LINEAR ALGEBRA, MATH 50003: Lecture Notes

Lecturer: Martin Liebeck

1 Course Overview

Most of the course will consist of basic results on matrices, vector spaces and linear maps. The last part of the course will have a more geometrical flavour.

1.1 Matrix results

Let's begin with a survey of some of the highlights among the matrix results in the course. We start with a definition.

Definition Let A, B be $n \times n$ matrices over a field F. We say A is *similar* to B if there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$.

Note that if we define a relation $\sim$ on $n \times n$ matrices by

$$A \sim B \Leftrightarrow A \text{ is similar to } B,$$

then $\sim$ is an equivalence relation (question on Problem Sheet 1).

Two similar matrices A, B share many basic properties: for example, they have

- the same determinant
- the same characteristic polynomial
- the same eigenvalues
- the same rank
- the same trace

(question on Problem Sheet 1). One of the major aims of the subject is:

Major Aim For an arbitrary $n \times n$ matrix A, find a "nice" matrix B such that $A \sim B$.

In the course we'll prove three famous therorems, in each of which the meaning of the word "nice" will be apparent.

Example Probably the nicest matrices are the diagonal ones. Recall that an $n \times n$ matrix A is *diagonalisable* if it is similar to a diagonal matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ (the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, the eigenvalues of A). This property can be used to do many computations with A, such as calculating any power A^k: a matrix P such that $D = P^{-1}AP$ can be computed (its columns are a basis of eigenvectors of A). Then $A = PDP^{-1}$, so

$$A^k = (PDP^{-1})\,(PDP^{-1}) \cdots (PDP^{-1}) = PD^kP^{-1},$$

and D^k is the diagonal matrix $D = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k)$.

However, many matrices are not diagonalisable, for example

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

To see this, suppose that A is diagonalisable. Then since the only eigenvalue is 1, there exist P such that $P^{-1}AP = \text{diag}(1,1) = I$, so $A = PIP^{-1} = I$, a contradiction.

So not every matrix can be diagonalised. However, every complex matrix can be *triangularised*. This is one of the first main results of the course:

Triangularisation Theorem *If A is an $n \times n$ matrix over $\mathbb{C}$, then A is similar to an upper triangular matrix, i.e. there exists P such that*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & & & \\ 0 & \lambda_2 & & * & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & 0 & & & \lambda_n \end{pmatrix}.$$

Note that this result does not hold for matrices aver arbitrary fields: for example over the real numbers $\mathbb{R}$, the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has complex eigenvalues $\pm i$, so is not similar to a real upper triangular matrix.

The theorem has a more serious drawback though: there is nothing unique about an upper triangular matrix similar to A. For example, for any $a, b, a', b' \neq 0$,

$$\begin{pmatrix} 1 & a & b \\ & 1 & 0 \\ & & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & a' & b' \\ & 1 & 0 \\ & & 1 \end{pmatrix},$$

(question on Sheet 1), so if A is similar to one such matrix, it is similar to all of them.

It is very desirable to have a *unique* matrix of a nice form that is similar to A, and that is provided by the next main result.

Jordan Canonical Form Theorem *If A is an $n \times n$ matrix over $\mathbb{C}$, then A is similar to a matrix of the form*

$$J = \begin{pmatrix} J_1 & & & & \\ & J_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & J_k \end{pmatrix},$$

a block-diagonal matrix with blocks

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda_i & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{pmatrix}$$

(these are called Jordan blocks). The collection of Jordan blocks $J_1, \dots, J_k$ is uniquely determined by A.

We call the matrix J the Jordan Canonical Form (JCF) of A. Its uniqueness is a vital part of the theorem, since it gives a powerful test for the similarity of two arbitrary complex matrices A and B: find the JCFs of A and B, call them J and J'. If J and J' are the same (apart from changing the order in which the Jordan blocks appear), then $A \sim B$; if not, then $A \not\sim B$. This test can be programmed very efficiently, and can be used for huge matrices.

The Jordan Canonical Form Theorem is an ideal result for complex matrices. But what about matrices over other fields, such as $\mathbb{R}$ or $\mathbb{Q}$ or the finite field $\mathbb{F}_p$ (the field of prime order p consisting of the integers $0, 1, \ldots, p-1$ with addition and multiplication modulo p)? The JCF theorem does not hold for arbitrary matrices over these fields, for the same reason that the Triangularisation theorem does not hold.

