Linear Algebra

David Pierce

April 15, 2019

Matematik Bölümü
Mimar Sinan Güzel Sanatlar Üniversitesi
mat.msgsu.edu.tr/~dpierce/
polytropy.com

Contents

Introduction

References for these notes include Hoffman and Kunze [1], Koç [2], Lang [3, 4], and Roman [5], but I may not follow them closely.

Since in set theory the letter ω denotes the set $\{0, 1, 2, \dots\}$ of natural numbers, I let $\mathbb{N}$ denote the set $\{1, 2, 3, \dots\}$ of counting numbers. For notational convenience, each n in $\mathbb{N}$ is the set $\{0, \dots, n-1\}$, which has n elements. The expressions $i < n$ and $i \in n$ are interchangeable.

An expression like

$$\bigwedge_{i<n} \varphi(i)$$

means $\varphi(i)$ holds whenever $i < n$; that is,

$$i < n \implies \varphi(i).$$

The notation $f\colon A \to B$ is to be read as a sentence, "f is a function from A to B."

1 Determinants

1.1 Matrix multiplication

The structures $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, and $\mathbb{Z}/(n)$, where $n \in \mathbb{N}$, where

$$\mathbb{N} = \{x \in \mathbb{Z} \colon x > 0\},$$

are *commutative rings.*

For us, a **ring** will be a structure $(R, \cdot, 1)$, where

1) R is an abelian group, written additively,
2) $\cdot$ is a **multiplication** on R, that is, a binary operation on R that distributes from each side over addition,
3) $\cdot$ is associative, and
4) 1 is a two-sided identity with respect to $\cdot$.

We usually write $(R, \cdot, 1)$ as R.

A **unit** of a ring is an **invertible** element, that is, an element with a left inverse and a right inverse. When these one-sided inverses exist, they are equal. The units of a ring R compose a multiplicative group, denoted by

$$R^{\times}.$$

A ring is **commutative** if its multiplication is commutative. We gave examples above. For an example of a group of units, we note that, for all n in $\mathbb{N}$,

$$|\mathbb{Z}/(n)^{\times}| = |x \in \mathbb{Z}/(n) \colon \gcd(x, n) = 1\}| = \phi(n).$$

A commutative ring R is a **field** if $R^\times = R \smallsetminus \{0\}$. If p is prime, then $\mathbb{Z}/(p)$ is the field $\mathbb{F}_p$, and

$$\mathbb{F}_p{}^\times \cong \mathbb{Z}_{p-1},$$

where in general $\mathbb{Z}_n$ is the cyclic group of order n, and $\mathbb{Z}/(n)$ means $(\mathbb{Z}_n, \cdot, 1)$.

In this chapter, we shall work with an arbitrary commutative ring K. The definition of a **module** over K is the same as the definition of a vector space, when K is a field. An abelian group is a module over $\mathbb{Z}$.

If $(m, n) \in \mathbb{N} \times \mathbb{N}$, then $K^{m\times n}$ and K^n are modules over K, and

$$(X, \boldsymbol{y}) \mapsto X\boldsymbol{y} \colon K^{m\times n} \times K^n \to K^m,$$

defined as follows.

If Ω is a set, we denote by

$$K^\Omega$$

the K-module of functions from Ω to K. This defines K^n when we understand n as the n-element set $\{0, \dots, n-1\}$. An arbitrary element of K^n is one of

$$(a^0, \dots, a^{n-1}), \qquad (a^j : j \in n), \qquad \boldsymbol{a}.$$

The superscripts are row numbers, when we think of $\boldsymbol{a}$ as the $1 \times n$ matrix

$$\begin{pmatrix} a^0 \\ \vdots \\ a^{n-1} \end{pmatrix}.$$

Many persons understand K^n as $K^{[n]}$, where $[n]$ is the set $\{1, \dots, n\}$ with n elements. What is important is that the entries of an element of K^n be functions into K from a linearly ordered set with n elements.

An element A of $K^{m\times n}$ is a matrix of m rows and n columns, having entries a^i_j from K, where $i \in m$ and $j \in n$, so

$$A = \begin{pmatrix} a^0_0 & \cdots & a^0_{n-1} \\ \vdots & \ddots & \vdots \\ a^{m-1}_0 & \cdots & a^{m-1}_{n-1} \end{pmatrix} = (a^i_j)^{i\in m}_{j\in n}.$$

If one prefers, one may work instead with elements of $E^{[m]\times[n]}$, and one may write a_{ij} for a^i_j. If also $\boldsymbol{b} \in K^n$, we define

$$A\boldsymbol{b} = \left(\sum_{j\in n} a^i_j b^j : i \in m\right), \tag{1.1}$$

an element of K^m. As in (1.1) with j, when an index appears twice, once raised and once lowered, it is usually being summed over. When $j \in n$, we define

$$\mathbf{e}_j = (\delta^i_j : i \in n) \tag{1.2}$$

in the module K^n, where

$$\delta^i_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \tag{1.3}$$

Then

$$A\mathbf{e}_j = \left(\sum_{k\in n} a^i_k \delta^k_j : i \in n\right) = (a^i_j : i \in n) = \boldsymbol{a}_j, \tag{1.4}$$

this being column j of A. If $\boldsymbol{b} \in K^n$, then

$$\boldsymbol{b} = \sum_{j\in n} b^j \mathbf{e}_j. \tag{1.5}$$

We denote by

$$\tau_A$$

the function $\boldsymbol{x} \mapsto A\boldsymbol{x}$ from K^n to K^m.

To say that a function φ from K^n to K^m is a **linear transformation** means that φ is a homomorphism of modules over K, that is,

$$\varphi(\boldsymbol{b}+\boldsymbol{c}) = \varphi(\boldsymbol{b}) + \varphi(\boldsymbol{c}), \qquad \varphi(d \cdot \boldsymbol{b}) = d \cdot \varphi(\boldsymbol{b}).$$

The linear transformations from K^n to K^m compose a module over K denoted by

$$\mathscr{L}(K^n, K^m).$$

Theorem 1. $X \mapsto \tau_X \colon K^{m\times n} \cong \mathscr{L}(K^n, K^m)$.

Proof. We have to check that

(1) $\tau_A \in \mathscr{L}(K^n, K^m)$ for each A in $K^{m\times n}$;

(2) $X \mapsto \tau_X$ is a homomorphism;

(3) if $\tau_A = 0$, then $A = 0$;

(4) every member of $\mathscr{L}(K^n, K^m)$ is τ_A for some A in $K^{m\times n}$.

Each step in the verification of the first two points uses the definition of a K-module or a property of K as a commutative ring. If $\tau_A = 0$, this means in each case $\mathbf{0} = A\mathbf{e}_j$, which is column j of A by (1.4); so $A = 0$.

