Spectral theory

Preliminaries *(Secs. 24–26)*

DIRECT SUM OF OPERATORS

DEFINITION: Given $\underline{A} \in \mathcal{L}(\mathcal{V};\mathcal{V})$, a subspace $\mathcal{U} \subset \mathcal{V}$ is *invariant [under $\underline{A}$]* if $\vec{v} \in \mathcal{U} \Rightarrow \underline{A}\vec{v} \in \mathcal{U}$. In other words, ran $\left(\underline{A}\big|_{\mathcal{U}}\right) \subseteq \mathcal{U}$.

NOTE: $\underline{A}$ does *not necessarily* leave the individual vectors in $\mathcal{U}$ invariant. (I.e., $\vec{v} = \underline{A}\vec{v} \quad \forall \vec{v} \in \mathcal{U}$ is not required).

DEFINITION: A square matrix A is *block-diagonal* if it partitions as

$$\begin{array}{c|c|c|c} A_1 & 0 & \dots & 0 \\ \hline 0 & A_2 & \dots & 0 \\ \hline \vdots & \vdots & \ddots & 0 \\ \hline 0 & 0 & \dots & A_L \end{array}$$

Example:

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

where all the (possibly) nonzero blocks along the diagonal are square.

We want to give an "intrinsic" description of this structure. Interpret A as the matrix of an endomorphism $\underline{A}$ with respect to some basis $\{\vec{v}_j\}_{j=1}^n$ for $\mathcal{V}$. Then each diagonal block A_μ is associated with some segment of the basis (a subsequence of adjacent vectors),

$$\{\vec{v}_{j_\mu}, \vec{v}_{j_\mu+1}, \dots, \vec{v}_{j_{\mu+1}-1}\}.$$

We have $j_{\mu+1} = j_\mu + n_\mu$, if A_μ is an $n_\mu \times n_\mu$ matrix. Let $\mathcal{V}_\mu$ be the span of these vectors (which has dimension n_μ). Then $\underline{A}$ maps $\mathcal{V}_\mu$ into itself ($\mathcal{V}_\mu$ is $\underline{A}$-invariant);

$$\underline{A}\big|_{\mathcal{V}_\mu} \equiv \underline{A}_\mu : \mathcal{V}_\mu \to \mathcal{V}_\mu$$

is represented by the matrix A_μ. Furthermore, $\mathcal{V} = \mathcal{V}_1 \oplus \cdots \oplus \mathcal{V}_L$, and each $\underline{A}_\mu$ can be identified with an operator $\hat{\underline{A}}_\mu : \mathcal{V} \to \mathcal{V}$ defined by

$$(*) \qquad \begin{cases} \hat{\underline{A}}_\mu \vec{v} \equiv \underline{A}_\mu \vec{v} = \underline{A}\vec{v} & \text{if } \vec{v} \in \mathcal{V}_\mu, \\ \hat{\underline{A}}_\mu \vec{v} \equiv \vec{0} & \text{if } \vec{v} \in \mathcal{V}_\nu, \quad \nu \neq \mu. \end{cases}$$

(If $\vec{v}$ is a nontrivial sum of vectors from several $\mathcal{V}_\rho$, then $\underline{A}\vec{v}$ is defined by linearity.) The matrix of $\hat{\underline{A}}_\mu$ is

$$\left(\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & A_\mu & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

Theorem 24.1′. *In the situation just described, we have*

$$\underline{A} = \sum_{\mu=1}^{L} \hat{\underline{A}}_\mu \qquad \textit{and} \qquad \hat{\underline{A}}_\mu \hat{\underline{A}}_\nu = \underline{0} \quad \text{if } \mu \neq \nu. \tag{$\dagger$}$$

