Appendix A

Algebraic concepts

A.1 Number fields

A *number field* $\mathbb{F}$ contains at least two elements 0 and 1. It is an Abelian group under addition, with 0 as the unit element. The non-zero elements form an Abelian group under multiplication, with 1 as unit element. Addition and multiplication together obey the *left and right distributive laws*,

$$a(b+c) = ab + ac \;, \qquad (a+b)c = ac + bc \qquad \forall\, a, b, c \in \mathbb{F} \;. \tag{A.1}$$

In our notation, multiplication takes priority over addition, so that for example $ab+ac$ means $(ab) + (ac)$.

We write $-a$ for the additive inverse of a, and a^{-1} or $1/a$ for the multiplicative inverse. And we define $a - b = a + (-b)$. The distributive law holds also for subtraction (just write it as $a(d-c) = ad - ac$ with $d = b + c$, $b = d - c$). It follows that $0a = a0 = a(0-0) = a0 - a0 = 0$, and $a(-b) = a(0-b) = a0 - ab = -ab$. Thus, division by 0 is not allowed, the equation $0a = b$ has no solution for a unless $b = 0$, in which case every $a \in \mathbb{F}$ is a solution.

The *characteristic* of a number field is the order of the number 1 as an element of the addition group. In other words, it is the smallest positive integer n such that the sum of n times 1 is 0,

$$n \times 1 = 1 + 1 + \cdots + 1 = 0 \;. \tag{A.2}$$

If no such n exists, then $\mathbb{F}$ is said to have characteristic 0.

The field of rational numbers, $\mathbb{Q}$, consisting of the integers, $\mathbb{Z}$, and all ratios of integers, is contained in every field of characteristic 0.

Essentially the only number fields of interest for physical applications are the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$. To get $\mathbb{R}$ from $\mathbb{Q}$, join to $\mathbb{Q}$ the irrational numbers, constructed as limits of sequences of rational numbers. To get $\mathbb{C}$ from $\mathbb{R}$, join to $\mathbb{R}$ the square root of -1, that is, one particular root $x = \mathrm{i}$ of the equation $x^2 + 1 = 0$. The general complex number is $c = a + b\,\mathrm{i}$ with $a, b \in \mathbb{R}$, and its complex conjugate is $c^* = a - b\,\mathrm{i}$.

A basic property of the complex numbers is their algebraic completeness.

Theorem A.1 (The fundamental theorem of algebra) *The complex number field $\mathbb{C}$ is algebraically complete. That is, every polynomial of degree n,*

$$p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 \;, \tag{A.3}$$

with complex coefficients $c_0, c_1, \ldots, c_{n-1}$*, has exactly* n *complex roots* $r_1, r_2, \ldots, r_n$ *and may be factorized as*

$$p(x) = (x - r_1)(x - r_2)\cdots(x - r_n)\ . \tag{A.4}$$

A.2 Algebras and division algebras

Definition A.2 *An* "associative algebra"*, or simply an* "algebra"*,* $\mathcal{A}$ *over the number field* $\mathbb{F}$ *has associative operations of addition and multiplication, as well as scalar multiplication, satisfying left and right distributive laws. The special element* $0 \in \mathcal{A}$ *is the unit for addition.*

Thus, for all $A, B \in \mathcal{A}$ *and* $a, b \in \mathbb{F}$ *we have* $A + B \in \mathcal{A}$*,* $AB \in \mathcal{A}$*,* $aA = Aa \in \mathcal{A}$*,* $a(A + B) = aA + aB$*, and* $(a + b)A = aA + bA$*.*

A "division algebra" *is an algebra with multiplicative unit* I *(such that* $IA = AI = A$*) in which every non-zero element is invertible. The inverse of* A *is called* A^{-1}*, and* $A^{-1}A = AA^{-1} = I$*.*

Theorem A.3 *The only (finite dimensional) division algebra over the complex number field* $\mathbb{C}$ *is* $\mathbb{C}$ *itself.*

Proof. Let A belong to a finite dimensional complex division algebra. The finite dimensionality implies that there exists an integer $n \geq 1$ such that A^n is a linear combination of the powers A^k for $k = 0, 1, \ldots, n-1$ (we define $A^0 = I$). In other words, A is the root of a polynomial equation

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I = 0\ , \tag{A.5}$$

with complex coefficients a_i. Since the complex numbers are algebraically complete, we may factorize this polynomial into linear factors, to obtain the equation

$$(A - b_1 I)(A - b_2 I)\cdots(A - b_n I) = 0\ , \tag{A.6}$$

with complex coefficients b_i. In a division algebra there are now two possibilities. Either $A = b_1 I$, or the first factor $A - b_1 I$ is invertible and can be divided out, leaving the equation

$$(A - b_2 I)\cdots(A - b_n I) = 0\ . \tag{A.7}$$

Repeating the argument, we end up with $A = b_k I$ for some b_k.

QED

There are three division algebras over the real number field $\mathbb{R}$. They are $\mathbb{R}$ itself, the complex numbers $\mathbb{C}$, and the *quaternions* $\mathbb{H}$.

The general quaternion is $q = a + b\,\mathrm{i} + c\,\mathrm{j} + d\,\mathrm{k}$, with a, b, c, d real numbers, $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$, $\mathrm{ij} = -\mathrm{ji} = \mathrm{k}$, $\mathrm{jk} = -\mathrm{kj} = \mathrm{i}$, and $\mathrm{ki} = -\mathrm{ik} = \mathrm{j}$. The eight special quaternions $\pm 1, \pm\mathrm{i}, \pm\mathrm{j}, \pm\mathrm{k}$ form a group, called the *quaternion group*. We may define $q^* = a - b\,\mathrm{i} - c\,\mathrm{j} - d\,\mathrm{k}$ and $|q|^2 = q^* q = q q^* = a^2 + b^2 + c^2 + d^2$. The unit quaternions, with $|q| = 1$, also form a group, which we recognize as the Lie group SU(2).

A.3 Algebraic numbers and algebraic integers

Definition A.4 *An* “algebraic number” *is a root of a polynomial equation*

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \tag{A.8}$$

with rational coefficients $a_{n-1}, \ldots, a_1, a_0$.

An “algebraic integer” *is a root of a polynomial equation with integer coefficients.*

The transcendental real numbers are those (like e and π) that are not algebraic numbers. The concept of algebraic integers is useful in the representation theory of finite groups, since character values are always algebraic integers.

Of course, the linear polynomial $x + a_0$ has the integer root $x = -a_0$, thus every integer is an algebraic integer.

Theorem A.5 *If an algebraic integer is a rational number, it is an integer.*

Proof. Assume that $x = b/c$ where b and c are integers, $c > 1$, and that

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \tag{A.9}$$

with integers $a_{n-1}, \ldots, a_1, a_0$. Multiplication of the equation by c^{n-1} proves that b^n/c is an integer,

$$\frac{b^n}{c} = -a_{n-1}b^{n-1} - a_{n-2}b^{n-2}c - \cdots - a_1 bc^{n-2} - a_0 c^{n-1} \ . \tag{A.10}$$

Hence, if p is a prime factor of c, it is also a prime factor of b^n. But every prime factor of b^n is a prime factor of b, and so we have $x = b'/c'$ where $b' = b/p$ and $c' = c/p$ are integers. We repeat the same argument until we have divided out every prime factor of c.

QED

Theorem A.6 *If x and y are algebraic integers, then $x + y$ and xy are algebraic integers.*

If x is a root of a polynomial of the form

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \tag{A.11}$$

with coefficients $a_{n-1}, \ldots, a_1, a_0$ that are algebraic integers, then x is an algebraic integer.

These properties of the algebraic integers are not immediately obvious, but a couple of examples will illustrate how they can be proved. We will need some linear algebra, see Appendix B.

Suppose, for example, that x is a root of the equation

$$x^2 + 5x - 3 = 0 \ . \tag{A.12}$$

We know of course how to solve this quadratic equation, the roots are the algebraic integers

$$x = \frac{-5 \pm \sqrt{37}}{2} \ . \tag{A.13}$$

Pick any one of the roots and introduce the vector

$$\mathbf{u} = \begin{pmatrix} 1 \\ x \end{pmatrix} \in \mathbb{C}^2 \,. \tag{A.14}$$

Multiplication of this vector by the number x gives

$$x\mathbf{u} = \begin{pmatrix} x \\ x^2 \end{pmatrix} = \begin{pmatrix} x \\ -5x+3 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 3 & -5 \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \mathbf{X}\mathbf{u} \,, \tag{A.15}$$

where $\mathbf{X}$ is a 2×2 matrix with integer matrix elements. Thus, $\mathbf{u}$ is an eigenvector of the matrix $\mathbf{X}$ with eigenvalue x, and x is a root of the characteristic polynomial

$$\det(x\mathbf{I} - \mathbf{X}) = \begin{vmatrix} x & -1 \\ -3 & x+5 \end{vmatrix} = x^2 + 5x - 3 \,. \tag{A.16}$$

Suppose further that y is a root of the equation

$$y^3 - 2y^2 - 3y + 7 = 0 \,, \tag{A.17}$$

and define

$$\mathbf{v} = \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix} \in \mathbb{C}^3 \,. \tag{A.18}$$

Then

$$y\mathbf{v} = \begin{pmatrix} y \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} y \\ y^2 \\ 2y^2 + 3y - 7 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -7 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix} = \mathbf{Y}\mathbf{v} \,, \tag{A.19}$$

where $\mathbf{Y}$ is a 3×3 matrix with integer matrix elements. Now introduce the vector

$$\mathbf{w} = \mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} 1 \\ y \\ y^2 \\ x \\ xy \\ xy^2 \end{pmatrix} \in \mathbb{C}^2 \otimes \mathbb{C}^3 = \mathbb{C}^6 \,. \tag{A.20}$$