Finally, since each τ_A is linear, from (1.4) and (1.5) we have

$$A\boldsymbol{b} = \sum_{j\in n} b^j \boldsymbol{a}_j.$$

If $T \in \mathscr{L}(K^n, K^m)$, by defining

$$T\mathbf{e}_j = \boldsymbol{a}_j,$$

we obtain A, and then

$$T = \tau_A. \qquad \square$$

If still $A \in K^{m\times n}$, and now also $C \in K^{n\times s}$, then we define

$$AC = \left(\sum_{j\in n} a^i_j c^j_k\right)^{i\in m}_{k\in s}, \tag{1.6}$$

an element of $K^{m\times s}$. We shall let M denote the special case $K^{n\times n}$, which is closed under matrix multiplication. We have

$$\mathrm{I}A = A = A\mathrm{I},$$

where

$$\mathrm{I} = (\delta^i_j)^{i\in n}_{j\in n}. \tag{1.7}$$

Theorem 2. *When $A \in K^{m\times n}$ and $C \in K^{n\times s}$, then*

$$\tau_{AC} = \tau_A \circ \tau_C.$$

Thus for any matrices A, B, and C for which either of the products $(AB)C$ and $A(BC)$ is defined, then both are defined, and they are equal. In particular, the structure $(M, \cdot, \mathrm{I})$ is a ring, and $X \mapsto \tau_X$ from M to $\mathscr{L}(K^n, K^n)$ is an isomorphism of rings.

1.2 Determinants

We use the possibility of Gauss–Jordan elimination to motivate the so-called Leibniz formula (1.19) for the determinant.

1.2.1 Desired Properties

Let M be the ring $K^{n\times n}$. We want to define a **determinant** function,

$$X \mapsto \det X,$$

from M to K so that

$$\det X \in K^\times \iff X \in M^\times. \tag{1.8}$$

If K is the two-element field $\mathbb{F}_2$, then (1.8) is equivalent to

$$\det X = \begin{cases} 1, & \text{if } X \in M^\times, \\ 0, & \text{otherwise.} \end{cases} \tag{1.9}$$

Moreover, with this definition,

$$\det(XY) = \det X \det Y. \tag{1.10}$$

However, over any K, we also want

$$\det X = f\bigl(x^i_j \colon (i,j) \in n \times n\bigr) \tag{1.11}$$

for some *polynomial* f (namely an element of the free abelian group generated by products of the variables x^i_j). In general then, (1.9) will fail. We still want (1.10) to hold, and this and (1.8) imply

$$\det \mathrm{I} = 1. \tag{1.12}$$

1.2.2 Additional properties

In seeking a determinant function satisfying (1.8), (1.10), and (1.11), and therefore (1.12), we consider what we know about $M^\times$. An element A of M is in $M^\times$ just in case A is row-equivalent to I. This means, for some *elementary* matrices E_i,

$$A = E_1 \cdots E_n \mathrm{I}. \tag{1.13}$$

Thus, if (1.10) and (1.12) hold, then $\det A$ will determined by the $\det E_i$.

We recall that an **elementary matrix** is the result of applying to I an **elementary row operation.** If Φ is such, then

$$\Phi(I)A = \Phi(A).$$

Here Φ does one of the following:

1) add to one row another row, scaled by some a in K;
2) interchange two rows;
3) scale a row by an element s of $K^{\times}$.

Let us denote the specific instance of Φ respectively by

$$\Phi_a, \qquad \Psi, \qquad \Theta_s.$$

We do not specify the row or rows involved. We draw the following conclusions about determinants.

1. If (1.11) is to hold, then, for some single-variable polynomial f,

$$\det \Phi_a(\mathrm{I}) = f(a).$$

If also (1.10) is to hold, then, since

$$\Phi_a(\mathrm{I}) \cdot \Phi_b(\mathrm{I}) = \Phi_{a+b}(\mathrm{I}),$$

we must have

$$f(a) \cdot f(b) = f(a+b).$$

In particular, $f(x)^n = f(nx)$ for all n in $\mathbb{N}$, and so, since $f \neq 0$, we must have

$$\det \Phi_a(\mathrm{I}) = 1. \tag{1.14}$$

2. If, again, (1.10) is to hold, then, since

$$\Psi(\mathrm{I}) \cdot \Psi(\mathrm{I}) = \mathrm{I},$$

we should have $\det \Psi(\mathrm{I}) = \pm 1$; we choose

$$\det \Psi(\mathrm{I}) = -1. \tag{1.15}$$

3. If, again (1.11) is to hold, then, for some single-variable polynomial g,
$$\det \Theta_s(\mathrm{I}) = g(s).$$
If also (1.10) is to hold, then, since
$$\Theta_s(\mathrm{I}) \cdot \Theta_t(\mathrm{I}) = \Theta_{st}(\mathrm{I}),$$
we must have
$$g(s) \cdot g(t) = g(st).$$
In particular, $g(x)^n = g(x^n)$, so $\det \Theta_s(\mathrm{I})$ must be a power of s; we choose
$$\det \Theta_s(\mathrm{I}) = s. \tag{1.16}$$

The definitions, or choices, (1.14), (1.15), and (1.16) will follow if $X \mapsto \det X$ is an *alternating multilinear form.*

We can understand any module $K^{m\times n}$ as $(K^m)^n$ or $(K^n)^m$, treating an element A as one of
$$\big((a^i_j \colon i \in m) \colon j \in n\big), \qquad \big((a^i_j \colon j \in n) \colon i \in m\big).$$
Given a module V over K and n in $\mathbb{N}$, we can form the module V^n. For each k in n, we let π_k the function from V^n to V given by
$$\pi_k(\boldsymbol{x}_j \colon j \in n) = \boldsymbol{x}_k.$$
Suppose now
$$\varphi \colon V^n \to K.$$
Given k in n and a function $j \mapsto \boldsymbol{a}_j$ from $n \smallsetminus \{k\}$ to V, we let ι be the function from V to V^n given by the rule that, for each j in n,
$$\pi_j(\iota(\boldsymbol{x})) = \begin{cases} \boldsymbol{x}, & \text{if } j = k,\\ \boldsymbol{a}_j, & \text{if } j \in n \smallsetminus \{k\}.\end{cases}$$

If the function $\boldsymbol{x} \mapsto \varphi(\iota(\boldsymbol{x}))$ is always linear, then φ itself is a **multilinear form,** specifically an n-**linear** form, on V. If, further, whenever $i < j < n$,

$$\boldsymbol{x}_i = \boldsymbol{x}_j \implies \varphi(\boldsymbol{x}_k \colon k \in n) = 0,$$

then φ is **alternating** as a multilinear form.

We let the group of permutations of a set Ω be

$$\mathrm{Sym}(\Omega).$$

If Ω is finite, then $\mathrm{Sym}(\Omega)$ is generated by transpositions. If $\sigma \in \mathrm{Sym}(n)$, we define

$$\mathrm{sgn}(\sigma) = (-1)^{|(i,j)\in n\times n\colon i<j \;\&\; \sigma(i)>\sigma(j)\}|}, \tag{1.17}$$

one of the elements of $\mathbb{Z}^{\times}$.