Definition: If $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \cdots$ and each $\mathcal{V}_\mu$ is invariant under $\underline{A}$, then $\underline{A}$ is the *direct sum* of the operators $\hat{\underline{A}}_\mu$ defined by $(*)$. [This makes sense even if the $\mathcal{V}_\mu$ are infinite-dimensional or if there are infinitely many of them; $(\dagger)$ still applies in such cases.] There is another version of this definition corresponding to the other definition of the direct sum of vector spaces: Let a sequence of vector spaces $\{\mathcal{V}_\mu\}$ be given, and let $\underline{A}_\mu$ be an endomorphism of $\mathcal{V}_\mu$. The direct sum, $\mathcal{V}$, of the $\mathcal{V}_\mu$ is defined as their Cartesian product with the obvious addition and scalar multiplication operations. Then $\hat{\underline{A}}_\mu$ is defined by $(*)$, and the direct-sum operator, $\underline{A}$, is defined by $(\dagger)$.

Remark: The converse of Theorem 24.1′ is *false*: $(\dagger)$ does not imply that $\mathcal{V} = \mathcal{V}_1 \oplus \cdots \oplus \mathcal{V}_L$ with $\mathcal{V}_\mu$ $\underline{A}$-invariant and $(*)$ holding. Counterexample:

$$\hat{A}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \hat{A}_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The book's definition of "direct sum of operators" is therefore objectionable.

Theorem 24.2. *In the finite-dimensional case,*

$$\det A = \prod_{\mu=1}^{L} \det A_\mu\,.$$

(Note that the factors are *not* $\det \hat{A}_\mu$, which are 0.)

Definitions: If $\underline{A}\vec{v} = \lambda\vec{v} \quad (\lambda \in \mathcal{F}, \quad \vec{v} \in \mathcal{V}, \quad \underline{A} \in \mathcal{L}(\mathcal{V};\mathcal{V}))$, and $\vec{v} \neq \vec{0}$, then

$\vec{v}$ is an *eigenvector* of $\underline{A}$,
λ is an *eigenvalue* of $\underline{A}$.

Note: $\lambda = 0$ is allowed! The restriction $\vec{v} \neq 0$ is needed since otherwise *every* λ would be an eigenvalue of every $\underline{A}$.

Definition: $\sigma(\underline{A}) =$ *spectrum* of $\underline{A} \quad =$ (if $\dim \mathcal{V} < \infty$) the set of all eigenvalues of $\underline{A}$. (In infinite-dimensional spaces, $\sigma(\underline{A})$ is defined to contain the eigenvalues and more besides.)

Remark: If one of the $\mathcal{V}_\mu$ in a direct-sum decomposition of $\underline{A}$ is one-dimensional, then the vectors in $\mathcal{V}_\mu$ are eigenvectors.

Theorem 25.1. *$\forall\lambda \in \sigma(\underline{A}), \quad \mathcal{V}(\lambda) \equiv \{\, \vec{v} \in \mathcal{V} : \underline{A}\vec{v} = \lambda\vec{v}$ or $\vec{v} = 0, \}$ is a subspace of $\mathcal{V}$ and is $\underline{A}$-invariant. (On $\mathcal{V}(\lambda)$, $\underline{A}$ acts as a multiple of the identity.)*

Proof: should be obvious.

Definitions: $\mathcal{V}(\lambda) =$ *eigenspace* corresponding to λ.
$\dim \mathcal{V}(\lambda) =$ *geometric multiplicity* of λ.

Theorem 25.2′. *$\dim \mathcal{V} < \infty \Rightarrow$*

$$\lambda \in \sigma(\underline{A}) \quad \Longleftrightarrow \quad \underline{A} - \lambda \quad [\equiv \underline{A} - \lambda\underline{1}] \quad \text{is singular.}$$

Furthermore, $\mathcal{V}(\lambda) = \ker(\underline{A} - \lambda)$. [Again obvious.]