However we will prove another canonical form theorem – the Rational Canonical Form – that holds over arbitrary fields. To state this, we need a bit of notation. Let F be a field, and denote by $F[x]$ the set of polynomials in x over F. We can add and multiply polynomials (indeed, under addition and multiplication they form what is called a *ring*).

We call a polynomial $p(x) \in F[x]$ *monic* if it has degree $r \geq 1$ and its leading coefficient is 1, i.e.

$$p(x) = x^r + a_{r-1}x^{r-1} + \cdots + a_0. \tag{1}$$

Definition Let $p(x)$ be a monic polynomial of degree r as in (1). The *companion matrix* of $p(x)$ is the $r \times r$ matrix $C(p(x))$ defined as follows:

$$C(p(x)) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{r-1} \end{pmatrix}.$$

For example,

$$C(x^3 - x + 1) = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note that $C(p(x))$ has characteristic polynomial $p(x)$ (question on Sheet 1).

Rational Canonical Form Theorem *Let A be an $n \times n$ matrix over F, with characteristic polynomial $p(x)$.*

(i) There exists a factorization $p(x) = p_1(x) \cdots p_k(x)$ such that A is similar to a block-diagonal matrix with blocks $C(p_i(x))$ for $i = 1, \ldots k$.

(ii) Under some conditions, the polynomials $p_1(x), \ldots, p_k(x)$ are uniquely determined by A.

The "conditions" in part (ii) will be spelled out when we state and prove the theorem in the lectures.

1.2 Geometry

The last part of the course will be concerned with some geometrical aspects of linear algebra.

Recall the *dot product* on $\mathbb{R}^n$: if $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, then

$$u.v = \sum_{i=1}^{n} u_i v_i.$$

Much of the geometry of $\mathbb{R}^n$ is based on the dot product. For example, the length $||u|| = \sqrt{u.u}$, and the distance between u and v is $||u - v||$. Various types of $n \times n$ matrices fit naturally into this geometrical picture, for example

- P is *orthogonal* if $P^T P = I$ (which implies that $Pu.Pv = u.v$ for all u, v)
- A is *symmetric* if $A^T = A$ (which implies that $Au.v = u.Av$ for all u, v).

It is useful to axiomatise the basic properties of the dot product, to obtain the theory of *inner product spaces*: an inner product space is a real vector space with a map sending any pair of vectors u, v to a scalar (u, v) satisfying the following axioms:

(1) the map is linear in each variable u, v

(2) the map is symmetric, i.e. $(v, u) = (u, v)$ for all u, v

(3) $(u, u) > 0$ for all nonzero vectors u.

We shall develop the theory of inner product spaces. In order to extend the geometrical notions to vector spaces over arbitrary fields, we shall also develop the theory of bilinear forms.

2 Some revision from 1st Year Linear Algebra

This chapter is a summary of some of the theory of matrices and linear maps from the 1st year course that we'll need.

Let V be a finite dimensional vector space over a field F and $T : V \to V$ a linear map. If $B = \{v_1, \ldots, v_n\}$ is a basis of V, let

$$\begin{aligned} T(v_1) &= a_{11}v_1 + \ldots + a_{n1}v_n, \\ &\vdots \\ T(v_n) &= a_{1n}v_1 + \ldots + a_{nn}v_n \end{aligned}$$

where all the coefficients $a_{ij} \in F$. The *matrix of T with respect to B* is

$$[T]_B = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Proposition 2.1 *Let $S : V \to V$ and $T : V \to V$ be linear transformations and let B be a basis of V. Then*

$$[ST]_B = [S]_B[T]_B,$$

where ST is the composition of S and T.

As a consequence of the proposition, the map $T \to [T]_B$ from linear maps to $n \times n$ matrices has many nice properties. For example, if $[T]_B = A$ then $[T^2]_B = A^2$ and similarly $\left[T^k\right]_B = A^k$ for any positive integer k. More generally, for a polynomial $q(x) = a_r x^r + \cdots + a_1 x + a_0$ ($a_i \in F$), define

$$q(A) = a_r A^r + \cdots + a_1 A + a_0 I$$

and

$$q(T) = a_r T^r + \cdots + a_1 T + a_0 I_V$$

where $I_V : V \to V$ is the identity map. Then Proposition 2.1 implies that

$$[q(T)]_B = q(A).$$

Change of basis

Let V be n-dimensional, and let bases $E = \{e_1, \ldots, e_n\}$ and $F = \{f_1, \ldots, f_n\}$ be two bases of V. Write

$$\begin{aligned} f_1 &= p_{11} e_1 + \cdots + p_{n1} e_n, \\ &\vdots \\ f_n &= p_{1n} e_1 + \cdots + p_{nn} e_n. \end{aligned}$$

and define P to be the $n \times n$ matrix (p_{ij}). We call P the *change of basis matrix* from E to F.