Theorem 3. *For every n in $\mathbb{N}$, the function $\xi \mapsto \mathrm{sgn}(\xi)$ on* $\mathrm{Sym}(n)$

1) is given by

$$\mathrm{sgn}(\sigma) = \prod_{i\in j\in n} \frac{\sigma(i)-\sigma(j)}{i-j}, \tag{1.18}$$

2) is a homomorphism onto $\mathbb{Z}^{\times}$, and

3) takes every transposition to -1.

Proof. 1. Since

$$\prod_{i\in j\in n} \frac{\sigma(i)-\sigma(j)}{i-j} = \frac{\prod_{i\in j\in n}(\sigma(i)-\sigma(j))}{\prod_{i\in j\in n}(i-j)} = \pm 1,$$

(1.17) follows from (1.18).

2. Note

$$\begin{aligned}
\operatorname{sgn}(\tau\sigma) &= \prod_{i\in j\in n} \frac{\tau\sigma(i)-\tau\sigma(j)}{i-j}\\
&= \prod_{i\in j\in n}\left(\frac{\tau\sigma(i)-\tau\sigma(j)}{\sigma(i)-\sigma(j)}\cdot\frac{\sigma(i)-\sigma(j)}{i-j}\right)\\
&= \prod_{i\in j\in n}\frac{\tau(i)-\tau(j)}{i-j}\cdot\operatorname{sgn}(\sigma)=\operatorname{sgn}(\tau)\cdot\operatorname{sgn}(\sigma).
\end{aligned}$$

3. Letting

$$\tau = (0\ \ 1),$$

since every transposition is $\sigma^{-1}\cdot\tau\cdot\sigma$ for some σ, it is enough to note that

$$\operatorname{sgn}(\tau) = -1,$$

since

$$\frac{\tau(i)-\tau(j)}{i-j}\begin{cases} >0, & \text{when } (i,j)\neq(0,1),\\ <0, & \text{when } (i,j)=(0,1).\end{cases}$$

□

An element σ of $\operatorname{Sym}(n)$ is **even** if $\operatorname{sgn}(\sigma)=1$; this means σ is a product of an even number of transpositions. The even permutations compose the subgroup of $\operatorname{Sym}(n)$ denoted by

$$\operatorname{Alt}(n).$$

Theorem 4. *For any module V over K, for any n in $\mathbb{N}$, for any n-linear form φ on V, for each σ in* $\operatorname{Sym}(n)$*,*

$$\varphi(\boldsymbol{x}_{\sigma(j)}\colon j\in n)=\operatorname{sgn}(\sigma)\cdot\varphi(\boldsymbol{x}_j\colon j\in n).$$

Proof. Every permutation of a finite set being a product of transpositions, we need only prove the claim when $n = 2$ and σ is the nontrivial permutation of 2. Assuming

$$\boldsymbol{x} = \boldsymbol{y} \implies \varphi(\boldsymbol{x}, \boldsymbol{y}) = 0,$$

we have $0 = \varphi(\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y})$, but the latter is

$$\varphi(\boldsymbol{x}, \boldsymbol{x}) + \varphi(\boldsymbol{x}, \boldsymbol{y}) + \varphi(\boldsymbol{y}, \boldsymbol{x}) + \varphi(\boldsymbol{y}, \boldsymbol{y}),$$

which reduces to $\varphi(\boldsymbol{x}, \boldsymbol{y}) + \varphi(\boldsymbol{y}, \boldsymbol{x})$. □

In particular, if $\sigma \in \mathrm{Alt}(n)$, then

$$\varphi(\boldsymbol{x}_{\sigma(j)} \colon j \in n) = \varphi(\boldsymbol{x}_j \colon j \in n).$$

1.2.3 Existence and uniqueness

Theorem 5. *There is at most one alternating multilinear function $X \mapsto \det X$ from M to K that satisfies* (1.12)*, and if it does exist, it satisfies satisfies* (1.8) *and* (1.10)*.*

Proof. The hypotheses ensure (1.14), (1.15), and (1.16), as well as (1.12). Then (1.10) holds when X is elementary, and therefore it holds for all X, and also (1.8) holds by the analysis (1.13) and since every non-invertible matrix is row-equivalent to one with a zero row. □

We now show that there is at least one function $X \mapsto \det X$ as desired. We define

$$\det X = \sum_{\sigma \in \mathrm{Sym}(n)} \mathrm{sgn}(\sigma) \prod_{i \in n} x^i_{\sigma(i)}. \tag{1.19}$$

Thus (1.11) holds.

Theorem 6. *For all A in M,*

$$\det(A^{\mathrm{t}}) = \det A.$$

Proof. Since $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$, we compute

$$\begin{aligned}\det(A^{\mathrm{t}}) &= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma)\prod_{i\in n} a_i^{\sigma(i)}\\ &= \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma^{-1})\prod_{i\in n} a^i_{\sigma^{-1}(i)},\end{aligned}$$

which is $\det A$. □

Theorem 7. *The function given by* (1.19) *is n-linear and alternating, and satisfies* (1.12).

Proof. By (1.7), since

$$\prod_{i\in n} \delta^i_{\sigma(i)} = 0 \iff \sigma \neq \mathrm{id}_n,$$

(1.12) holds. For multilinearity, Suppose matrices A, B, and C agree everywhere but in some row k, and $a^k_j = s\cdot b^k_j + t\cdot c^k_j$ for each j in n, for some s and t in K. Then

$$\begin{aligned}\det A = \sum_{\sigma\in\operatorname{Sym}(n)} \operatorname{sgn}(\sigma) \prod_{i\in n\smallsetminus\{k\}} a^i_{\sigma(i)}\cdot(s\cdot b^k_{\sigma(k)} + t\cdot c^k_{\sigma(k)})\\ = s\cdot\det B + t\cdot\det C.\end{aligned}$$

Finally, if $i<j<n$, and τ in $\operatorname{Sym}(n)$ transposes i and j, then $\tau^{-1}=\tau$, and $\xi\mapsto\xi\circ\tau$ is a bijection between $\operatorname{Alt}(n)$ and $\operatorname{Sym}(n)\smallsetminus\operatorname{Alt}(n)$, so

$$\det A = \sum_{\sigma\in\operatorname{Alt}(n)}\left(\prod_{k\in n} a^k_{\sigma(k)} - \prod_{k\in n} a^{\tau(k)}_{\sigma(k}\right).$$

If moreover $a^i_k = a^j_k$ for each k in n, then $\det A = 0$. □

1.3 Inversion

We know from Theorems 5 and 7 that (1.8) holds. In particular, if $\det A \in K^{\times}$, then A^{-1} exists in M. We can compute A^{-1} by the reduction in (1.13); but we now develop another method.