Multiplication of $\mathbf{w}$ by the number x gives

$$x\mathbf{w} = (x\mathbf{u}) \otimes \mathbf{v} = (\mathbf{X}\mathbf{u}) \otimes \mathbf{v} = (\mathbf{X} \otimes \mathbf{I}_3)\, \mathbf{w} \,, \tag{A.21}$$

where $\mathbf{X} \otimes \mathbf{I}_3$ is a 6×6 matrix with integer matrix elements,

$$\mathbf{X} \otimes \mathbf{I}_3 = \begin{pmatrix} 0 & 1 \\ 3 & -5 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 3 & 0 & 0 & -5 & 0 & 0 \\ 0 & 3 & 0 & 0 & -5 & 0 \\ 0 & 0 & 3 & 0 & 0 & -5 \end{pmatrix} . \tag{A.22}$$

Similarly, multiplication of $\mathbf{w}$ by the number y gives

$$y\mathbf{w} = \mathbf{u} \otimes (y\mathbf{v}) = \mathbf{u} \otimes (\mathbf{Y}\mathbf{v}) = (\mathbf{I}_2 \otimes \mathbf{Y})\,\mathbf{w}\;, \tag{A.23}$$

where $\mathbf{I}_2 \otimes \mathbf{Y}$ is another 6×6 matrix with integer matrix elements,

$$\mathbf{I}_2 \otimes \mathbf{Y} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -7 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -7 & 3 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -7 & 3 & 2 \end{pmatrix}. \tag{A.24}$$

Yet other 6×6 matrix with integer matrix elements are, for example, the matrix sum $\mathbf{A} = (\mathbf{X} \otimes \mathbf{I}_3) + (\mathbf{I}_2 \otimes \mathbf{Y})$ and the matrix product $\mathbf{B} = (\mathbf{X} \otimes \mathbf{I}_3)(\mathbf{I}_2 \otimes \mathbf{Y}) = \mathbf{X} \otimes \mathbf{Y}$.

We see that $\mathbf{w}$ is an eigenvector of the matrix $\mathbf{A}$ with eigenvalue $x + y$, and is also an eigenvector of the matrix $\mathbf{B}$ with eigenvalue xy. This proves that $x + y$ and xy are algebraic integers, since they are roots of the characteristic polynomials of matrices with integer matrix elements.

Suppose still further that x and y are the same algebraic integers as above, and that z is a root of the polynomial equation

$$z^2 + xz + y = 0\;. \tag{A.25}$$

Introduce the vectors

$$\mathbf{t} = \begin{pmatrix} 1 \\ z \end{pmatrix} \tag{A.26}$$

and

$$\mathbf{s} = \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{t} = \begin{pmatrix} 1 \\ x \end{pmatrix} \otimes \begin{pmatrix} 1 \\ y \\ y^2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ z \end{pmatrix} \in \mathbb{C}^2 \otimes \mathbb{C}^3 \otimes \mathbb{C}^2 = \mathbb{C}^{12}\;. \tag{A.27}$$

Since

$$z\mathbf{t} = z\begin{pmatrix} 1 \\ z \end{pmatrix} = \begin{pmatrix} z \\ z^2 \end{pmatrix} = \begin{pmatrix} z \\ -xz - y \end{pmatrix} = (\mathbf{C} - x\mathbf{D} - y\mathbf{E})\,\mathbf{t}\;, \tag{A.28}$$

with

$$\mathbf{C} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{E} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{A.29}$$

we have that

$$\begin{aligned} z\mathbf{s} &= \mathbf{u} \otimes \mathbf{v} \otimes (\mathbf{C}\mathbf{t}) - (\mathbf{X}\mathbf{u}) \otimes \mathbf{v} \otimes (\mathbf{D}\mathbf{t}) - \mathbf{u} \otimes (\mathbf{Y}\mathbf{v}) \otimes (\mathbf{E}\mathbf{t}) \\ &= (\mathbf{I}_2 \otimes \mathbf{I}_3 \otimes \mathbf{C} - \mathbf{X} \otimes \mathbf{I}_3 \otimes \mathbf{D} - \mathbf{I}_2 \otimes \mathbf{Y} \otimes \mathbf{E})\,\mathbf{s}\;. \end{aligned} \tag{A.30}$$

Thus, $\mathbf{s}$ is an eigenvector with eigenvalue z of a 12×12 matrix having integer matrix elements, which shows that z is an algebraic integer.

Appendix B

Finite dimensional linear algebra

This appendix is a summary of the theory of finite dimensional vector spaces over a number field $\mathbb{F}$. The complex numbers, $\mathbb{F} = \mathbb{C}$, will be our standard example, for two good reasons. From the mathematical point of view, working with complex numbers is convenient, for example in eigenvalue problems, for the reason that $\mathbb{C}$ is algebraically complete: every polynomial of degree n with complex coefficients has exactly n roots. From the physical point of view, the Hilbert space of state vectors in quantum mechanics is complex.

The second example of physical interest is the real numbers, $\mathbb{F} = \mathbb{R}$. For example, vectors representing points in (Euclidean) space, or electromagnetic fields, are real. Both $\mathbb{C}$ and $\mathbb{R}$ are number fields of characteristic 0 (or ∞, which is the same), that is, they contain the ring of integers, $\mathbb{Z}$.

We exclude the interesting case $\mathbb{F} = \mathbb{H}$, the quaternions, in which multiplication of numbers is non-commutative. The pleasure of working out the theory of quaternionic vector spaces is left as an exercise. Quantum mechanics based on quaternionic Hilbert spaces is a theoretical possibility which has so far found no physical application.

Many definitions and results of the finite dimensional theory are equally valid in the infinite dimensional case. The main difference is that in an infinite dimensional vector space one needs to consider infinite sums as limits of finite sums, which means that some kind of topology is required. In a Hilbert space, for example, the scalar product defines the topology. The finite dimensional case is simpler because there is usually no need to mention topology.

B.1 Vector spaces

Definition B.1 *A* "vector space" V *over the number field* $\mathbb{F}$ *is a commutative group under vector addition. The zero vector* $\mathbf{0}$ *is the unit element of the group, and* $-\mathbf{v}$ *is the additive inverse of* $\mathbf{v}$.

There is also defined an operation of multiplication of scalars with vectors, $a\mathbf{v} \in V$ *when* $a \in \mathbb{F}$ *and* $\mathbf{v} \in V$. *In particular,* $1\mathbf{v} = \mathbf{v}$.

The following associative and distributive laws for multiplication and addition hold for all vectors $\mathbf{u}, \mathbf{v} \in V$ *and all scalars* $a, b \in \mathbb{F}$,

$$(ab)\mathbf{v} = a(b\mathbf{v}) \,, \qquad (a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v} \,, \qquad a(\mathbf{u}+\mathbf{v}) = a\mathbf{u} + a\mathbf{v} \,. \tag{B.1}$$

As always, multiplications are carried out before additions, so that $a\mathbf{u} + b\mathbf{v}$ means $(a\mathbf{u}) + (b\mathbf{v})$.

We define $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$. The distributive laws hold also if "+" is replaced by "−": simply write them as

$$(c - b)\mathbf{v} = c\mathbf{v} - b\mathbf{v} \;, \qquad a(\mathbf{w} - \mathbf{v}) = a\mathbf{w} - a\mathbf{v} \;, \tag{B.2}$$

with $c = a + b$, $\mathbf{w} = \mathbf{u} + \mathbf{v}$. It follows immediately that $a\mathbf{v} = \mathbf{0}$ if $a = 0$ or $\mathbf{v} = \mathbf{0}$,

$$0\mathbf{v} = (0 - 0)\mathbf{v} = 0\mathbf{v} - 0\mathbf{v} = \mathbf{0} \;, \qquad a\mathbf{0} = a(\mathbf{0} - \mathbf{0}) = a\mathbf{0} - a\mathbf{0} = \mathbf{0} \;. \tag{B.3}$$

The other way around, if $a\mathbf{v} = \mathbf{0}$, then either $a = 0$, or else $\mathbf{v} = a^{-1}a\mathbf{v} = \mathbf{0}$.

It also follows that $(-1)\mathbf{v} = (0 - 1)\mathbf{v} = 0\mathbf{v} - 1\mathbf{v} = -\mathbf{v}$.

From the group theoretical point of view, the distributive law $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$ means that every scalar a acts as a group homomorphism from V into itself. The associative law $a(b\mathbf{v}) = (ab)\mathbf{v}$ means that the composition of such homomorphisms is the same as multiplication in the number field $\mathbb{F}$. Thus, the multiplicative group of the non-zero elements of $\mathbb{F}$ acts as a group of automorphisms over the addition group V.

Basis vectors

Definition B.2 *A* "linear combination" *of the m vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in V$ is a sum*

$$\mathbf{v} = \sum_{i=1}^{m} a_i \mathbf{u}_i = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \ldots + a_m\mathbf{u}_m \;, \tag{B.4}$$

with coefficients $a_i \in \mathbb{F}$.