Proposition 2.2 (i) *The change of basis matrix P is invertible.*

(ii) *If $T : V \to V$ is a linear map, then $[T]_F = P^{-1}[T]_E P$ (so $[T]_E$ and $[T]_F$ are similar matrices).*

Determinants

As we already noted in Chapter 1, if A, B are similar $n \times n$ matrices, then they have the same determinant. Hence if $T : V \to V$ is a linear map, and E, F are two bases of V, then the matrices $[T]_E$ and $[T]_F$ have the same determinant (by Proposition 2.2(ii)). Therefore we can define the determinant $\det(T)$ of a linear map T to be the determinant of the matrix $[T]_E$ for any basis E of V. The *characteristic polynomial* of T is defined to be $\det(xI_V - T)$. This is a polynomial in x of degree $n = \dim V$.

Proposition 2.3 (i) *The eigenvalues of T are the roots of the characteristic polynomial of T.*

(ii) *If λ is an eigenvalue of T, the eigenvectors corresponding to λ are the nonzero vectors in*

$$E_\lambda = \{v \in V : (\lambda I_V - T)(v) = 0\} = \ker(\lambda I_V - T).$$

(iii) *The matrix $[T]_B$ is a diagonal matrix iff B consists of eigenvectors of T.*

Definition We call E_λ the λ-*eigenspace* of T. Note that E_λ is a subspace of V (since it is the kernel of the linear map $\lambda I_V - T$).

Proposition 2.4 *Let V a finite-dimensional vector space over $\mathbb{C}$, and let $T : V \to V$ be a linear map. Then T has an eigenvalue $\lambda \in \mathbb{C}$.*

Proof The characteristic polynomial of T has a root $\lambda \in \mathbb{C}$ by the Fundamental theorem of Algebra. $\square$

Note that Proposition 2.4 is not necessarily true for vector spaces over other fields. For example $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1, x_2) = (x_2, -x_1)$ has characteristic polynomial $x^2 + 1$, which has no real roots.

Diagonalisation

Recall that a linear map $T : V \to V$ is diagonalisable iff there exists a basis of V consisting of eigenvectors of T. Here is a very useful result on eigenvectors.

Proposition 2.5 *Let $T : V \to V$ be a linear map. Suppose $v_1, \ldots, v_k$ are eigenvectors of T corresponding to* distinct *eigenvalues $\lambda_1, \ldots, \lambda_k$. Then $v_1, \ldots, v_k$ are linearly independent.*

Corollary 2.6 *Let V be n-dimensional over F, and let $T : V \to V$ a linear map. Suppose the characteristic polynomial of T has n distinct roots in F. Then T is diagonalisable.*

Example Let

$$A = \begin{pmatrix} \lambda_1 & & & \\ 0 & \lambda_2 & * & \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

be upper triangular, with diagonal entries $\lambda_1, \ldots, \lambda_n$, all distinct. The characteristic polynomial of A is $\prod_{i=1}^{n}(x - \lambda_i)$, which has roots $\lambda_1, \ldots, \lambda_n$. Hence by Corollary 2.6, A is diagonalisable, so there exists P such that $P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Note that this is not necessarily true if the diagonal entries are not distinct, e.g. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalisable.

As a final point about diagonalisation, it is sometimes important to specify which field we are working over. If A is an $n \times n$ matrix over a field F, we say A is diagonalisable *over F* if it is similar to a diagonal matrix with entries in F. For example, the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is not diagonalisable over $\mathbb{R}$, but it *is* diagonalisable over $\mathbb{C}$.

3 Algebraic and geometric multiplicities of eigenvalues

In this chapter we introduce and study two types of eigenvalue multiplicity.