If $\dim \mathcal{V} = \infty$, then $\underline{A} - \lambda$ can be singular without failing to be injective. This is the essential reason why $\sigma(\underline{A})$ is larger than the set of eigenvalues. Crudely speaking, $\sigma(\underline{A})$ is defined so that Theorem 25.2′ remains true in infinite dimensions, for a suitable interpretation of "singular".

Until further notice, $\dim \mathcal{V} = N < \infty$. Theorem 25.2′ implies

$$\lambda \in \sigma(\underline{A}) \iff \det(\underline{A} - \lambda) = 0.$$

Here $\det(\underline{A} - \lambda)$ is a polynomial in λ, of degree N, called the *characteristic polynomial* of $\underline{A}$. The eigenvalues are its roots.

Theorem. *The determinant of an endomorphism is independent of the basis used to define the endomorphism's matrix (as long as the same basis is used for both domain and codomain).*

PROOF: I assume the basic facts about (a) determinants and (b) basis changes are recalled from an earlier course. [Some of these facts will be reviewed later.] When the basis is changed, the matrix of $\underline{A}$ changes from A to SAS^{-1} (for some matrix S). But

$$\det(SAS^{-1}) = (\det S)(\det A)(\det S^{-1}) = \det A$$

since $\det S^{-1} = (\det S)^{-1}$.

Consequently, the *invariant* number $\det A$ deserves to be called $\det \underline{A}$ (determinant of the *operator*, not just of the matrix). This is not the case for operators in $\mathcal{L}(\mathcal{V};\mathcal{U})$ with $\mathcal{V} \neq \mathcal{U}$ (even if the dimensions are the same), since the bases in $\mathcal{V}$ and $\mathcal{U}$ can be changed independently.

The coefficients of the characteristic polynomial provide additional *fundamental invariants* of $\underline{A}$:

DEFINITION: μ_j is defined by

$$\begin{aligned}\det(\underline{A} - \lambda) &= \sum_{j=0}^{N} \mu_j (-1)^{N-j} \lambda^{N-j} \\ &= (-1)^N \lambda^N + (-1)^{N-1} \mu_1 \lambda^{N-1} + \cdots + \mu_N\,.\end{aligned}$$

(Note that $\mu_0 = 1$ always.)

REMARK: In studying the invariants of an endomorphism, it is helpful to use the following definition of the determinant:

$$\det A = \sum (\pm 1) A^1{}_{j_1} A^2{}_{j_2} \cdots A^N{}_{j_N}\,,$$

where the sum is over all permutations $(j_1, \ldots, j_N)$ of $(1, \ldots, N)$ and the ± 1 is the parity of the permutation. I.e., each term in $\det A$ contains exactly one factor from each row and one factor from each column, and every possible such term appears, with coefficient ± 1 according to whether an even or an odd number of transpositions is needed to convert the list of row indices into the list of column indices. To review (or learn) this and other properties of determinants, I recommend the book of Kolman.

Let's examine the fundamental invariants for operators on spaces of low dimension:

$\underline{N=1}$ is trivial: $\det(\underline{A} - \lambda) = A - \lambda$, so $\mu_1 = \mu_N = A$.

$\underline{N=2}$: $\det(\underline{A}-\lambda) = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix}$

$$= (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + ad - bc.$$

Therefore,

$$\mu_2 = \mu_N = \det \underline{A}, \qquad \mu_1 = a+d \equiv \operatorname{tr} \underline{A}.$$

In general, the *trace* of $\underline{A}$ is the sum of the diagonal elements of its matrix:

Definition: $\operatorname{tr} A \equiv \sum_{j=1}^{N} A^j{}_j$.

THEOREM. $\operatorname{tr} AB = \operatorname{tr} BA$. *Hence* tr *is invariant under any* **cyclic** *permutation: e.g.,* $\operatorname{tr} ABC = \operatorname{tr} BCA$ *but not necessarily* $= \operatorname{tr} ACB$.