The vectors $\mathbf{u}_i$ are "linearly independent" *if the equation*

$$\sum_{i=1}^{m} a_i \mathbf{u}_i = \mathbf{0} \;, \tag{B.5}$$

considered as an equation for the coefficients $a_i \in \mathbb{F}$, has only the trivial solution

$$a_1 = a_2 = \ldots = a_m = 0 \;. \tag{B.6}$$

They form a (finite) "basis" *for V if, for every vector $\mathbf{v} \in V$, eq. (B.4) has one and only one solution for the coefficients $a_i \in \mathbb{F}$.*

Obviously, linearly independent vectors must all be non-zero. Taking $\mathbf{v} = \mathbf{0}$ in eq. (B.4), we see that basis vectors must be linearly independent. On the other hand, linear independence of the vectors $\mathbf{u}_i$ is enough to guarantee the uniqueness of the solution of eq. (B.4), for

$$\mathbf{v} = \sum_{i=1}^{m} a_i \mathbf{u}_i = \sum_{i=1}^{m} b_i \mathbf{u}_i \tag{B.7}$$

implies that

$$\sum_{i=1}^{m} (a_i - b_i)\mathbf{u}_i = \mathbf{0} \;, \tag{B.8}$$

with the unique solution $a_i - b_i = 0$ for $i = 1, 2, \ldots, m$.

Theorem B.3 *k vectors $\mathbf{v}_j$ that are all linear combinations of m vectors $\mathbf{u}_i$,*

$$\mathbf{v}_j = \sum_{i=1}^{m} a_{ij}\mathbf{u}_i \ , \tag{B.9}$$

can not be linearly independent if $k > m$.

Proof. We have to prove that the equation

$$\mathbf{0} = \sum_{j=1}^{k} b_j\mathbf{v}_j = \sum_{i=1}^{m}\sum_{j=1}^{k} a_{ij}b_j\mathbf{u}_i \tag{B.10}$$

has non-trivial solutions for the k coefficients b_j. It is in fact solved by any solution of the set of m equations

$$\sum_{j=1}^{k} a_{ij}b_j = 0 \ . \tag{B.11}$$

Each of these equations can be used to eliminate at most one unknown. When $k > m$, there are more unknowns than equations, hence there exists a solution where at least $k - m$ of the coefficients b_j are undetermined and may have arbitrary values.

QED

Theorem B.4 *Different finite bases for the vector space V must have the same number of basis vectors.*

In fact, no basis can have more basis vectors than another, because the vectors of the first basis are linearly independent, and they are linear combinations of the vectors of the second basis.

Definition B.5 *The* "dimension" *of a vector space V over $\mathbb{F}$, $\dim V$, is either infinite or the number of vectors in a finite basis.*

If the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are not linearly independent, it is because either $\mathbf{u}_1 = \mathbf{0}$, or at least one vector $\mathbf{u}_k$, where $k \geq 2$, is a linear combination of the $k-1$ previous vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}$. In fact, if eq. (B.5) has a non-trivial solution, take k to be the largest index i such that $a_i \neq 0$. If $k = 1$, then $\mathbf{u}_1 = \mathbf{0}$, otherwise $\mathbf{u}_k$ is a linear combination

$$\mathbf{u}_k = -a_k^{-1}\sum_{i=1}^{k-1} a_i\mathbf{u}_i \ . \tag{B.12}$$

Theorem B.6 *The m non-zero vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are linearly independent if and only if no vector $\mathbf{u}_k$ with $k = 2, \ldots, m$ is a linear combination of the $k-1$ vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}$.*

Therefore, given any set of linearly independent vectors in V which is not a basis for V, we may always join to it another linearly independent vector, so as to get a larger set of linearly independent vectors. The process of extending a set of linearly independent vectors will stop when we get a basis for V, or it will continue indefinitely. This construction shows that every vector space has a finite basis, unless it is large enough to contain an infinite sequence of vectors where none of the vectors is a linear combination of the preceding vectors.

Theorem B.7 *If the vector space V has finite dimension n, then any set of n linearly independent vectors in V is a basis for V. An arbitrary set of fewer than n linearly independent vectors in V can always be extended to a basis.*

The case of infinite dimensional vector spaces needs special treatment. There it is natural to introduce infinite sums of vectors, defined as limits of finite sums with respect to some topology. In this appendix we discuss only finite dimensional vector spaces, where there is usually no need for infinite sums, and no need to mention topology.

The vector space $\mathbb{F}^n$

The most trivial example of a vector space is the zero dimensional space $\{\mathbf{0}\}$. Next on the list is the one dimensional vector space $\mathbb{F}$, in which any non-zero number is a basis. The obvious generalization is $\mathbb{F}^n$, this is the standard example of an n dimensional vector space over $\mathbb{F}$. A vector $\mathbf{v} \in \mathbb{F}^n$ is an $n \times 1$ matrix (a column matrix) with components $v_i \in \mathbb{F}$,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} . \tag{B.13}$$

Vector addition, multiplication by scalars, and the zero vector in $\mathbb{F}^n$ are defined as follows,

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix} , \qquad a\mathbf{v} = \begin{pmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{pmatrix} , \qquad \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} . \tag{B.14}$$

The general vector $\mathbf{v} \in \mathbb{F}^n$ can be written in a unique way as a linear combination

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{d}_i = v_1 \mathbf{d}_1 + v_2 \mathbf{d}_2 + \ldots + v_n \mathbf{d}_n \tag{B.15}$$

of the n natural basis vectors

$$\mathbf{d}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} , \quad \mathbf{d}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} , \quad \ldots , \quad \mathbf{d}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} . \tag{B.16}$$

The components of these basis vectors are the Kronecker delta symbols,

$$(\mathbf{d}_i)_j = \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \, , \\ 0 & \text{if} \quad i \neq j \, . \end{cases} \tag{B.17}$$

By eq. (B.4), a basis for an n dimensional vector space V identifies the vector $\mathbf{v} \in V$ with a matrix of coordinates,

$$\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{u}_i \in V \quad \leftrightarrow \quad \mathbf{a} = \sum_{i=1}^{n} a_i \mathbf{d}_i = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{F}^n . \tag{B.18}$$

B.2 Linear transformations

Definition B.8 *Given two vector spaces U and V over $\mathbb{F}$. A transformation $A : U \to V$ is* "linear" *if it preserves the vector space structure, that is, if*

$$A(a\mathbf{u} + b\mathbf{v}) = a(A\mathbf{u}) + b(A\mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in U, \quad a, b \in \mathbb{F} \; . \tag{B.19}$$

A linear transformation is also called a "linear operator"*, or simply an* "operator"*.*

When A is linear, its function value at $\mathbf{u}$ *is written as* $A\mathbf{u}$ *instead of* $A(\mathbf{u})$.

A linear transformation $A : U \to V$ is a "homomorphism" *from the vector space U into the vector space V. The set of all such homomorphisms is called* $\mathrm{Hom}(U, V)$.

Two vector spaces U and V are "isomorphic" *if there exists an invertible linear transformation (an* "isomorphism"*) from U to V.*

Recall that a group homomorphism transforms the unit element of one group into the unit element of the other group. Since a linear transformation $A : U \to V$ is a group homomorphism in the sense that it preserves vector addition, it must transform the zero vector $\mathbf{0} = \mathbf{0}_U \in U$ into the zero vector $\mathbf{0} = \mathbf{0}_V \in V$. That is, we must have $A\mathbf{0} = \mathbf{0}$.

If $A : U \to V$ is linear and invertible, its inverse $A^{-1} : V \to U$ is also linear and invertible. In fact, $(A^{-1})^{-1} = A$. Thus, two isomorphic vector spaces are identical from an abstract point of view. In particular, an isomorphism between U and V transforms a basis of U into a basis of V, and vice versa.

For example, the mapping defined by eq. (B.18) is an isomorphism between V and $\mathbb{F}^n$, by which the basis vectors of V correspond to the natural basis vectors of $\mathbb{F}^n$.

Theorem B.9 *Two isomorphic vector spaces must have the same dimension.*

Every n dimensional vector space over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.

The simplest example of a linear transformation from V to itself is the identity function $I = I_V : \mathbf{v} \mapsto I\mathbf{v} = \mathbf{v}$. The next simplest example is multiplication by a scalar $a \in \mathbb{F}$, $\mathbf{v} \mapsto A\mathbf{v} = a\mathbf{v}$, this is linear as long as multiplication in the number field $\mathbb{F}$ is commutative. In fact,

$$a(b\mathbf{v}) = (ab)\mathbf{v} = (ba)\mathbf{v} = b(a\mathbf{v}) \; . \tag{B.20}$$

$\mathrm{Hom}(U, V)$ is itself a vector space, where the vector sum $A + B$ and scalar multiple aA is defined in the natural way,

$$(A + B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}, \quad (aA)\mathbf{u} = a(A\mathbf{u}) \quad \forall \mathbf{u} \in U, \quad a \in \mathbb{F} \; . \tag{B.21}$$

The definition of aA simply means that a product such as $aA\mathbf{u}$ is associative and can be written without parentheses. The zero vector in this space is the transformation Z such that $Z\mathbf{u} = \mathbf{0}\ \forall \mathbf{u} \in U$.

Definition B.10 *Given a basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m \in U$ and a basis $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n \in V$. The* "matrix elements" *of the linear transformation $A : U \to V$, relative to these two bases, are the numbers A_{ji}, with $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, defined by the expansion*

$$A\mathbf{e}_i = \sum_{j=1}^{n} A_{ji}\mathbf{f}_j \; . \tag{B.22}$$

By definition, the first index j of the matrix element A_{ji} is associated with V and the second index i with U. Given an arbitrary vector

$$\mathbf{u} = \sum_{i=1}^{m} u_i \mathbf{e}_i \in U \ . \tag{B.23}$$

we have that

$$A\mathbf{u} = \sum_{i=1}^{m} u_i A\mathbf{e}_i = \sum_{i=1}^{m} u_i \sum_{j=1}^{n} A_{ji}\mathbf{f}_j = \sum_{j=1}^{n} \sum_{i=1}^{m} A_{ji} u_i \mathbf{f}_j \ . \tag{B.24}$$

Thus, the nm matrix elements A_{ji} define uniquely the linear transformation A.