Definition Let $T : V \to V$ be a linear map with characteristic polynomial $p(x)$. Let λ be a root of $p(x)$ (i.e. an eigenvalue of T). Then there is a positive integer $a(\lambda)$ such that

$$p(x) = (x - \lambda)^{a(\lambda)} q(x),$$

where λ is not a root of $q(x)$. We call $a(\lambda)$ the *algebraic multiplicity* of λ as an eigenvalue of T.

The *geometric multiplicity* of λ is defined to be

$$g(\lambda) = \dim E_\lambda,$$

where E_λ is the λ-eigenspace of T.

We adopt similar definitions for $n \times n$ matrices.

Example For $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, we have

$$a(1) = g(1) = 1, \;\; a(2) = g(2) = 1.$$

And for $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we have

$$a(1) = 2, g(1) = 1.$$

Proposition 3.1 *If λ is an eigenvalue of $T : V \to V$, then $g(\lambda) \leq a(\lambda)$.*

Proof Let $r = g(\lambda) = \dim E_\lambda$ and let $v_1, \dots, v_r$ be a basis of E_λ. Extend to a basis of V:

$$B = \{v_1, \dots, v_r, w_1, \dots, w_s\}.$$

We work out the matrix $[T]_B$:

$$\begin{aligned} T(v_1) &= \lambda v_1, \\ &\vdots \\ T(v_r) &= \lambda v_r, \\ T(w_1) &= a_{11}v_1 + \cdots + a_{r1}v_r + b_{11}w_1 + \cdots + b_{s1}w_s, \\ &\vdots \\ T(w_s) &= a_{1s}v_1 + \cdots + a_{rs}v_r + b_{1s}w_1 + \cdots + b_{ss}w_s. \end{aligned}$$

So

$$[T]_B = \left(\begin{array}{cccc|ccc} \lambda & 0 & \cdots & 0 & a_{11} & \cdots & a_{1s} \\ 0 & \lambda & \cdots & 0 & \vdots & & \vdots \\ \vdots & \vdots & \ddots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & a_{r1} & \cdots & a_{rs} \\ \hline 0 & \cdots & \cdots & 0 & b_{11} & \cdots & b_{1s} \\ \vdots & & & \vdots & \vdots & & \vdots \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & b_{s1} & \cdots & b_{ss} \end{array}\right) = \begin{pmatrix} \lambda I_r & A \\ 0 & B \end{pmatrix}.$$

The characteristic polynomial of this is

$$p(x) = \det \left(\begin{array}{c|c} (x-\lambda)I_r & -A \\ \hline 0 & xI_s - B \end{array}\right).$$

Using Q4 on Sheet 1, this is

$$p(x) = \det((x-\lambda)I_r)\det(xI_s - B) = (x-\lambda)^r s(x),$$

where $s(x)$ is the characteristic polynomial of B. Hence the algebraic multiplicity $a(\lambda) \geq r = g(\lambda)$. $\square$

Using this we can prove the following basic criterion for diagonalisation.

Theorem 3.2 *Let* $\dim V = n$*, let* $T : V \to V$ *be a linear map, let* $\lambda_1, \dots, \lambda_r$ *be the distinct eigenvalues of* T*, and let the characteristic polynomial of* T *be*

$$p(x) = \prod_{i=1}^{r} (x - \lambda_i)^{a(\lambda_i)}$$

(so $\sum_{i=1}^{r} a(\lambda_i) = n$*). The following statements are equivalent:*

(1) T *is diagonalisable.*

(2) $\sum_{i=1}^{r} g(\lambda_i) = n$*.*

(3) $g(\lambda_i) = a(\lambda_i)$ *for all* i*.*

Proof We first prove (1) $\Rightarrow$ (2). Suppose (1) holds, so V has a basis B consisting of eigenvectors of T. Each vector in B is in some eigenspace E_{λ_i}, so

$$\sum_{i=1}^{r} g(\lambda_i) = \sum_{i=1}^{r} \dim E_{\lambda_i} \geq |B| = n.$$

By 3.1, $\sum_{i=1}^{r} g(\lambda_i) \leq \sum_{i=1}^{r} a(\lambda_i) = n$. Hence $\sum g(\lambda_i) = n$.

Next we show that (2) $\Leftrightarrow$ (3). This is easy, as

$$\sum g(\lambda_i) = n \Leftrightarrow \sum g(\lambda_i) = \sum a(\lambda_i) \Leftrightarrow g(\lambda_i) = a(\lambda_i) \ \forall i$$

(using 3.1 for the last implication).

