarios. Convengamos en considerar que la matriz se elige siempre de tal manera que el valor ν_H queda mínimo. Recordemos que en todo caso $\nu_H \geqslant 1$.

Si B es una matriz normal y, además, hermitiana o unitaria, entonces H siempre puede elegirse unitaria. En este caso $\nu_H = 1$ y, por lo tanto,

$$|\lambda_i - \lambda| \leqslant \| \varepsilon_B \|_2. \tag{89.3}$$

Examinemos más detalladamente el caso de una matriz hermitiana B con la perturbación hermitiana ε_B. Ahora podemos mostrar que

En todo círculo con el centro en λ_i, de radio $\| \varepsilon_B \|_2$ está contenido por lo menos un solo valor propio de la matriz perturbada.

Efectivamente, consideraremos *convencionalmente* la matriz $B + \varepsilon_B$ como «exacta» y la matriz $B = (B + \varepsilon_B) - \varepsilon_B$, como «perturbada» cuya perturbación es igual a $-\varepsilon_B$. Repitiendo textualmente todos los razonamientos, obtendremos una fórmula, análoga a la (89.3) con la particularidad de que en la primera los valores propios de las matrices B y $B + \varepsilon_B$ cambian sus papeles. Esto significa que para todo valor propio λ_i de la matriz «perturbada» B existe infaliblemente por lo menos un solo valor propio λ de la matriz «exacta» $B + \varepsilon_B$, para el cual la desigualdad (89.3) tiene lugar.

Si los valores propios de la matriz B son simples, entonces siendo la perturbación ε_B suficientemente pequeña, todos los círculos *se separan* y todo círculo contendrá *uno*, y *sólo un* valor propio de la matriz perturbada.

La fórmula (89.3) muestra que los valores propios de las matrices normales poseen una estabilidad considerable a la perturbación. No obstante, en el problema general de la definición de los valores propios este fenómeno es más bien una excepción que una regla.

A título de ejemplo consideraremos un caso, «límite» en cierto sentido, en que la matriz B se compone de una sola caja canónica de Jordan. Se puede considerar convencionalmente que todos los vectores propios de tal matriz son colineales, la matriz de vectores propios es degenerada y, por consiguiente, su número de condicionalidad es igual a «infinito». Así pues, supongamos que la matriz B de orden m tiene por expresión

$$B=\begin{pmatrix} \lambda_0 & 1 & & & & 0\\ & \lambda_0 & 1 & & & \\ & & \lambda_0 & 1 & & \\ & & & \ddots & \ddots & \\ & & & & & \\ 0 & & & & \lambda_0 & 1\\ & & & & & \lambda_0 \end{pmatrix}.$$

Es evidente que su polinomio característico es $(\lambda-\lambda_0)^m$.

Tomemos ahora tal matriz de perturbación ε_B en la cual sólo un elemento, dispuesto en posición $(m, 1)$ es diferente de cero y es igual al número ε. El polinomio característico de la matriz perturbada es igual a $(\lambda-\lambda_0)^m-\varepsilon$. Por ello, los valores propios de la matriz perturbada se encuentran a la distancia de $|\varepsilon|^{1/m}$ de los valores propios de la matriz exacta. Si, por ejemplo, $m=20$, $\varepsilon=10^{-10}$ y λ_0 es de orden uno, entonces no se puede hablar de ninguna estabilidad práctica.

Es importante comprender que la inestabilidad de los valores propios no está ligada obligatoriamente a la presencia de valores propios múltiples y tampoco, con la mayor razón, a la presencia de las cajas de Jordan. Examinemos una matriz de orden 20

$$B=\begin{pmatrix} 20 & 20 & & 0 & & \\ & 19 & 20 & & & \\ & & 18 & 20 & & \\ & & & \ddots & \ddots & \\ & & & & & \\ 0 & & & & 2 & 20\\ & & & & & 1 \end{pmatrix}.$$

Es una matriz triangular y, por eso, sus valores propios se constituirán de elementos diagonales. A primera vista, están suficientemente bien separados y parecería que no hay ninguna razón de esperar la inestabilidad. Pero, agreguemos la perturbación ε al elemento nulo dispuesto en la posición (20.1). El término independiente del polinomio característico variará en este caso a una magnitud de $20^{19}\,\varepsilon$.

Dado que el producto de los valores propios es igual al término independiente, los mismos valores propios deben alterarse en un grado considerable.

Unos problemas, más complejos aún, surgen al estudiar la estabilidad de los vectores propios. Está claro que si un valor propio λ de la matriz B es inestable ante una perturbación, entonces el vector propio x, que corresponde a dicho valor, no puede ser estable a ciencia cierta, puesto que B, λ, x están ligados entre sí mediante la correlación lineal $Bx = \lambda x$.

Sin embargo, resulta importante subrayar que si incluso los valores propios no varían como resultado de una perturbación, los vectores propios no sólo pueden ser inestables, sino que su número puede variar. Por ejemplo, la primera de las matrices

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & \varepsilon \\ 0 & 0 & 1 \end{pmatrix}$$

tiene tres vectores propios linealmente independientes y la segunda, dos vectores, aunque los valores propios de estas matrices son iguales. Desde el punto de vista teórico, este fenómeno está relacionado sólo con el hecho de presencia de valores propios múltiples de la matriz inicial. Pero, *cuando una matriz se da aproximadamente, es difícil y frecuentemente imposible decir cuáles valores propios deben considerarse múltiples y cuáles, simples.*

Los problemas referentes al estudio de la estabilidad de los valores propios, los vectores propios y radicales aparecen como los más complejos en los apartados del álgebra relacionados con los cálculos.

Ejercicios.

1. Supongamos que una matriz B es de estructura simple, pero tiene valores propios múltiples. Demuéstrese que para cualquier número $\varepsilon > 0$, tan pequeño como se quiera, existe tal perturbación ε_B, la cual satisface la condición $\| \varepsilon_B \| < \varepsilon$, que la matriz $B + \varepsilon_B$ ya no será de estructura simple.

2. Supongamos que una matriz B tiene valores propios distintos dos a dos y que $d > 0$ es la distancia mínima entre los valores propios. Demuéstrese que existe tal perturbación ε_B, la cual satisface la condición $\| \varepsilon_B \|_2 > d$, que la matriz $B + \varepsilon_B$ no será de estructura simple.

3. Supongamos ahora que la matriz B es hermitiana. Demuéstrese que si una perturbación hermitiana ε_B satisface la condición $\| \varepsilon_B \|_2 < d/2$, la matriz $B + \varepsilon_B$ tiene valores propios distintos dos a dos.

4. Supongamos, por fin, que la matriz B no es hermitiana y tiene valores propios distintos dos a dos. Demuéstrese que existe tal número r, el cual satisface la condición $0 < r \leqslant d$, que la matriz $B + \varepsilon_B$ tiene estructura simple, siempre que $\| \varepsilon_B \|_2 \leqslant r$.

5. Trátese de establecer una relación más estrecha entre los números r y d.