Given also a basis $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p \in W$, and a linear transformation $B : V \to W$ with matrix elements B_{kj},

$$B\mathbf{f}_i = \sum_{k=1}^{p} B_{kj}\mathbf{g}_k \ . \tag{B.25}$$

The composition $BA : U \to W$ of the two linear transformations $A : U \to V$ and $B : V \to W$ is again a linear transformation, and

$$BA\mathbf{e}_i = B\sum_{j=1}^{n} A_{ji}\mathbf{f}_j = \sum_{j=1}^{n} A_{ji} \sum_{k=1}^{p} B_{kj}\mathbf{g}_k = \sum_{k=1}^{p} \sum_{j=1}^{n} B_{kj} A_{ji} \mathbf{g}_k \ . \tag{B.26}$$

We see that the matrix elements of the composite linear transformation $D = BA$ are

$$D_{ki} = \sum_{j=1}^{n} B_{kj} A_{ji} \ . \tag{B.27}$$

B.3 Matrices

A general $n \times m$ matrix $\mathbf{A}$ has nm matrix elements $A_{ij} \in \mathbb{F}$ arranged in a rectangle of n rows and m columns,

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m} \\ A_{21} & A_{22} & \ldots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nm} \end{pmatrix} . \tag{B.28}$$

Interchanging rows and columns, we get the *transposed* matrix

$$\mathbf{A}^\top = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1m} & A_{2m} & \ldots & A_{nm} \end{pmatrix} . \tag{B.29}$$

When the number field is the complex numbers, $\mathbb{F} = \mathbb{C}$, we often combine transposition with complex conjugation, which we denote by *. This gives the *Hermitean conjugate* matrix $\mathbf{A}^\dagger = \mathbf{A}^{*\top} = \mathbf{A}^{\top *}$.

We may think of the $n \times m$ matrix $\mathbf{A}$ as consisting of m column vectors,

$$\mathbf{A} = (\, \mathbf{A}_1 \quad \mathbf{A}_2 \quad \dots \quad \mathbf{A}_m \,) , \qquad \mathbf{A}_i = \begin{pmatrix} A_{1i} \\ A_{2i} \\ \vdots \\ A_{ni} \end{pmatrix} \in \mathbb{F}^n . \tag{B.30}$$

We may also think of it as a vector in $\mathbb{F}^{nm}$, obtained by stacking the column vectors on top of each other,

$$\mathbf{A} \leftrightarrow \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_m \end{pmatrix} , \qquad \mathbf{A}_i \in \mathbb{F}^n . \tag{B.31}$$

The matrix product of a $p \times n$ matrix $\mathbf{B}$ and an $n \times m$ matrix $\mathbf{A}$ is the $p \times m$ matrix $\mathbf{D} = \mathbf{BA}$ with matrix elements

$$D_{ik} = \sum_{j=1}^{n} B_{ij} A_{jk} . \tag{B.32}$$

Matrix multiplication is associative, that is, $\mathbf{C}(\mathbf{BA}) = (\mathbf{CB})\mathbf{A}$ when $\mathbf{C}$ is a $q \times p$ matrix. The proof consists in just writing out the definitions and using the distributive and associative laws for the addition and multiplication of numbers,

$$[\mathbf{C}(\mathbf{BA})]_{hk} = \sum_{i=1}^{p} \sum_{j=1}^{n} C_{hi} B_{ij} A_{jk} = [(\mathbf{CB})\mathbf{A}]_{hk} . \tag{B.33}$$

Matrix multiplication is also *bilinear*, that is, linear in each of the two factors,

$$\mathbf{A}(a\mathbf{C} + b\mathbf{D}) = a\mathbf{AC} + b\mathbf{AD} , \qquad (a\mathbf{A} + b\mathbf{B})\mathbf{C} = a\mathbf{AC} + b\mathbf{BC} . \tag{B.34}$$

Here $a, b \in \mathbb{F}$, while $\mathbf{A}$ and $\mathbf{B}$ are $p \times n$ matrices, $\mathbf{C}$ and $\mathbf{D}$ are $n \times m$ matrices.

An $n \times m$ matrix $\mathbf{A}$ defines a linear transformation $A : \mathbb{F}^m \to \mathbb{F}^n$, transforming $\mathbf{u} \in \mathbb{F}^m$ into $A\mathbf{u} = \mathbf{Au} \in \mathbb{F}^n$, where $\mathbf{Au}$ is the matrix product. In the same way, a $p \times n$ matrix $\mathbf{B}$ defines a linear transformation $B : \mathbb{F}^n \to \mathbb{F}^p$, and the composite linear transformation $D = BA : \mathbb{F}^m \to \mathbb{F}^p$ corresponds to the matrix product $\mathbf{D} = \mathbf{BA}$,

$$D\mathbf{u} = (BA)\mathbf{u} = B(A\mathbf{u}) = \mathbf{B}(\mathbf{Au}) = (\mathbf{BA})\mathbf{u} . \tag{B.35}$$

The identity function $I = I_m$ on $\mathbb{F}^m$ is linear and is given by the identity $m \times m$ square matrix

$$\mathbf{I} = \mathbf{I}_m = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} . \tag{B.36}$$

$\mathbf{I}_m$ is a right unit and $\mathbf{I}_n$ is a left unit for the $n \times m$ matrix $\mathbf{A}$, that is, $\mathbf{AI}_m = \mathbf{I}_n\mathbf{A} = \mathbf{A}$. By definition, an $m \times n$ matrix $\mathbf{B}$ is a left inverse of an $n \times m$ matrix $\mathbf{A}$ if $\mathbf{BA} = \mathbf{I}_m$, and it is a right inverse if $\mathbf{AB} = \mathbf{I}_n$.

What happens if $\mathbf{A}$ is an $n \times m$ matrix having both a left inverse $\mathbf{B}$, with $\mathbf{BA} = \mathbf{I}_m$, and a right inverse $\mathbf{C}$, with $\mathbf{AC} = \mathbf{I}_n$? In that case we have one unique matrix which is simultaneously a left and right inverse of $\mathbf{A}$, since

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_m\mathbf{C} = \mathbf{C}\ . \tag{B.37}$$

When this unique inverse of $\mathbf{A}$ exists, we call it $\mathbf{A}^{-1}$, and we say that $\mathbf{A}$ is invertible.

An invertible $n \times m$ matrix $\mathbf{A}$ defines an isomorphism between $\mathbb{F}^n$ and $\mathbb{F}^m$. But this is possible only if $n = m$, since isomorhic vector spaces must have the same dimension. This proves that only square matrices can be invertible.

Matrix transposition and matrix inversion both invert the order of matrix products,

$$(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top, \qquad (\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}\ . \tag{B.38}$$

Matrix representation and change of basis

A basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m \in U$ defines a one to one correspondence between a vector $\mathbf{u} \in U$ and the $m \times 1$ matrix of its coordinates relative to that basis,

$$\mathbf{u} = \sum_{i=1}^{m} a_i \mathbf{e}_i \quad \leftrightarrow \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}\ , \tag{B.39}$$

which is an isomorphism between U and $\mathbb{F}^m$. A formal, but rather natural and useful way to read this relation is as a matrix product $\mathbf{u} = [\mathbf{e}]\,\mathbf{a}$, where

$$[\mathbf{e}] = (\,\mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \quad \mathbf{e}_m\,) \tag{B.40}$$

is the $1 \times m$ matrix of basis vectors. In the special case $U = \mathbb{F}^m$, each basis vector $\mathbf{e}_i$ is an $m \times 1$ matrix, and the matrix $[\mathbf{e}]$ of basis vectors is an $m \times m$ square matrix. With the natural basis vectors of $\mathbb{F}^m$ we get simply the identity matrix,

$$[\mathbf{d}] = (\,\mathbf{d}_1 \quad \mathbf{d}_2 \quad \ldots \quad \mathbf{d}_m\,) = \mathbf{I}\ . \tag{B.41}$$

With these definitions, $[\mathbf{e}]$ is a linear transformation $\mathbb{F}^m \to U$, and the fact that it is invertible is precisely the definition of a basis. Hence the coordinate matrix is

$$\mathbf{a} = [\mathbf{e}]^{-1}\mathbf{u}\ . \tag{B.42}$$

An equivalent definition of the linear transformation $[\mathbf{e}] : \mathbb{F}^m \to U$ is that it defines a correspondence between the natural basis vector $\mathbf{d}_i \in \mathbb{F}^m$ and the basis vector $\mathbf{e}_i \in U$,

$$\mathbf{e}_i = [\mathbf{e}]\,\mathbf{d}_i, \qquad \mathbf{d}_i = [\mathbf{e}]^{-1}\mathbf{e}_i\ . \tag{B.43}$$

Let $[\mathbf{e}]$ be a basis for U, as before, let $A : U \to V$ be a linear transformation, and let

$$[\,\mathbf{f}\,] = (\,\mathbf{f}_1 \quad \mathbf{f}_2 \quad \ldots \quad \mathbf{f}_n\,) \tag{B.44}$$

be a basis for V. By Definition B.10, the linear transformation A is repesented relative to these two bases by an $n \times m$ matrix $\mathbf{A}$, and we may write the definition in eq. (B.22) as

$$A[\mathbf{e}] = (\,A\mathbf{e}_1 \quad A\mathbf{e}_2 \quad \ldots \quad A\mathbf{e}_m\,) = [\,\mathbf{f}\,]\,\mathbf{A}\ . \tag{B.45}$$

Or,

$$\mathbf{A} = [\,\mathbf{f}\,]^{-1}A[\mathbf{e}]\ . \tag{B.46}$$

We will see now what happens when we change the bases in U and in V. The new basis vectors $\widetilde{\mathbf{e}}_i$ in U must be linear combinations of the old basis vectors $\mathbf{e}_i$,

$$\widetilde{\mathbf{e}}_i = \sum_{j=1}^{m} S_{ji}\mathbf{e}_j\ . \tag{B.47}$$

The coefficients S_{ji} are the matrix elements with respect to the old basis of a linear transformation S which transforms the old basis vectors into the new ones,

$$\widetilde{\mathbf{e}}_i = S\mathbf{e}_i\ . \tag{B.48}$$

S must be an isomorphism of U with itself, indeed it must be invertible since the old basis vectors must also be expressible as linear combinations of the new ones.