PROOF: $\operatorname{tr} AB = A^j{}_k B^k{}_j = B^k{}_j A^j{}_k = \operatorname{tr} BA$.

Corollary. $\operatorname{tr} \underline{A}$ *is an invariant of the endomorphism* $\underline{A}$.

PROOF: $\operatorname{tr}(SAS^{-1}) = \operatorname{tr}(S^{-1}SA) = \operatorname{tr} A$.

$\underline{N=3}$: $\det(A-\lambda) = \begin{vmatrix} A^1{}_1-\lambda & A^1{}_2 & A^1{}_3 \\ A^2{}_1 & A^2{}_2-\lambda & A^2{}_3 \\ A^3{}_1 & A^3{}_2 & A^3{}_3-\lambda \end{vmatrix} =$

$$-\lambda^3 + \lambda^2(A^1{}_1 + A^2{}_2 + A^3{}_3)$$
$$- \lambda\left[\begin{vmatrix} A^2{}_2 & A^2{}_3 \\ A^3{}_2 & A^3{}_3 \end{vmatrix} + 2 \text{ similar terms}\right] + \det A.$$

The thing in brackets is the trace of the matrix of 2×2 minors of A. (For larger N, μ_2 will still be the sum of all the distinct "diagonal" 2×2 minors, although there will be too many of the latter to fit into an $N\times N$ matrix.)

In general (for any N), we can now easily see that

$$\mu_1 = \operatorname{tr} \underline{A}, \qquad \mu_N = \det \underline{A}.$$

What can we say about the other μ's?

CLAIM: $\mu_2 = \frac{1}{2}[(\operatorname{tr} \underline{A})^2 - \operatorname{tr}(\underline{A}^2)]$.

PROOF: $\operatorname{tr}(A^2) = \sum_j \sum_k A^j{}_k A^k{}_j = \sum_j A^j{}_j A^j{}_j + \sum_{j\neq k} A^j{}_k A^k{}_j\,.$

$$(\operatorname{tr} A)^2 = \Bigl(\sum_j A^j{}_j\Bigr)\Bigl(\sum_k A^k{}_k\Bigr) = \sum_j A^j{}_j A^j{}_j + \sum_{j\neq k} A^j{}_j A^k{}_k\,.$$

Therefore,

$$\begin{aligned} \operatorname{tr} A^2 - (\operatorname{tr} A)^2 &= \sum_{j\neq k} \bigl(A^j{}_k A^k{}_j - A^j{}_j A^k{}_k\bigr) \\ &= -2\sum_{j<k} \begin{vmatrix} A^j{}_j & A^j{}_k \\ A^k{}_j & A^k{}_k \end{vmatrix} = -2\mu_2\,, \end{aligned}$$

according to previous observation.

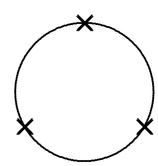
FEYNMAN RULES FOR THE INVARIANTS

Fact. *For each μ_j there is a formula in terms of traces of powers of $\underline{A}$. (Each term is a homogeneous polynomial of degree j in the matrix elements of $\underline{A}$.)*

These formulas enable one to avoid complicated and ugly alternative formulas in terms of various minors of the matrix A. The formulas can be stated in a fashion reminiscent of the *Feynman diagrams* used by theoretical physicists to express the terms in various perturbative expansions:

Algorithm: Each potential *factor* is represented by a ring diagram, where each "$\times$" (*vertex*) represents a matrix A and the connecting lines represent sums over adjacent indices.

$$\operatorname{tr}(\underline{A}^3) = \sum_{jkl} A^j{}_k A^k{}_l A^l{}_j =$$



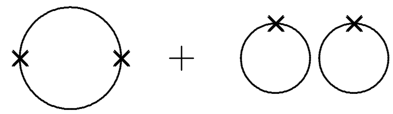
Each *term* in μ_j contains j vertices and any number of factors consistent with that constraint. The coefficient of the term is

$$\frac{(-1)^{j+p}}{\Bigl(\prod_{i=1}^{p} n_i\Bigr)\Bigl(\prod_{k=1}^{\max n_i} m_k!\Bigr)}$$

where $p =$ number of rings in the diagram,
$n_i =$ number of vertices in ring i,
$m_k =$ number of rings containing k vertices.