As in (1.17), if τ is a bijection from a finite ordered set S to a finite ordered set T, we can define

$$\operatorname{sgn}(\tau) = (-1)^{|(i,j)\in S\times S\colon i<j\ \&\ \sigma(i)>\sigma(j)\}|}.$$

There is a unique isomorphism φ from S to T, and then

$$\varphi^{-1}\circ\tau \in \operatorname{Sym}(S),$$
$$\operatorname{sgn}(\tau) = \operatorname{sgn}(\varphi^{-1}\circ\tau).$$

Suppose now $\sigma \in \operatorname{Sym}(n)$ and $k \in n$. Letting S be $n \smallsetminus \{k\}$ and T be $n \smallsetminus \{\sigma(k)\}$, we can define τ to be the restriction of σ to S, so that τ is a bijection from S to T. Then

$$\frac{\operatorname{sgn}(\sigma)}{\operatorname{sgn}(\tau)} = (-1)^{|\{j\in n\smallsetminus\{k\}\colon j>k \iff \sigma(j)<\sigma(k)\}|}.$$

Theorem 8. *In the notation above,*

$$\frac{\operatorname{sgn}(\sigma)}{\operatorname{sgn}(\tau)} = (-1)^{k+\sigma(k)}.$$

Proof. We may assume $k \leqslant \sigma(k)$. There are at least $\sigma(k) - k$ values of j greater than k and the condition

$$j > k \iff \sigma(j) < \sigma(k) \tag{1.20}$$

is satisfied. For every additional such value, there must be a value less than k for which (1.20) is satisfied. This proves the claim. □

For any (k, ℓ) in $n \times n$, assuming $n > 1$, we let $\hat{A}^k_\ell$ be the matrix that we obtain from A by deleting row k and column ℓ. Formally,

$$\hat{A}^k_\ell = \left(a^{[i,k]}_{[j,\ell]}\right)^{i \in n-1}_{j \in n-1},$$

where

$$[i,k] = \begin{cases} i, & \text{if } i < k, \\ i+1, & \text{if } k \leqslant i. \end{cases}$$

Theorem 9. *For any k in n,*

$$\det X = \sum_{j \in n} (-1)^{k+j} x^k_j \det \hat{X}^k_j.$$

Proof. We group the terms in (1.19), which are indexed by σ in $\mathrm{Sym}(n)$, according to the value of $\sigma(k)$:

$$\begin{aligned} \det X &= \sum_{j \in n} \sum_{\substack{\sigma \in \mathrm{Sym}(n) \\ \sigma(k)=j}} \operatorname{sgn}(\sigma) \prod_{i \in n} x^i_{\sigma(i)} \\ &= \sum_{j \in n} x^k_j \sum_{\substack{\sigma \in \mathrm{Sym}(n) \\ \sigma(k)=j}} \operatorname{sgn}(\sigma) \prod_{\substack{i \in n \\ i \neq k}} x^i_{\sigma(i)} \\ &= \sum_{j \in n} (-1)^{k+j} x^k_j \det \hat{X}^k_j \end{aligned}$$

by Theorem 8. □

We now define the operation $X \mapsto \operatorname{adj}(X)$ on M by

$$\operatorname{adj}(A) = \left((-1)^{i+j} \det \hat{A}^j_i\right)^{i \in n}_{j \in n}.$$

This is the **adjugate** of A.

Theorem 10. *For all A in M,*

$$A\,\mathrm{adj}(A) = \det A \cdot \mathrm{I}.$$

Proof. By Theorem 9, if $A\,\mathrm{adj}(A) = B$, then b^i_j is the determinant of the matrix that we obtain from A by replacing row j with row i. This determinant is

- $\det A$, if $i = j$;
- 0, if $i \neq j$, since $X \mapsto \det X$ is alternating. □

Theorem 11. *If* $\det A \in K^\times$*, then*

$$A^{-1} = (\det A)^{-1} \cdot \mathrm{adj}(A).$$

Proof. Assuming $\det A \in K^\times$, if we denote $(\det A)^{-1} \cdot \mathrm{adj}(A)$ by B, then by Theorem 10,

$$AB = \mathrm{I}.$$

Since A^{-1} does exist, we have

$$A^{-1} = A^{-1}(AB) = (A^{-1}A)B = \mathrm{I}B = B. \qquad \square$$

2 Polynomials

2.1 Characteristic values

We henceforth suppose K is a field; still M is $K^{n\times n}$. For any A in M, an element λ of K is a **characteristic value** or **eigenvalue** of A if, for some $\boldsymbol{b}$ in K^n,

$$A\boldsymbol{b} = \lambda \cdot \boldsymbol{b}. \tag{2.1}$$

In this case, $\boldsymbol{b}$ is a **characteristic vector** or **eigenvector** of A associated with λ. Rewriting (2.1) as

$$(A - \lambda \cdot \mathrm{I})\boldsymbol{b} = \mathbf{0}$$

shows that the characteristic values of A are precisely the zeroes of the polynomial

$$\det(A - x \cdot \mathrm{I}),$$

which is called the **characteristic polynomial** of A.

If λ is indeed a characteristic value of A, then the null-space of $A - \lambda \cdot \mathrm{I}$ is the **characteristic space** or **eigenspace** of A associated with λ.

Theorem 12. *Eigenvectors corresponding to distinct eigenvalues of any matrix are linearly independent.*

Proof. We prove the claim by induction on the number of eigenvectors. The empty set of eigenvectors is trivially linearly independent. Suppose $(\boldsymbol{v}_i : i < k)$ is linearly independent, each $\boldsymbol{v}_i$ being an eigenvector of A with associated eigenvalue λ_i, the λ_i being distinct. Let $\boldsymbol{v}_k$ be a an eigenvector associated with a new eigenvalue, λ_k. If

$$\sum_{i \leqslant k} x^i \boldsymbol{v}_i = \mathbf{0}, \tag{2.2}$$

then

$$\begin{aligned}\mathbf{0} = (A - \lambda_k \cdot \mathrm{I}) \sum_{i<m+1} x^i \boldsymbol{v}_i &= \sum_{i \leqslant k} (\lambda_i - \lambda_k) x^i \boldsymbol{v}_i \\ &= \sum_{i<k} (\lambda_i - \lambda_k) x^i \boldsymbol{v}_i,\end{aligned}$$

so $x^i = 0$ when $i < k$, and then also $x^k = 0$ by (2.2). □

If A in M has n linearly independent eigenvectors $\boldsymbol{b}_i$, each associated with an eigenvalue λ_i (possibly not distinct), then the eigenvectors are the columns of an element B of $M^\times$, and

$$AB = BL,$$

where

$$\ell^i_j = \begin{cases} \lambda_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}$$

or in short

$$L = \operatorname{diag}(\lambda_i : i \in n),$$

a **diagonal matrix.** Thus

$$B^{-1}AB = \operatorname{diag}(\lambda_i : i \in n),$$

and in particular A is **diagonalizable.**

It will be useful to recall that every matrix B in $M^\times$ is the change-of-basis matrix from the basis $({}_j\colon j \in n)$ of K^n consisting of the columns of B to the standard basis of K^n.

Every matrix of the form $P^{-1}AP$ for *some* P in $M^\times$ is **similar** to A (in group theory one says *conjugate*). Similarity of matrices is an equivalence relation, as is row-equivalence (mentioned first on page 9); but they are different relations. We want to characterize the diagonalizable matrices.