The matrix elements S_{ji} define an $m \times m$ matrix $\mathbf{S}$, and the above equations may be written as

$$[\widetilde{\mathbf{e}}] = S[\mathbf{e}] = (\,S\mathbf{e}_1 \quad S\mathbf{e}_2 \quad \ldots \quad S\mathbf{e}_m\,) = [\mathbf{e}]\,\mathbf{S}\ . \tag{B.49}$$

Or equivalently,

$$[\mathbf{e}] = S^{-1}[\widetilde{\mathbf{e}}] = [\widetilde{\mathbf{e}}]\,\mathbf{S}^{-1}\ . \tag{B.50}$$

With respect to the new basis in U, a vector $\mathbf{u} \in U$ has a new coordinate matrix

$$\widetilde{\mathbf{a}} = [\widetilde{\mathbf{e}}]^{-1}\mathbf{u} = \mathbf{S}^{-1}[\mathbf{e}]^{-1}\mathbf{u} = \mathbf{S}^{-1}\mathbf{a}\ . \tag{B.51}$$

In a similar way, a basis change in V is given by a linear transformation T, or an $n \times n$ matrix $\mathbf{T}$, such that

$$[\,\widetilde{\mathbf{f}}\,] = T[\,\mathbf{f}\,] = [\,\mathbf{f}\,]\,\mathbf{T}, \qquad [\,\mathbf{f}\,] = T^{-1}[\,\widetilde{\mathbf{f}}\,] = [\,\widetilde{\mathbf{f}}\,]\,\mathbf{T}^{-1}\ . \tag{B.52}$$

With respect to the new bases both in U and in V, a linear transformation $A : U \to V$ is represented by a new matrix

$$\widetilde{\mathbf{A}} = [\,\widetilde{\mathbf{f}}\,]^{-1}A[\widetilde{\mathbf{e}}] = \mathbf{T}^{-1}[\,\mathbf{f}\,]^{-1}A[\mathbf{e}]\,\mathbf{S} = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}\ . \tag{B.53}$$

As a special case, take $V = U$, $[\,\mathbf{f}\,] = [\mathbf{e}]$ and $\mathbf{T} = \mathbf{S}$. Then

$$\widetilde{\mathbf{A}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\ . \tag{B.54}$$

Definition B.11 *A transformation of the type* $\mathbf{A} \mapsto \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$, *corresponding to a change of basis, is called a* "similarity transformation" *of the square matrix* $\mathbf{A}$.

Active and passive transformations

We may think of a similarity transformation as an example of a *passive transformation.* If we think of the vector $\mathbf{u} \in V$ as representing the state of a physical system, then the basis $[\mathbf{e}]$ corresponds to a reference frame, or measurement apparatus, and the coordinate vector $\mathbf{a} = [\mathbf{e}]^{-1}\mathbf{u} \in \mathbb{F}^n$ describes the physical state $\mathbf{u}$ mathematically by a list of n numerical coordinates.

In this terminology, an *active transformation* is a linear transformation S which transforms the vector $\mathbf{u}$ (the state of the system) into $\widetilde{\mathbf{u}} = S\mathbf{u}$, without changing the basis $[\mathbf{e}]$ (the reference frame). Thus, it transforms the coordinate vector $\mathbf{a} = [\mathbf{e}]^{-1}\mathbf{u}$ into

$$\widetilde{\mathbf{a}} = [\mathbf{e}]^{-1}\widetilde{\mathbf{u}} = [\mathbf{e}]^{-1}S\mathbf{u} = [\mathbf{e}]^{-1}S[\mathbf{e}]\,\mathbf{a} = \mathbf{S}\mathbf{a}\ , \tag{B.55}$$

where the matrix $\mathbf{S} = [\mathbf{e}]^{-1}S[\mathbf{e}]$ represents the active transformation S.

The same mathematical transformation of coordinates results from a *passive transformation*, in which we do not transform the physical state $\mathbf{u}$, but transform instead the basis $[\mathbf{e}]$ into $[\widetilde{\mathbf{e}}]$. We see that we get $\widetilde{\mathbf{a}} = [\widetilde{\mathbf{e}}]^{-1}\mathbf{u}$ if we define

$$[\widetilde{\mathbf{e}}]^{-1} = [\mathbf{e}]^{-1}S = \mathbf{S}\,[\mathbf{e}]^{-1}\ . \tag{B.56}$$

Or equivalently,

$$[\widetilde{\mathbf{e}}] = S^{-1}[\mathbf{e}] = [\mathbf{e}]\,\mathbf{S}^{-1}\ . \tag{B.57}$$

Two active transformations in succession, first S_1 and then S_2, is the same as one single transformation $S = S_2S_1$, corresponding to the matrix product $\mathbf{S} = \mathbf{S}_2\mathbf{S}_1$.

What happens when we make two passive transformations in succession? We start with the basis $[\mathbf{e}]$, transforming first to a second basis $[\,\mathbf{f}\,] = [\mathbf{e}]\,\mathbf{S}_1^{-1}$, and afterwards to a third basis

$$[\mathbf{g}] = [\,\mathbf{f}\,]\,\mathbf{S}_2^{-1} = [\,\mathbf{e}\,]\,\mathbf{S}_1^{-1}\mathbf{S}_2^{-1} = [\,\mathbf{e}\,]\,(\mathbf{S}_2\mathbf{S}_1)^{-1}\ . \tag{B.58}$$

This is the same result as is produced by one single passive transformation, again given by the matrix product $\mathbf{S} = \mathbf{S}_2\mathbf{S}_1$.

B.4 Subspaces

Definition B.12 *A* “subspace” *of the vector space V is a subset U of V which is again a vector space under the same operations of vector addition and multiplication by scalars.*

Since a vector space is a group under vector addition, and a subspace is a subgroup, we know from the general group theory that the zero vector of U must be identical to the zero vector of V,

There are two trivial examples of subspaces. By definition, every vector space is a subspace of itself. And the zero dimensional vector space $\{\mathbf{0}\}$ is a subspace of every vector space.

Exercise B.13 *Let U be a subspace of a finite dimensional vector space V. Show that if $\dim U = \dim V$, then $U = V$.*

Theorem B.14 *A subset $U \subset V$ is a subspace of the vector space V if and only if*

$$a\mathbf{u} + b\mathbf{v} \in U \qquad \forall \mathbf{u}, \mathbf{v} \in U, \quad a, b \in \mathbb{F} . \tag{B.59}$$

The intersection $\bigcap U_\alpha$ of any collection of subspaces U_α is again a subspace. Therefore, given any subset $X \subset V$, the intersection of all subspaces containing X is a subspace. It is the smallest subspace containing X, it consists of all possible linear combinations of vectors in X, and is called the *linear span* of X.

In particular, if X consists of m linearly independent vectors, for example m out of n basis vectors, $m \leq n$, then the linear span of X is an m dimensional subspace.

The union $\bigcup U_\alpha$ of a collection of subspaces U_α is not a subspace, unless we define it in a suitable way. Let us define $\bigcup U_\alpha$ as the smallest subspace containing all the subspaces U_α.

Theorem B.15 *Given two finite dimensional subspaces $U_1, U_2 \subset V$. Then*

$$\dim(U_1 \cup U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 . \tag{B.60}$$

Proof. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be a basis for $U_1 \cap U_2$. By Theorem B.7 it may be extended to a basis $\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{f}_{m+1}, \ldots, \mathbf{f}_j$ for U_1, and to a basis $\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{g}_{m+1}, \ldots, \mathbf{g}_k$ for U_2. We may verify that $\mathbf{e}_1, \ldots, \mathbf{e}_m, \mathbf{f}_{m+1}, \ldots, \mathbf{f}_j, \mathbf{g}_{m+1}, \ldots, \mathbf{g}_k$ is a basis for $U_1 \cup U_2$. Hence $\dim U_1 = j$, $\dim U_2 = k$, $\dim(U_1 \cap U_2) = m$, and $\dim(U_1 \cup U_2) = j + k - m$.

QED

Definition B.16 *Given two vector spaces U and V, and a linear transformation $A : U \to V$. The* "kernel" *of A is the subset*

$$\operatorname{Ker} A = \{\, \mathbf{u} \in U \mid A\mathbf{u} = \mathbf{0} \,\} \subset U . \tag{B.61}$$

The "image" *of A is the subset*

$$\operatorname{Img} A = \{\, A\mathbf{u} \mid \mathbf{u} \in U \,\} \subset V . \tag{B.62}$$

Theorem B.17 $\operatorname{Ker} A$ *is a subspace of U,* $\operatorname{Img} A$ *is a subspace of V, and if U is finite dimensional, then*

$$\dim(\operatorname{Ker} A) + \dim(\operatorname{Img} A) = \dim U . \tag{B.63}$$

Proof. That $\operatorname{Ker} A$ and $\operatorname{Img} A$ are subspaces follows from Theorem B.14, by the linearity of A. To prove the last part of the theorem we take any basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ for $\operatorname{Ker} A$ and extend it to a basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ for U. This is possible by Theorem B.7. It is then easy to show that the $n - k$ vectors $A\mathbf{e}_{k+1}, A\mathbf{e}_{k+2}, \ldots, A\mathbf{e}_n$ form a basis for $\operatorname{Img} A$. Thus, $\dim(\operatorname{Ker} A) = k$, $\dim(\operatorname{Img} A) = n - k$ and $\dim U = n$.