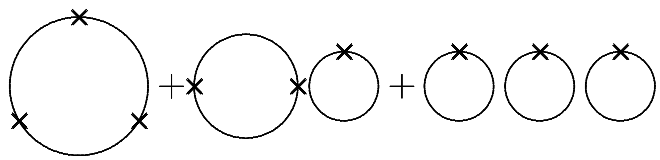
$$\left(\text{Thus}\ \sum_{k=1}^{\max n_i} m_k = p, \qquad \sum_{k=1}^{\max n_i} k\, m_k = j = \sum_{i=1}^{p} n_i\,.\right)$$

Thus:

$$\mu_2 =$$ 

$$\equiv \frac{-1}{2 \cdot 0! \cdot 1!} \text{tr } (A^2) + \frac{1}{1 \cdot 1 \cdot 2!} (\text{tr } A)^2$$
$$= \tfrac{1}{2}\big[-\text{tr } (A^2) + (\text{tr } A)^2\big],$$

in agreement with our previous result.

$$\mu_3 =$$ 

$$\equiv \frac{\text{tr } (A^3)}{3 \cdot (1 \cdot 1 \cdot 1)} \qquad + \qquad \frac{-\text{tr } (A^2) \text{ tr } A}{(2 \cdot 1) \cdot (1 \cdot 1)} \qquad + \qquad \frac{(\text{tr } A)^3}{(1 \cdot 1 \cdot 1) \cdot 3!}$$
$$= \frac{1}{3} \text{ tr } (A^3) - \frac{1}{2} \text{ tr } (A^2) \text{ tr } A + \frac{1}{6} (\text{tr } A)^3.$$

The more conventional formula for μ_3 would involve 3×3 minors of A.

Corollary. *This formula for* μ_j *is independent of* N *(the dimension). For* $j > N$, μ_j *must be* 0, *so one can conclude, e.g., that*

$$\frac{1}{3} \text{ tr } (A^3) - \frac{1}{2} \text{ tr } (A^2) \text{ tr } A + \frac{1}{6} (\text{tr } A)^3 = 0$$

for all 2×2 *matrices. Also,* $\mu_N = \det A$, *so, e.g.,*

$$\det A = \frac{1}{2}\big[-\text{tr } (A^2) + (\text{tr } A)^2\big], \quad \forall\, 2 \times 2 \textit{ matrices},$$
$$\det A = \frac{1}{3} \text{ tr } (A^3) - \frac{1}{2} \text{ tr } (A^2) \text{ tr } A + \frac{1}{6} (\text{tr } A)^3, \quad \forall\, 3 \times 3,$$

etc. (Verifying these last few formulas directly is a good exercise.)

Proof of algorithm: postponed till after treatment of Jordan canonical form.

Now return to the characteristic equation, $\det\,(\underline{A}-\lambda)=0$. **Assume $\mathcal{F}$ is $\mathbf{C}$.** Then there is a factorization

$$\det\,(\underline{A}-\lambda)=(-1)^N(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\,\cdots\,(\lambda-\lambda_L)^{d_L}$$

(where $\sum_{\nu=1}^{L} d_\nu = N$). The λ_ν are the eigenvalues: $\{\lambda_\nu\}_{\nu=1}^{L}=\sigma(\underline{A})$.

$d_\nu \equiv$ *algebraic multiplicity* of the eigenvalue λ_ν .