A matrix A in M is **triangular** if

$$\bigwedge_{j<i<n} a^i_j = 0. \tag{2.3}$$

A matrix similar to a triangular matrix is **triangularizable.**

Theorem 13. *A matrix A in M is triangularizable just in case, for some B in $M^\times$,*

$$\bigwedge_{j\in n} A\boldsymbol{b}_j \in \operatorname{span}\{\boldsymbol{b}_0, \dots, \boldsymbol{b}_j\}; \tag{2.4}$$

and in this case $B^{-1}AB$ is triangular.

Proof. The condition (2.3) on A for being triangular means precisely

$$\bigwedge_{j\in n} A\mathbf{e}_j = \sum_{i=0}^{j} a^i_j \mathbf{e}_i, \tag{2.5}$$

and thus that I is a matrix B as in the statement of the theorem. If $B^{-1}AB$ is triangular, then putting this matrix in place of A in (2.5) yields (2.4). Conversely, if B is as in the statement, then we can write (2.4) as

$$\bigwedge_{j\in n} AB\mathbf{e}_j \in \operatorname{span}\{B\mathbf{e}_0, \dots, B\mathbf{e}_j\},$$

and then

$$\bigwedge_{j\in n} B^{-1}AB\mathbf{e}_j \in \operatorname{span}\{\mathbf{e}_0,\dots,\mathbf{e}_j\},$$

so $B^{-1}AB$ is triangular. $\square$

Theorem 14. *Every matrix in M is triangularizable over an algebraically closed field.*

Proof. Given A in M, assuming K is algebraically closed, so that the characteristic polynomial of A has at least one zero, and therefore A has at least one eigenvector, we extend this to a basis of K^n that satisfies (2.4). Doing this will be enough, by Theorem 13.

We use induction on n. The claim is trivial when $n = 1$. Suppose it holds when $n = m$. Now let $n = m + 1$ and $A \in M$. There is a basis $(\boldsymbol{p}_0,\dots,\boldsymbol{p}_m)$ of K^n such that $\boldsymbol{p}_0$ is an eigenvector. Thus the basis satisfies the first conjunct of (2.4). We could satisfy the remaining conjuncts, by the inductive hypothesis, if we had

$$\bigwedge_{j=1}^{m} A\boldsymbol{p}_j \in \operatorname{span}\{\boldsymbol{p}_1,\dots,\boldsymbol{p}_m\}.$$

However, we may not actually have this. Nonetheless, there are matrices B and C such that

$$\tau_B\left(\sum_{i=0}^{m} x^i\boldsymbol{p}_i\right) = x_0\boldsymbol{p}_0, \quad \tau_C\left(\sum_{i=0}^{m} x^i\boldsymbol{p}_i\right) = \sum_{i=1}^{m} x^i\boldsymbol{p}_i. \qquad (2.6)$$

In words,

- τ_C is an endomorphism of $\operatorname{span}\{\boldsymbol{p}_1,\dots,\boldsymbol{p}_m\}$, and therefore so is τ_{CA};

- τ_B is a homomorphism from K^n to $\operatorname{span}\{\boldsymbol{p}_0\}$, and therefore so is τ_{BA}.

Now $\operatorname{span}\{\boldsymbol{p}_1, \dots, \boldsymbol{p}_m\}$ has a basis $(\boldsymbol{v}_1 \dots, \boldsymbol{v}_m)$ such that

$$\bigwedge_{j=1}^{m} CA\boldsymbol{v}_j \in \operatorname{span}\{\boldsymbol{v}_1 \dots, \boldsymbol{v}_j\},$$

by inductive hypothesis. Therefore now

$$\bigwedge_{j=1}^{m} (BA + CA)\boldsymbol{v}_j \in \operatorname{span}\{\boldsymbol{v}_0, \dots, \boldsymbol{v}_j\}.$$

From all of (2.6),

$$\tau_B + \tau_C = \mathrm{id}_{K^n},$$

and so

$$\bigwedge_{j=1}^{m} A\boldsymbol{v}_j \in \operatorname{span}\{\boldsymbol{v}_0, \dots, \boldsymbol{v}_j\}.$$

Finally, since $\boldsymbol{v}_0$ is an eigenvector of A,

$$\bigwedge_{j=0}^{m} A\boldsymbol{v}_j \in \operatorname{span}\{\boldsymbol{v}_0, \dots, \boldsymbol{v}_j\}.$$

Thus we have (2.4). This completes the induction. □

We can write out the foregoing proof entirely in terms of matrices as follows. We have

$$P^{-1}AP = \left(\begin{array}{c|c} \lambda & \boldsymbol{a} \\ \hline \mathbf{0} & D \end{array}\right)$$

for some $m \times m$ matrix D, where λ is the eigenvalue associated with $\boldsymbol{p}_0$. We choose B and C so that

$$P^{-1}BP = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & 0 \end{array}\right), \qquad P^{-1}CP = \left(\begin{array}{c|c} 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathrm{I} \end{array}\right).$$

Then

$$P^{-1}CAP = P^{-1}CPP^{-1}AP = \left(\begin{array}{c|c} 0 & \mathbf{0} \\ \hline \mathbf{0} & \mathrm{I} \end{array}\right) \left(\begin{array}{c|c} \lambda & \boldsymbol{a} \\ \hline \mathbf{0} & D \end{array}\right)$$
$$= \left(\begin{array}{c|c} 0 & \mathbf{0} \\ \hline \mathbf{0} & D \end{array}\right)$$

and

$$P^{-1}BAP = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & 0 \end{array}\right) \left(\begin{array}{c|c} \lambda & \boldsymbol{a} \\ \hline \mathbf{0} & D \end{array}\right) = \left(\begin{array}{c|c} \lambda & \boldsymbol{a} \\ \hline \mathbf{0} & 0 \end{array}\right).$$

Therefore

$$P^{-1}BAP + P^{-1}CAP = P^{-1}AP,$$
$$BA + CA = A.$$

By inductive hypothesis, for some Q, $Q^{-1}DQ$ is a triangular matrix T. Then

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right)^{-1} P^{-1}CAP \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right) = \left(\begin{array}{c|c} 0 & \mathbf{0} \\ \hline \mathbf{0} & T \end{array}\right),$$

while

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right)^{-1} P^{-1}BAP \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right)$$
$$= \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q^{-1} \end{array}\right) \left(\begin{array}{c|c} \lambda & \boldsymbol{a}Q \\ \hline \mathbf{0} & 0 \end{array}\right) = \left(\begin{array}{c|c} \lambda & \boldsymbol{a}Q \\ \hline \mathbf{0} & 0 \end{array}\right),$$

and therefore

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right)^{-1} P^{-1}AP \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q \end{array}\right) = \left(\begin{array}{c|c} \lambda & \boldsymbol{a}Q \\ \hline \mathbf{0} & T \end{array}\right),$$

a triangular matrix.