QED

Theorem B.18 *Given a linear transformation $A : U \to V$. It is invertible if and only if* $\operatorname{Ker} A = \{\mathbf{0}\}$ *and* $\operatorname{Img} A = V$.

If U and V are finite dimensional and $\dim U = \dim V$*, then these three conditions are equivalent:* (i) *A is invertible;* (ii) $\operatorname{Ker} A = \{\mathbf{0}\}$*;* (iii) $\operatorname{Img} A = V$.

Quotient spaces

The concept of quotient groups has its counterpart in the theory of vector spaces. Since the addition group of vectors is Abelian, every subspace is a normal subgroup and produces a quotient vector space.

Theorem B.19 *Let U be a subspace of the vector space V. The quotient space*

$$V/U = \{\, \mathbf{v} + U \mid \mathbf{v} \in V \,\} \,, \tag{B.64}$$

where $\mathbf{v} + U = \{\, \mathbf{v} + \mathbf{u} \mid \mathbf{u} \in U \,\}$, is a vector space under the natural operations of vector addition and scalar multiplication.

If V is finite dimensional, then $\dim(V/U) = \dim V - \dim U$.

The dimension of the quotient space V/U is called the *codimension* of U in V.

Projections

Definition B.20 *The subspaces $V_1, V_2, \ldots, V_m \subset V$ are* "complementary" *if every vector $\mathbf{v} \in V$ has a unique decomposition as a sum of vectors $\mathbf{v}_i \in V_i$,*

$$\mathbf{v} = \sum_{i=1}^{m} \mathbf{v}_i \,. \tag{B.65}$$

We say then that V is the "direct sum" *of these subspaces,*

$$V = V_1 \oplus V_2 \oplus \ldots \oplus V_m \,. \tag{B.66}$$

The unique decomposition in eq. (B.65) of an arbitrary vector $\mathbf{v} \in V$ makes it possible to define linear transformations $P_1, P_2, \ldots, P_m$ on V such that

$$P_j \mathbf{v} = \mathbf{v}_j \,. \tag{B.67}$$

To prove the linearity of P_j, take a linear combination $\mathbf{w} = a\mathbf{u} + b\mathbf{v}$ with $a, b \in \mathbb{F}$ of two vectors $\mathbf{u}, \mathbf{v} \in V$ with decompositions

$$\mathbf{u} = \sum_{i=1}^{m} \mathbf{u}_i, \qquad \mathbf{v} = \sum_{i=1}^{m} \mathbf{v}_i, \qquad \mathbf{u}_i, \mathbf{v}_i \in V_i \,. \tag{B.68}$$

The decomposition of $\mathbf{w}$ is

$$\mathbf{w} = \sum_{i=1}^{m} \mathbf{w}_i, \qquad \mathbf{w}_i = a\mathbf{u}_i + b\mathbf{v}_i \,, \tag{B.69}$$

implying that

$$P_j \mathbf{w} = \mathbf{w}_j = a\mathbf{u}_j + b\mathbf{v}_j = aP_j\mathbf{u} + bP_j\mathbf{v} \,. \tag{B.70}$$

The decomposition $\mathbf{v} = \sum_i P_i \mathbf{v}\ \forall \mathbf{v} \in V$ may be written as

$$I = \sum_{i=1}^{m} P_i \,. \tag{B.71}$$

By definition, $P_i\mathbf{v}_i = \mathbf{v}_i$, while $P_i\mathbf{v}_j = \mathbf{0}$ if $i \neq j$. This implies the relations

$$P_i{}^2 = P_i \quad \forall i = 1, 2, \ldots, m; \qquad P_iP_j = 0 \quad \forall i \neq j \ . \tag{B.72}$$

Conversely, assume that $P_1, P_2, \ldots, P_m$ are linear operators on V satisfying eq. (B.71) and eq. (B.72). Then $V_i = P_iV = \operatorname{Img} P_i$ with $i = 1, 2, \ldots, m$ are complementary subspaces of V. In fact, they are subspaces because each P_i is linear, and they are complementary because eq. (B.65) for the components $\mathbf{v}_i \in V_i$ of the arbitrary vector $\mathbf{v} \in V$ has the unique solution

$$\mathbf{v}_i = P_i\mathbf{v} \ . \tag{B.73}$$

To prove that eq. (B.73) is a solution of eq. (B.65), we use eq. (B.71),

$$\mathbf{v} = I\mathbf{v} = \sum_{i=1}^{m} P_i\mathbf{v} \ . \tag{B.74}$$

And to prove that the solution is unique, we note that, by definition, every vector $\mathbf{v}_j \in V_j$ is the image under P_j of some vector $\mathbf{w}_j \in V$, hence

$$P_i\mathbf{v}_j = P_iP_j\mathbf{w}_j = \delta_{ij}P_j\mathbf{w}_j = \delta_{ij}\mathbf{v}_j \ . \tag{B.75}$$

Therefore we may operate on eq. (B.65) with P_i, to get that

$$P_i\mathbf{v} = \sum_{j=1}^{m} P_i\mathbf{v}_j = \mathbf{v}_i \ . \tag{B.76}$$

Definition B.21 *A linear transformation $P : V \to V$ is a* "projection operator"*, or simply a* "projection"*, if it is idempotent, that is, if $P^2 = P$.*

The projections $P_i : V \to V$, $i = 1, 2, \ldots, m$, are "disjoint" *if they satisfy eq. (B.72).*

A set of disjoint projections $P_1, P_2, \ldots, P_m$ is "complete" *if eq. (B.71) is satisfied. A complete set of disjoint projections is said to be a* "decomposition of the identity".

We have proved above the following theorem.

Theorem B.22 *If $V_1, V_2, \ldots, V_m$ are complementary subspaces of V, then they define uniquely a decomposition of the identity.*

Conversely, a decomposition of the identity on V, as in eq. (B.71), splits V into m complementary subspaces $V_i = P_iV$.

$P = I$ and $P = 0$ are the two trivial examples of projections. To every projection P there corresponds a complementary projection $Q = I - P$, with the properties that

$$Q^2 = I - 2P + P^2 = I - P = Q, \qquad PQ = QP = P - P^2 = 0 \ . \tag{B.77}$$

A sum of disjoint projections,

$$P = \sum_{i=1}^{k} P_i \ , \tag{B.78}$$

is always a projection,

$$P^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} P_i P_j = \sum_{i=1}^{k} P_i = P \ . \tag{B.79}$$

A projection P on V defines a subspace $PV = \operatorname{Img} P$, and $\mathbf{u} \in PV$ if and only if $P\mathbf{u} = \mathbf{u}$. It also defines the complementary subspace $QV = \operatorname{Ker} P$, where $Q = I - P$, and $\mathbf{v} \in QV$ if and only if $P\mathbf{v} = \mathbf{0}$. Conversely, two complementary subspaces of V define uniquely a projection P projecting onto one of the subspaces. Thus, in order to define a projection P on V it is necessary and sufficient to specify the subspace PV as well as the complementary subspace QV.

B.5 Trace and determinant

Definition B.23 *The* "trace" *of an $n \times n$ matrix $\mathbf{A}$ is the sum of its diagonal elements,*

$$\operatorname{Tr} \mathbf{A} = \sum_{i=1}^{n} A_{ii} \ . \tag{B.80}$$

Theorem B.24 *Let $\mathbf{A}$ be an $n \times m$ matrix and $\mathbf{B}$ an $m \times n$ matrix. Then* $\operatorname{Tr}(\mathbf{AB}) = \operatorname{Tr}(\mathbf{BA})$.

Here $\mathbf{AB}$ is an $n \times n$ matrix, whereas $\mathbf{BA}$ an $m \times m$ matrix. An immediate consequence is that only square matrices can be invertible, for if $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_m$, then

$$n = \operatorname{Tr} \mathbf{I}_n = \operatorname{Tr}(\mathbf{AB}) = \operatorname{Tr}(\mathbf{BA}) = \operatorname{Tr} \mathbf{I}_m = m \ . \tag{B.81}$$

This is an elegant proof of the fact that $\mathbb{F}^m$ and $\mathbb{F}^n$ can be isomorphic only if $m = n$.

Note however that the relation $n = \operatorname{Tr} \mathbf{I}_n$ used here is generally valid only if the number field $\mathbb{F}$ has characteristic 0. In a field of characteristic p, e.g. in $\mathbb{F} = \mathbb{Z}_p$ with p prime, we have only the weaker relation $n \equiv \operatorname{Tr} \mathbf{I}_n \pmod p$.

By Theorem B.24 we have that

$$\operatorname{Tr}(\mathbf{S}^{-1}\mathbf{AS}) = \operatorname{Tr}((\mathbf{S}^{-1}\mathbf{A})\mathbf{S}) = \operatorname{Tr}(\mathbf{S}(\mathbf{S}^{-1}\mathbf{A})) = \operatorname{Tr} \mathbf{A} \ , \tag{B.82}$$

a similarity transformation does not change the trace of a matrix. This means that the trace is an intrinsic property of a linear transformation $A : V \to V$, independent of the choice of basis, although one needs a matrix representation relative to a particular basis in order to compute the value of the trace.

Theorem B.25 *Let $A : V \to W$ and $B : W \to V$ be linear transformations. Then* $\operatorname{Tr}(AB) = \operatorname{Tr}(BA)$.

Let $A : V \to V$ and $B : W \to W$ be linear transformations. If $B = SAS^{-1}$ where $S : V \to W$ is an invertible linear transformation, then $\operatorname{Tr} B = \operatorname{Tr} A$.