Expanding, and recalling the notation

$$\det(\underline{A}-\lambda)\equiv\sum_{j=1}^{N}(-1)^{N-j}\mu_j\lambda^{N-j},$$

we see that

$$\begin{gathered}\mu_1=d_1\lambda_1+d_2\lambda_2+\cdots d_L\lambda_L\,,\\ \mu_2=\lambda_1\lambda_2+{\lambda_1}^2\,[\text{if } {\scriptstyle d_1\geq 2}]\;+\;\cdots\;=\text{sum of all possible pairs},\\ \cdots,\\ \mu_N={\lambda_1}^{d_1}{\lambda_2}^{d_2}\,\cdots\,{\lambda_L}^{d_L}.\end{gathered}$$

Remark: It is essential to understand that L and N are different things! L is the number of *distinct* eigenvalues of A; it may be equal to N, but it also may be less. When I want to refer to the list of *distinct* numbers $\sigma(\underline{A})$, I shall index them with *Greek* letters: $\{\lambda_\nu\}_{\nu=1}^{L}$. Later, in dealing with an operator $\underline{A}$ with a matrix such as

$$\begin{pmatrix}2&0&0\\0&2&0\\0&0&3\end{pmatrix},$$

we shall want to list the eigenvalues corresponding to a basis of linearly independent eigenvectors of $\underline{A}$. This list may contain repetitions: in this example,

$$\lambda_1=2,\quad\lambda_2=2,\quad\lambda_3=3.$$

Thus there is an unavoidable notational ambiguity; we *must* retain the option to index the eigenvalues distinctly in some contexts, but according to their multiplicities in other contexts. In addition to stating in words, whenever there is a chance of confusion, which convention is in force, I shall use a *Latin* index whenever the list of eigenvalues is associated with a basis and hence may have repetitions: $\{\lambda_j\}_{j=1}^{N}$.

Let's check the results of the paragraph before the Remark against our earlier results. The simplest matrix with eigenvalues $\{\lambda_j\}_{j=1}^N$ — note that we are changing point of view in midstream, but in accordance with the Greek–Latin convention of the Remark! — is

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

We see that

$$\mu_1 = \mathrm{tr}\ D = \sum \lambda_j\,, \qquad \mu_N = \det\ D = \prod \lambda_j\,,$$

$$\begin{aligned} \mu_2 &= \text{sum of diagonal } 2\times 2 \text{ minors} \\ &= \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3 + \cdots \end{aligned}$$

or, alternatively,

$$\begin{aligned} \mu_2 &= \frac{1}{2}\bigl[(\mathrm{tr}\ D)^2 - \mathrm{tr}\ (D^2)\bigr] \\ &= \frac{1}{2}\bigl[(\lambda_1 + \lambda_2 + \cdots)^2 - ({\lambda_1}^2 + {\lambda_2}^2 + \cdots)\bigr] \\ &= \lambda_1\lambda_2 + \cdots . \end{aligned}$$

All this is consistent with what we knew already.

The functions we are dealing with here are called the *elementary symmetric polynomials* in the variables $\lambda_1 \ldots, \lambda_N$. In general, we have $\mu_j = \sum \lambda_{i_1} \cdots \lambda_{i_j}$, where the sum is over all lists of j distinct indices i_k , arranged in their natural order. This is clear upon contemplation of the process of calculating the determinant of the diagonal matrix $D - \lambda$.

The Cayley-Hamilton theorem

Theorem 26.1. *A matrix (or endomorphism) satisfies its own characteristic equation:*

$$\sum_{j=1}^{N} (-1)^{N-j} \mu_j \underline{A}^{N-j} = \underline{0}.$$

($\underline{A}^n$ means, of course, $\underbrace{\underline{A}\,\underline{A}\cdots\underline{A}}_{n \text{ factors}}$.)

Proof: See Bowen & Wang, pp. 154-155. Change "substitute (26.16) and (26.15) into (26.14)" to " substitute (26.16) into (26.15)". (After all, (26.14) is what we are trying to prove, not a given!)

Another proof will emerge later.