2.2 Polynomial functions of matrices

Although K is a field, the ring M is not commutative when $n > 1$. However, it has commutative sub-rings. Indeed, for every A in M, the smallest sub-ring of M that contains A is commutative. We may denote this sub-ring by

$$K[A].$$

This is also a vector space over K, spanned by the powers I, A, A^2, A^3, . . . , of A. Thus

$$K[A] = \{f(A) \colon f \in K[x]\},$$

where, if

$$f(x) = \sum_{i=0}^{m} f_i x^i \tag{2.7}$$

in $K[x]$ (and x^i is now the power $\prod_{k \in i} x$), we define

$$f(A) = \sum_{i=0}^{n} f_i A^i.$$

If $f(A)$ is the zero matrix, we may say A is a **zero** of f. However, theorems about zeroes in fields may not apply here. For example, since $K[A]$ may have zero divisors, the number of zeroes of f in M may exceed $\deg f$. Indeed, A itself may be a zero divisor, as for example when

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

since then A^2 is the zero matrix. In this case every scalar multiple $b \cdot A$ of A is a zero in $K[A]$ of the polynomial x^2.

2.3 Cayley–Hamilton Theorem

Given A in M, we are going to want to know that A is a zero of some nonzero polynomial over K. Suppose

$$f(x) = \det(x \cdot \mathrm{I} - A), \tag{2.8}$$

the characteristic polynomial of A. The equation remains correct automatically when we replace x with an element of K or of any field that includes K. Note however that, for a matrix B in M, while $f(B) \in M$, we have

$$\det(B\mathrm{I} - A) \in K.$$

Since $A\mathrm{I} - A$ is the zero matrix, we have $\det(A\mathrm{I} - A) = 0$. This observation is not enough to ensure that $f(A)$ is the zero matrix. Nonetheless, we shall show that it is, in two ways.

Theorem 15 (Cayley–Hamilton)**.** *Over any field, every matrix is a zero of its characteristic polynomial.*

First proof. By Theorem 10, with f as in (2.8) we have

$$f(x) \cdot \mathrm{I} = (x \cdot \mathrm{I} - A) \operatorname{adj}(x \cdot \mathrm{I} - A). \tag{2.9}$$

Moreover,

$$f(x) \cdot \mathrm{I} = \sum_{j=0}^{n} x^j \cdot F_j,$$

where, in the notation of (2.7),

$$F = f_j \cdot \mathrm{I}.$$

Likewise, for some matrices B_j in M,

$$\operatorname{adj}(x \cdot \mathrm{I} - A) = \sum_{j=0}^{n-1} x^j \cdot B_j.$$

Thus (2.9) becomes

$$\sum_{j=0}^{n} x^j \cdot f_j \cdot \mathrm{I} = (x \cdot \mathrm{I} - A) \sum_{j=0}^{n-1} x^j \cdot B_j. \tag{2.10}$$

This then will be true when x is replaced by an element of M that commutes with A. Since A is such an element, and the right member of (2.10) becomes 0 when x is replaced with A, the same is true for the left member; but this just means $f(A) = 0$. □

Second proof. Letting f be the characteristic polynomial of A in M as in (2.8), we want to show $f(A) = 0$. Since the determinant function is multiplicative, for every P in $M^{\times}$,

$$\begin{aligned}\det(x \cdot \mathrm{I} - A) &= \det\bigl(P^{-1} \cdot (x \cdot \mathrm{I} - A) \cdot P\bigr) \\ &= \det(x \cdot \mathrm{I} - P^{-1}AP).\end{aligned}$$

By Theorem 14, for some matrix P, $P^{-1}AP$ is a triangular matrix. It does not matter that entries of P may come from the algebraic closure of K, possibly not K itself. We may assume A is triangular. The characteristic polynomial of A is now

$$\prod_{i<n} (x - a^i_i).$$

Since the product is independent of the order of the factors, so is the product $\prod_{i<n}(A - a^i_i \cdot \mathrm{I})$. We have to show that this product is 0. Column j of the product is

$$\prod_{i<n} (A - a^i_i \cdot \mathrm{I})\mathbf{e}_j.$$

However, by (2.5),

$$(A - a^j_j \cdot \mathrm{I})\mathbf{e}_j = A\mathbf{e}_j - a^j_j\mathbf{e}_j = \sum_{i<j} a^j_i \mathbf{e}_i, \tag{2.11}$$

and in particular

$$(A - a^j_j \cdot \mathrm{I})\mathbf{e}_j \in \operatorname{span}\{\mathbf{e}_i \colon i < j\}.$$

By induction then,

$$\prod_{i \leqslant j}(A - a^i_i \cdot \mathrm{I})\mathbf{e}_j = \mathbf{0}.$$

Finally

$$\prod_{i \in n}(A - a^i_i \cdot \mathrm{I})\mathbf{e}_j = \prod_{j<i<n}(A - a^i_i \cdot \mathrm{I}) \prod_{i \leqslant j}(A - a^i_i \cdot \mathrm{I})\mathbf{e}_j = \mathbf{0}. \quad \square$$

2.4 Minimal polynomial

Theorem 16. *For any A in M, the subset*

$$\{f \in K[x] \colon f(A) = 0\}$$

of $K[x]$ is a nonzero ideal.

Proof. The set is easily an ideal. It is nontrivial for containing the characteristic polynomial of A; alternatively, since $\dim M = n^2$, there must be some coefficients f_i, not all 0, for which

$$f_0 + f_1 \cdot A + \cdots + f_{n^2} \cdot A^{n^2} = 0. \quad \square$$

Since $K[x]$ is a principal-ideal domain, the ideal of the theorem has a monic generator, called the **minimal polynomial** of A. This polynomial therefore is a factor of the characteristic polynomial of A. In particular, every zero in K of the minimal polynomial is a zero of the characteristic polynomial.

Theorem 17. *In a field, every zero of the characteristic polynomial of a square matrix over the field is a zero of the minimal polynomial. Hence every irreducible factor of the characteristic polynomial is a factor of the minimal polynomial.*

Proof. A zero of the characteristic polynomial of A is just an eigenvalue of A. Let λ be an eigenvalue, with corresponding eigenvector $\boldsymbol{b}$. Thus

$$\begin{aligned} A\boldsymbol{b} &= \lambda\boldsymbol{b}, \\ A^j\boldsymbol{b} &= \lambda^j\boldsymbol{b}, \\ f(A)\boldsymbol{b} &= f(\lambda)\boldsymbol{b} \end{aligned}$$

for all $f(x)$ in $K[x]$. In particular,

$$f(A) = 0 \implies f(\lambda) = 0.$$

If f is the minimal polynomial of A, then $f(A) = 0$, so $f(\lambda) = 0$. □

Theorem 18. *A square matrix is diagonalizable if and only if its minimal polynomial is the product of distinct linear factors.*

Proof. Suppose A in M is diagonalizable, so that, for some B in $M^{\times}$, for some λ_j in K,

$$AB = B\operatorname{diag}(\lambda_j \colon j \in n).$$

Letting column j of B be $\boldsymbol{b}_j$, we have

$$A\boldsymbol{b}_j = \lambda_j \boldsymbol{b}_j,$$
$$(A - \lambda_j \cdot \mathrm{I})\boldsymbol{b}_j = \mathbf{0}.$$

Letting $m = |\{\lambda_j : j \in n\}|$, we may assume

$$\{\lambda_j : j \in n\} = \{\lambda_i : i \in m\}.$$

For all j in n, we have

$$\prod_{i \in m} (A - \lambda_i \cdot \mathrm{I})\boldsymbol{b}_j = \mathbf{0}.$$

The $\boldsymbol{b}_j$ being linearly independent, letting

$$f(x) = \prod_{i \in m} (x - \lambda_i), \tag{2.12}$$

we conclude $f(A) = 0$, so the minimal polynomial of A is a factor of $f(x)$. (It is the same as $f(x)$, since the λ_i are eigenvectors of A, and each of these must be a zero of the minimal polynomial, by Theorem 17.)