Definition B.26 *The* "determinant" *of an $n \times n$ matrix $\mathbf{A}$ is*

$$\det \mathbf{A} = \sum_{p \in S_n} \operatorname{sgn}(p) \prod_{i=1}^{n} A_{p(i)i} \ , \tag{B.83}$$

where the sum is over all permutations $p \in S_n$.

Exercise B.27 *Show that transposition of a matrix leaves the determinant unchanged,*

$$\det(\mathbf{A}^\top) = \det \mathbf{A}\ . \tag{B.84}$$

Alternatively, let $f : X \to X$ be any function on the index set $X = \{1, 2, \ldots, n\}$, not necessarily a permutation, and define

$$\epsilon(f(1), f(2), \ldots, f(n)) = \begin{cases} \operatorname{sgn}(f) & \text{if } f \text{ is a permutation,} \\ 0 & \text{if } f \text{ is non-invertible.} \end{cases} \tag{B.85}$$

Then we may define the determinant as

$$\det \mathbf{A} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \epsilon(i_1, i_2, \ldots, i_n) A_{i_1 1} A_{i_2 2} \ldots A_{i_n n}\ . \tag{B.86}$$

If we exclude the possibility that the number field $\mathbb{F}$ has characteristic 2, then it is easy to see that the antisymmetry of the function ϵ defines it uniquely, up to a proportionality constant. In particular, if we define

$$\alpha(j_1, j_2, \ldots, j_n) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \epsilon(i_1, i_2, \ldots, i_n) A_{i_1 j_1} A_{i_2 j_2} \ldots A_{i_n j_n}\ , \tag{B.87}$$

then α is completely antisymmetric and must be proportional to ϵ. Taking $j_k = k$ for $k = 1, 2, \ldots, n$ we see that the proportionality factor is the determinant of $\mathbf{A}$, so that

$$\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \epsilon(i_1, i_2, \ldots, i_n) A_{i_1 j_1} A_{i_2 j_2} \ldots A_{i_n j_n} = (\det \mathbf{A})\, \epsilon(j_1, j_2, \ldots, j_n)\ . \tag{B.88}$$

Using this relation we may compute in a straightforward way the determinant of a matrix product.

Theorem B.28 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $n \times n$ *matrices. Then* $\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B})$*.*

It follows that $\det(\mathbf{A}^{-1}) = (\det \mathbf{A})^{-1}$, since

$$(\det(\mathbf{A})^{-1})(\det \mathbf{A}) = \det(\mathbf{A}^{-1}\mathbf{A}) = \det \mathbf{I} = 1\ . \tag{B.89}$$

Like the trace, the determinant is invariant under similarity transformations,

$$\det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}) = (\det \mathbf{S})^{-1}(\det \mathbf{A})(\det \mathbf{S}) = \det \mathbf{A}\ . \tag{B.90}$$

Therefore it is independent of any particular basis used for computing it, and is an intrinsic property of a linear transformation.

n arbitrary vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{F}^n$ may be put together into an $n \times n$ matrix, and the determinant $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ may be interpreted as the *volume* of the parallelepiped spanned by these vectors. Since

$$\det(\mathbf{A}\mathbf{v}_1, \mathbf{A}\mathbf{v}_2, \ldots, \mathbf{A}\mathbf{v}_n) = \det\left(\mathbf{A}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)\right) = (\det \mathbf{A})(\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n))\ , \tag{B.91}$$

we see that the linear transformation $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ on $\mathbb{F}^n$ changes volumes by the factor $\det \mathbf{A}$.

Exercise B.29 *Prove the following results about* $\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ *regarded as a function of* n *arguments (column vectors)* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

– *It is linear in any one argument,*

$$\det(\ldots, a\mathbf{u} + b\mathbf{v}, \ldots) = a \det(\ldots, \mathbf{u}, \ldots) + b \det(\ldots, \mathbf{v}, \ldots) . \tag{B.92}$$

– *It is antisymmetric in any pair of arguments,*

$$\det(\ldots, \mathbf{u}, \ldots, \mathbf{v}, \ldots) = -\det(\ldots, \mathbf{v}, \ldots, \mathbf{u}, \ldots) . \tag{B.93}$$

– *Therefore, for arbitrary coefficients* a_i*,*

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = \det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j - \sum_{i \neq j} a_i \mathbf{v}_i, \ldots, \mathbf{v}_n) . \tag{B.94}$$

Use these results to prove the following theorem.

Theorem B.30 *The* n *vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{F}^n$ *are linearly independent (that is, they form a basis for* $\mathbb{F}^n$*) if and only if*

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) \neq 0 . \tag{B.95}$$

Theorem B.31 *The* $n \times n$ *matrix* $\mathbf{A}$ *is invertible if and only if* $\det \mathbf{A} \neq 0$*.*

Proof. If $\mathbf{A}$ is invertible, then $\det \mathbf{A} \neq 0$, because $(\det(\mathbf{A}^{-1}))(\det \mathbf{A}) = 1$.

Conversely, if $\det \mathbf{A} \neq 0$, and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ is a basis for $\mathbb{F}^n$, then $\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2, \ldots, \mathbf{A}\mathbf{e}_n$ is also a basis, since

$$\det(\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2, \ldots, \mathbf{A}\mathbf{e}_n) = (\det \mathbf{A})(\det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)) \neq 0 . \tag{B.96}$$

Therefore $\mathbf{A}$ is invertible.

QED

The famous *Cramer's rule* is an explicit formula for the inverse $\mathbf{B} = \mathbf{A}^{-1}$ of a matrix $\mathbf{A}$ with $\det \mathbf{A} = 0$. A matrix element B_{jk} is given by the *cofactor* of A_{kj} in the determinant of $\mathbf{A}$,

$$(\det \mathbf{A}) B_{jk} = \sum_{p \in S_n,\, p(j)=k} \operatorname{sgn}(p) \prod_{i \neq j} A_{p(i)i} . \tag{B.97}$$

A matrix is said to be *singular* if its determinant is zero. A singular matrix has at least one eigenvector of eigenvalue zero.

B.6 Eigenvalues and eigenvectors

Definition B.32 *Given a linear transformation* $A : V \to V$*. A vector* $\mathbf{u} \in V$ *is an* "eigenvector" *of* A *with* "eigenvalue" $\lambda \in \mathbb{F}$ *if* $\mathbf{u} \neq \mathbf{0}$ *and* $A\mathbf{u} = \lambda\mathbf{u}$*.*

A "complete set" *of eigenvectors for* A *is a basis* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ *for* V *such that all the basis vectors are eigenvectors,*

$$A\mathbf{u}_i = \lambda_i \mathbf{u}_i . \tag{B.98}$$

Theorem B.33 *If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are eigenvectors of A corresponding to m different eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, then they are linearly independent.*

Proof. Assume that

$$\sum_{i=1}^{m} a_i \mathbf{u}_i = \mathbf{0} \; . \tag{B.99}$$

We want to prove that all the coefficients a_i must vanish. From the fact that

$$(A - \lambda_j I)\mathbf{u}_i = (\lambda_i - \lambda_j)\mathbf{u}_i \; , \tag{B.100}$$

follows that

$$\left(\prod_{j=2}^{m} (A - \lambda_j I) \right) \left(\sum_{i=1}^{m} a_i \mathbf{u}_i \right) = a_1 \left(\prod_{j=2}^{m} (\lambda_1 - \lambda_j) \right) \mathbf{u}_1 = \mathbf{0} \; . \tag{B.101}$$

Since, by assumption, $\mathbf{u}_1 \neq \mathbf{0}$ and the eigenvalues are all different, it follows that $a_1 = 0$. In the same way we prove that $a_2 = \ldots = a_m = 0$.

QED

Assume now that the dimension of V is finite, $\dim V = n$. Then no more than n vectors in V can be linearly independent, hence A can have at most n different eigenvalues. Furthermore, since n linearly independent vectors must form a basis for V, we have the following result.

Theorem B.34 *If A has $n = \dim V$ different eigenvalues, then it has a complete set of eigenvectors.*

We will come back to other conditions which also guarantee the existence of a complete set of eigenvectors, for example that A be either Hermitean or unitary.

Since every vector space of dimension n is isomorphic to $\mathbb{F}^n$, there is no loss of generality if we assume that $V = \mathbb{F}^n$ and identify A with an $n \times n$ matrix $\mathbf{A}$. To see exactly how a general eigenvalue problem is related to a matrix eigenvalue problem, recall that in a given basis [$\mathbf{e}$] the linear transformation A is represented by a matrix $\mathbf{A}$ such that

$$A\mathbf{e}_i = \sum_{j=1}^{n} A_{ji} \mathbf{e}_j \; . \tag{B.102}$$

Relative to a basis [$\mathbf{u}$] of eigenvectors, A is represented by a diagonal matrix

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} . \tag{B.103}$$

If every eigenvector $\mathbf{u}_i$ is expressed in the basis [$\mathbf{e}$] as

$$\mathbf{u}_i = \sum_{j=1}^{n} S_{ji} \mathbf{e}_j \; , \tag{B.104}$$

then the matrices $\mathbf{A}$ and $\mathbf{D}$ are related by the similarity transformation

$$\mathbf{D} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\ . \tag{B.105}$$

In this way the eigenvalue problem: to find a complete set of eigenvectors and eigenvalues for a linear transformation A on an n dimensional vector space, is equivalent to the so called *algebraic* eigenvalue problem: to find an $n \times n$ matrix $\mathbf{S}$ which diagonalizes an $n \times n$ matrix $\mathbf{A}$ by a similarity transformation.