Suppose conversely $f(x)$ as given by (2.12), where again the λ_i are all distinct, is the minimal polynomial of A. In particular, $f(A) = 0$. If $j \in m$, we can define $g_j(x)$ in $K[x]$ by

$$(x - \lambda_j) g_j(x) = f(x). \tag{2.13}$$

The λ_j being distinct, the greatest common divisor of the $g_j(x)$ in $K[x]$ is unity. Since $K[x]$ is a Euclidean domain, by Bézout's Lemma there are $q_j(x)$ in $K[x]$ such that

$$\sum_{j \in m} g_j(x) q_j(x) = 1.$$

Then

$$\sum_{j \in m} g_j(A) q_j(A) = \mathrm{I}.$$

Thus for every $\boldsymbol{v}$ in K^n, when we define

$$g_j(A) q_j(A) \boldsymbol{v} = \boldsymbol{w}_j, \tag{2.14}$$

we have

$$\sum_{j \in m} \boldsymbol{w}_j = \boldsymbol{v}.$$

But then since $f(A) = 0$, from (2.13) and (2.14) we have

$$\mathbf{0} = f(A) q_j(A) \boldsymbol{v} = (A - \lambda_j) \boldsymbol{w}_j,$$

so that $\boldsymbol{w}_j$ belongs to the eigenspace associated with λ_j. In particular, by Theorem 12, there must be n linearly independent eigenvectors, so A is diagonalizable. □

3 Jordan Normal Form

The presentation here is based mainly on Lang [3].

3.1 Cyclic spaces

Supposing λ is an eigenvalue of the $n \times n$ matrix A, we let

$$B_\lambda = A - \lambda \cdot \mathrm{I}. \tag{3.1}$$

If $\boldsymbol{v}_0$ is a corresponding eigenvector, this means

$$\boldsymbol{v}_0 \neq \mathbf{0}, \qquad B_\lambda \boldsymbol{v}_0 = \mathbf{0}.$$

If possible now, let $B_\lambda \boldsymbol{v}_1 = \boldsymbol{v}_0$. Then

$$A\boldsymbol{v}_1 = \lambda \boldsymbol{v}_1 + \boldsymbol{v}_0, \qquad {B_\lambda}^2 \boldsymbol{v}_1 = \mathbf{0}.$$

Suppose, in this way, for some s, when $0 < k < s$,

$$A\boldsymbol{v}_k = \lambda \boldsymbol{v}_k + \boldsymbol{v}_{k-1}, \qquad {B_\lambda}^{k+1} \boldsymbol{v}_k = \mathbf{0}.$$

Then defining P as the $n \times s$ matrix

$$\left(\ \boldsymbol{v}_0 \ \middle| \ \cdots \ \middle| \ \boldsymbol{v}_{s-1} \ \right),$$

we have

$$\begin{aligned} AP &= \left(\ A\boldsymbol{v}_0 \ \middle| \ \cdots \ \middle| \ A\boldsymbol{v}_{s-1} \ \right) \\ &= \left(\ \lambda \boldsymbol{v}_0 \ \middle| \ \boldsymbol{v}_0 + \lambda \boldsymbol{v}_1 \ \middle| \ \cdots \ \middle| \ \boldsymbol{v}_{s-2} + \lambda \boldsymbol{v}_{s-1} \ \right) = PJ, \end{aligned} \tag{3.2}$$

where J is the $s \times s$ matrix

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix}.$$

Theorem 19. *The columns of the matrix P just defined are linearly independent.*

Proof. Writing $\boldsymbol{v}$ for $\boldsymbol{v}_{s-1}$ and B for B_λ, we have

$$P = \left(\, B^{s-1}\boldsymbol{v} \, \middle| \, \cdots \, \middle| \, B^0\boldsymbol{v}. \, \right).$$

We show the columns are linearly independent. Suppose for some scalars c^i,

$$\sum_{i<s} c^i \cdot B^{s-i}\boldsymbol{v} = \mathbf{0}.$$

Then $f(B)\boldsymbol{v} = \mathbf{0}$, where

$$f(x) = \sum_{i<s} c^i x^{s-i}.$$

However, also $g(B)\boldsymbol{v} = \mathbf{0}$, where

$$g(x) = x^s.$$

Letting h be the greatest common factor of f and g, we have

$$h(B)\boldsymbol{v} = \mathbf{0}.$$

Also, $h(x) = x^r$ for some r, where $r \leqslant s$. When $r < s$, we have

$$B^r\boldsymbol{v} = \boldsymbol{v}_{s-1-r},$$

which is not $\mathbf{0}$. Thus $h(x) = x^s$, and therefore $f = 0$. □

In the proof, $\operatorname{span}\{\boldsymbol{v}_k \colon k \in s\}$ is a B-**cyclic** subspace of K^n, because it is, for some one vector $\boldsymbol{v}$, spanned by the vectors $B^k\boldsymbol{v}$. The space is then B-**invariant,** because closed under multiplication by B.

3.2 Direct sums

Suppose V is a vector space over K, and for some m in $\mathbb{N}$, and for each j in n, V_j is a subspace of V. If the homomorphism

$$(v_i \colon i < n) \mapsto \sum_{i<n} v_i$$

from $\prod_{i<n} V_i$ to V is surjective, then V is the **sum** of the subspaces V_i, and we may write

$$V = V_0 + \cdots + V_{n-1} = \sum_{i<n} V_i.$$

If, further, the homomorphism is injective, then V is the **direct sum** of the V_i, and we may write

$$V = V_0 \oplus \cdots \oplus V_{n-1} = \bigoplus_{i<n} V_i.$$

Given B in M, we shall understand

$$\ker B = \{\boldsymbol{x} \in K^n \colon B\boldsymbol{x} = \mathbf{0}\}.$$

Lemma 1. *If f and g in $K[x]$ are co-prime, then for all A in M,*

$$\ker(f(A)g(A)) = \ker f(A) \oplus \ker g(A).$$

Proof. By Bézout's Lemma for some q and r in $K[x]$,

$$qf + rg = 1,$$
$$q(A)f(A) + r(A)g(A) = \mathrm{I}.$$

For all $\boldsymbol{v}$ in K^n then,

$$q(A)f(A)\boldsymbol{v} + r(A)g(A)\boldsymbol{v} = \boldsymbol{v}.$$

Suppose now

$$\boldsymbol{w} \in \ker(f(A)g(A)).$$

Then

$$r(A)g(A)\boldsymbol{w} \in \ker f(A), \qquad q(A)f(A)\boldsymbol{w} \in \ker g(A),$$

and so

$$\boldsymbol{w} \in \ker f(A) + \ker g(A).$$

Conversely, suppose

$$\boldsymbol{u} \in \ker f(A), \qquad \boldsymbol{v} \in \ker g(A).$$

Then

$$\boldsymbol{u} = q(A)f(A)\boldsymbol{u} + r(A)g(A)\boldsymbol{u} \\ = r(A)g(A)\boldsymbol{u} = r(A)g(A)(\boldsymbol{u} + \boldsymbol{v})$$

and likewise

$$\boldsymbol{v} = q(A)f(A)(\boldsymbol{u} + \boldsymbol{v}).$$

This shows $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \boldsymbol{u} + \boldsymbol{v}$ is injective. □