We have already shown that an $n \times n$ matrix $\mathbf{A}$ can have at most n different eigenvalues. This result follows also from the next theorem, which in turn follows from Theorem B.18 and Theorem B.31.

Theorem B.35 *λ is an eigenvalue of $\mathbf{A}$, that is, the eigenvalue equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$ has at least one solution $\mathbf{u} \neq \mathbf{0}$, if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$.*

Definition B.36 *The* "characteristic polynomial" *of an $n \times n$ matrix $\mathbf{A}$ is*

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})\ . \tag{B.106}$$

The equation $p(\lambda) = 0$ is called the "characteristic equation" *of $\mathbf{A}$.*
An eigenvalue λ is "degenerate" *if it is a multiple root of the characteristic polynomial.*

If we write

$$p(\lambda) = p_0 - p_1\lambda + \ldots + p_{n-1}(-\lambda)^{n-1} + (-\lambda)^n\ , \tag{B.107}$$

then $p_0 = \det \mathbf{A}$ and $p_{n-1} = \mathrm{Tr}\,\mathbf{A}$. We have seen that the trace and the determinant are both invariants of a matrix, in the sense that they are invariant under similarity transformations (which correspond to changes of basis). The same is true for all the coefficients p_k. In fact, the characteristic polynomial as a whole is invariant under a similarity transformation $\mathbf{A} \mapsto \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$,

$$\det(\mathbf{S}^{-1}\mathbf{A}\mathbf{S} - \lambda\mathbf{I}) = \det(\mathbf{S}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{S}) = \det(\mathbf{A} - \lambda\mathbf{I})\ . \tag{B.108}$$

That the roots of the characteristic polynomial, the eigenvalues, are invariant, can also be seen directly. For if $\mathbf{u}$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ, $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, then $\mathbf{S}^{-1}\mathbf{u}$ is an eigenvector of $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ with the same eigenvalue,

$$(\mathbf{S}^{-1}\mathbf{A}\mathbf{S})(\mathbf{S}^{-1}\mathbf{u}) = \mathbf{S}^{-1}\mathbf{A}\mathbf{u} = \lambda(\mathbf{S}^{-1}\mathbf{u})\ . \tag{B.109}$$

p is a polynomial of degree n and so has at most n roots. The number of roots depends not only on the polynomial p but also on the number field $\mathbb{F}$. Let us assume that $\mathbb{F}$ is algebraically complete, for example that $\mathbb{F} = \mathbb{C}$, so that every polynomial of degree n has n roots and can be factored in linear factors, in particular

$$p(\lambda) = \prod_{i=1}^{n}(\lambda_i - \lambda)\ . \tag{B.110}$$

The roots $\lambda_1, \ldots, \lambda_n$ need not be all different, but if they are, in other words if no eigenvalue is degenerate, then there exists a complete set of eigenvectors.

Example. The simplest example of a matrix which has no complete set of eigenvectors, is the 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} . \tag{B.111}$$

Its characteristic polynomial is

$$p(\lambda) = \begin{vmatrix} -\lambda & 1 \\ 0 & -\lambda \end{vmatrix} = \lambda^2 , \tag{B.112}$$

with the double root $\lambda = 0$. It has one eigenvector

$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{B.113}$$

with the eigenvalue 0, but not two linearly independent eigenvectors. Note that $p(\mathbf{A}) = \mathbf{A}^2 = 0$. This is an instance of a general theorem.

Theorem B.37 (Cayley—Hamilton) *Every $n \times n$ matrix $\mathbf{A}$ satisfies its own characteristic equation. That is,*

$$p(\mathbf{A}) = p_0\mathbf{I} - p_1\mathbf{A} + \ldots + p_{n-1}(-\mathbf{A})^{n-1} + p_n(-\mathbf{A})^n = 0 . \tag{B.114}$$

Proof. We must prove that $p(\mathbf{A})\mathbf{u} = \mathbf{0}\ \forall \mathbf{u} \in \mathbb{F}^n$. If $\mathbf{u}$ is an eigenvector, $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, then λ is a root of $p(\lambda)$, and $p(\mathbf{A})\mathbf{u} = p(\lambda)\mathbf{u} = \mathbf{0}$. The theorem follows if there exists a complete set of eigenvectors, otherwise a more elaborate proof is needed.

The factorization

$$\mathbf{A}^k - \lambda^k\mathbf{I} = (\mathbf{A}^{k-1} + \lambda\mathbf{A}^{k-2} + \ldots + \lambda^{k-1}\mathbf{I})(\mathbf{A} - \lambda\mathbf{I}) \tag{B.115}$$

is valid for $k = 1, 2, \ldots$, and gives that

$$p(\mathbf{A}) - p(\lambda)\mathbf{I} = \sum_{k=0}^{n} p_k(-1)^k(\mathbf{A}^k - \lambda^k\mathbf{I}) = \mathbf{B}(\lambda)(\mathbf{A} - \lambda\mathbf{I}) , \tag{B.116}$$

where the matrix $\mathbf{B}(\lambda)$ is a polynomial in λ of degree $n-1$. A factorization of this kind holds for any polynomial. For the characteristic polynomial p we have in addition the factorization

$$p(\lambda)\mathbf{I} = (\det(\mathbf{A} - \lambda\mathbf{I}))\mathbf{I} = \mathbf{C}(\lambda)(\mathbf{A} - \lambda\mathbf{I}) , \tag{B.117}$$

with $\mathbf{C}(\lambda)$ a polynomial in λ of degree $n-1$. In fact, $\mathbf{C}(\lambda)$ is the transposed of the matrix of cofactors in the determinant $\det(\mathbf{A} - \lambda\mathbf{I})$ (this is Cramer's rule for inverting the matrix $\mathbf{A} - \lambda\mathbf{I}$). Consequently,

$$p(\mathbf{A}) = (\mathbf{B}(\lambda) + \mathbf{C}(\lambda))(\mathbf{A} - \lambda\mathbf{I}) . \tag{B.118}$$

But the matrix $p(\mathbf{A})$ is independent of the variable λ, hence the only way the equation can be satisfied is that $\mathbf{B}(\lambda) + \mathbf{C}(\lambda) = 0$ and $p(\mathbf{A}) = 0$.

QED

It may happen of course that an $n \times n$ matrix $\mathbf{A}$ is the root of a polynomial of lower degree than n. For example, if $\mathbf{A}$ is an element of a finite group of linear transformations, then $\mathbf{A}^m = \mathbf{I}$ where m is the order of $\mathbf{A}$ as a group element. The following theorem may then be useful.

Theorem B.38 *If $\mathbf{A}$ is diagonal and has m different eigenvalues $\mu_1, \mu_2, \ldots, \mu_m$, then the polynomial of lowest degree of which $\mathbf{A}$ is a root, is*

$$q(\lambda) = \prod_{i=1}^{m} (\lambda - \mu_i) \ . \tag{B.119}$$

Conversely, if $\mathbf{A}$ is the root of a polynomial of degree m with m different complex roots, then $\mathbf{A}$ is fully diagonalizable, and its eigenvalues are roots of the same polynomial.

Proof. We prove only the "converse" part of the theorem. Assume that

$$\prod_{i=1}^{m} (\mathbf{A} - \mu_i \mathbf{I}) = \mathbf{0} \ , \tag{B.120}$$

where the roots $\mu_1, \mu_2, \ldots, \mu_m$ are all different. The case $m = 1$ is trivial, so we assume that $m \geq 2$. Then the m matrices

$$\mathbf{P}_i = \prod_{h \neq i} \frac{\mathbf{A} - \mu_h \mathbf{I}}{\mu_i - \mu_h} \tag{B.121}$$

are projection operators satisfying the equations

$$\mathbf{P}_i \mathbf{P}_j = \delta_{ij} \mathbf{P}_j , \qquad \sum_{i=1}^{m} \mathbf{P}_i = \mathbf{I} \ , \qquad \sum_{i=1}^{m} \mu_i \mathbf{P}_i = \mathbf{A}. \tag{B.122}$$

To see this, note that $(\mathbf{A} - \mu_i \mathbf{I})\mathbf{P}_i = \mathbf{0}$, or equivalently, $\mathbf{A}\mathbf{P}_i = \mu_i \mathbf{P}_i$. Hence,

$$\mathbf{P}_i \mathbf{P}_j = \left(\prod_{h \neq i} \frac{\mathbf{A} - \mu_h \mathbf{I}}{\mu_i - \mu_h} \right) \mathbf{P}_j = \left(\prod_{h \neq i} \frac{\mu_j - \mu_h}{\mu_i - \mu_h} \right) \mathbf{P}_j = \delta_{ij} \mathbf{P}_j \ . \tag{B.123}$$

Furthermore, $\mathbf{P}_i$ is a polynomial in $\mathbf{A}$ of degree $m-1$, $\mathbf{P}_i = p_i(\mathbf{A})$. Since $\lambda = \mu_i$ is a root of the polynomial equation $p_i(\lambda) = 1$, we have a factorization

$$1 - p_i(\lambda) = (\lambda - \mu_i) q_i(\lambda) \ , \tag{B.124}$$

where $q_i(\lambda)$ is a polynomial in λ of degree $m-2$. It follows from the above relations that

$$\mathbf{I} - \sum_{i=1}^{m} \mathbf{P}_i = \prod_{i=1}^{m} (\mathbf{I} - \mathbf{P}_i) = \prod_{i=1}^{m} (\mathbf{A} - \mu_i \mathbf{I}) \prod_{i=1}^{m} q_i(\mathbf{A}) = \mathbf{0} \ . \tag{B.125}$$

Finally, multiplying this equation from the left with $\mathbf{A}$, we get that

$$\mathbf{A} = \sum_{i=1}^{m} \mu_i \mathbf{P}_i \ . \tag{B.126}$$

QED