Theorem 20. *If each of some f in $K[x]$ is prime to the others, then for all A in M,*

$$\ker \prod_f f(A) = \bigoplus_f \ker f(A).$$

3.3 Kernels

Suppose A in M has characteristic polynomial f, and K is algebraically closed. Then

$$f = \prod_{j<m} (x - \lambda_j)^{r_j}$$

for some λ_j in K and r_j in $\mathbb{N}$. By the Cayley–Hamilton Theorem,

$$\ker\bigl(f(A)\bigr) = K^n.$$

Letting

$$B_j = A - \lambda_j \cdot \mathrm{I},$$

we have now, by Theorem 20,

$$K^n = \bigoplus_{j<m} \ker\left({B_j}^{r_j}\right). \tag{3.3}$$

Theorem 21. *For all B in M, for all s in $\mathbb{N}$, $\ker(B^s)$ is the direct sum of B-cyclic subspaces.*

Proof. We shall prove the claim for every B-invariant subspace of $\ker(B^s)$. We use induction on the dimension of the subspace. If the dimension is 0, the claim is vacuously true. Suppose V is a B-invariant subspace of $\ker(B^s)$ having positive dimension. Then

$$V \nsubseteq \ker(B^0), \quad \ker(B^0) \subseteq \ldots \subseteq \ker(B^s), \quad V \subseteq \ker(B^s),$$

so for some r,

$$V \nsubseteq \ker(B^{r-1}), \qquad\qquad V \subseteq \ker(B^r).$$

Then

$$VB \subseteq V \cap \ker(B^{r-1}) \subset V.$$

This shows

$$VB \subset V.$$

As an inductive hypothesis, we assume

$$VB = \bigoplus_{i<m} W_i, \tag{3.4}$$

where each W_i is B-cyclic. Then for some $\boldsymbol{w}_i$ in V, for some r_i in $\mathbb{N}$,

$$W_i = \operatorname{span}\{B^j \boldsymbol{w}_i \colon j < r_i\}, \qquad \mathbf{0} = B^{r_i}\boldsymbol{w}_i. \tag{3.5}$$

For some $\boldsymbol{v}_i$ in V,

$$\boldsymbol{w}_i = B\boldsymbol{v}_i. \tag{3.6}$$

Now let

$$V_i = \operatorname{span}\{B^j \boldsymbol{v}_i \colon i \leqslant r_i\}.$$

Then V_i is a B-cyclic space, since $B^{r_i+1}\boldsymbol{v}_i = \mathbf{0}$. We shall show that the sum of the V_i is direct. An arbitrary element of V_i is $f_i(B)\boldsymbol{v}_i$ for some f_i in $K[x]$ such that

$$\deg f_i \leqslant r_i. \tag{3.7}$$

Suppose

$$\mathbf{0} = \sum_{i<m} f_i(B)\boldsymbol{v}_i.$$

Then by (3.6),

$$\mathbf{0} = \sum_{i<m} f_i(B)\boldsymbol{w}_i. \tag{3.8}$$

But then by (3.4),

$$\mathbf{0} = f_i(B)\boldsymbol{w}_i,$$

so by (3.7), and (3.5), and Theorem 19,

$$f_i = c_i x^{r_i}$$

for some c_i in K. In this case, we can write (3.8) as

$$\mathbf{0} = \sum_{i<m} c_i B^{r_i-1} \boldsymbol{w}_i,$$

which implies that each c_i is 0. Thus $f_i = 0$.

Now we can let

$$V' = \bigoplus_{i<m} V_i.$$

Then $V' \subseteq V$. By construction, $V_i B = W_i$, so

$$V'B = W = VB.$$

Therefore

$$V = V' + \ker B.$$

Each element of $\ker B$ constitutes a basis of a one-dimensional B-cyclic space. Then V is the direct sum of some of these spaces, along with the V_i, as desired. □

In the notation of (3.3), there are n_j in $\mathbb{N}$, and then there are $\boldsymbol{v}_{jk}$ in $\ker({B_j}^{r_j})$ and s_{jk} in $\mathbb{N}$ such that

$${B_j}^{s_{jk}-1} \boldsymbol{v}_{jk} \neq \mathbf{0}, \qquad {B_j}^{s_{jk}} \boldsymbol{v}_{jk} = \mathbf{0},$$

and

$$\ker({B_j}^{r_j}) = \bigoplus_{k<n_j} \operatorname{span}\{{B_j}^i \boldsymbol{v}_{jk} \colon i < s_{jk}\}.$$

Now we may let

$$P = \left(\, P_0 \mid \cdots \mid P_{m-1} \, \right),$$

where, for each j in m,

$$P_j = \left(\ P_{j0} \,\middle|\, \cdots \,\middle|\, P_{j,n_j-1}\ \right),$$

where, for each k in n_j,

$$P_{jk} = \left(\ {B_j}^{s_{jk}-1}\boldsymbol{v}_{jk} \,\middle|\, \cdots \,\middle|\, \boldsymbol{v}_{j,k}\ \right).$$

Then PAP^{-1} is a **Jordan normal form** for A. Indeed, by the considerations yielding (3.2),

$$PAP^{-1} = \operatorname{diag}(\Lambda_0, \ldots, \Lambda_{m-1}),$$

where, for each j in m,

$$\Lambda_j = \operatorname{diag}(\Lambda_{j0}, \ldots, \Lambda_{j,n_j-1}),$$

where, for each k in n_j, $\Lambda_{j,k}$ is the $s_{jk} \times s_{jk}$ matrix

$$\begin{pmatrix} \lambda_j & 1 & 0 & \ldots & 0 \\ 0 & \lambda_j & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda_j & 1 \\ 0 & \ldots & \ldots & 0 & \lambda_j \end{pmatrix}.$$

Bibliography

[1] Kenneth Hoffman and Ray Kunze. *Linear algebra*. Second edition. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

[2] Cemal Koç. Topics in linear algebra. Printed by the Department of Mathematics, Middle East Technical University, 2010. No ISBN.

[3] Serge Lang. *Linear Algebra*. Addison-Wesley, Reading MA, second edition, 1970. World Student Series edition, second printing, 1972.

[4] Serge Lang. *Algebra*. Addison-Wesley, Reading MA, third edition, 1993. Reprinted with corrections, 1997.

[5] Steven Roman. *Advanced linear algebra*, volume 135 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2005.