To complete the proof, we show that (2) $\Rightarrow$ (1). Suppose (2) holds, so $\sum_{i=1}^{r} \dim E_{\lambda_i} = n$. Let B_i be a basis of E_{λ_i} and let $B = \bigcup_{i=1}^{r} B_i$, so $|B| = n$ (the B_i's are disjoint as they consist of eigenvectors for different eigenvalues).

We claim that B is a basis of V (hence (1) holds). Since $|B| = n = \dim V$, it is enough to show that B is linearly independent. Suppose there is a linear relation on the vectors in B, and write it as

$$\sum_{a \in B_1} \alpha_a a + \cdots + \sum_{z \in B_r} \alpha_z z = 0. \tag{2}$$

Write

$$\begin{aligned} v_1 &= \textstyle\sum_{a \in B_1} \alpha_a a, \\ &\vdots \\ v_r &= \textstyle\sum_{z \in B_r} \alpha_z z, \end{aligned}$$

so $v_i \in E_{\lambda_i}$ and $v_1 + \cdots + v_r = 0$. As $\lambda_1, \dots, \lambda_r$ are distinct, the set of nonzero v_i's is linearly independent by 2.5. Therefore there can't be any nonzero v_i's, and so $v_i = 0$ for all i. Then $v_1 = \sum_{a \in B_1} \alpha_a a = 0$, so as B_1 is linearly independent (it is a basis of E_{λ_1}) all the coefficients $\alpha_a = 0$. Similarly all the other α's in (2) are 0. This completes the proof that B is linearly independent, hence a basis of V. $\square$

Using 3.2 we obtain a test to check whether a given $n \times n$ matrix or linear map is diagonalisable:

1. Find the characteristic polynomial, and factorise it as

$$\prod_{i=1}^{r} (x - \lambda_i)^{a(\lambda_i)}.$$

2. Calculate each $g(\lambda_i) = \dim E_{\lambda_i}$.

3. If $g(\lambda_i) = a(\lambda_i)$ for all i, YES.
   If $g(\lambda_i) < a(\lambda_i)$ for some i, NO.

Example Let $A = \begin{pmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{pmatrix}$. Check that

(1) Characteristic polynomial is $(x+2)^2(x-4)$.

(2) For eigenvalue 4: $a(4) = 1, g(4) = 1$ (as it is $\leq a(4)$).
For eigenvalue -2: $a(-2) = 2$, $g(-2) = \dim E_{-2} = 1$.

So $g(-2) < a(-2)$ and A is not diagonalisable.

4 Direct sums

Recall that if $U_1, \ldots, U_k$ are subspaces of a vector space V, we can form their *sum*

$$U_1 + \cdots + U_k = \{u_1 + \cdots + u_k : u_i \in U_i \text{ for all } i\},$$

which is another subspace of V. A *direct sum* of subspaces is a particular case of this, defined as follows.

Definition Let V be a vector space, and let $V_1, \ldots, V_k$ be subspaces of V. We write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k \tag{3}$$

if every vector $v \in V$ can be expressed as $v = v_1 + \cdots + v_k$ for *unique* vectors $v_i \in V_i$. The uniqueness statement means that if $v_1 + \cdots + v_k = v_1' + \cdots + v_k'$ with $v_i, v_i' \in V_i$, then $v_i = v_i'$ for all i. If (3) holds, we say that V is the *direct sum* of the subspaces $V_1, \ldots, V_k$.

As an obvious first example, $\mathbb{R}^2 = \mathrm{Sp}(1,0) \oplus \mathrm{Sp}(0,1)$. (Here, and throughout these notes, "Sp" is an abbreviation for "Span".)

It will be important for us to be able to check quickly whether the direct sum condition (3) holds. For a direct sum of two subspaces (the case $k = 2$), this is easy:

Proposition 4.1 *The following statements are equivalent:*

(1) $V = V_1 \oplus V_2$.

(2) $V_1 \cap V_2 = \{0\}$ *and* $\dim V_1 + \dim V_2 = \dim V$.

Proof First we show (1) $\Rightarrow$ (2). Assume (1), so that $V = V_1 \oplus V_2$. If there exists $0 \neq v \in V_1 \cap V_2$, then

$$v = v + 0 = 0 + v$$

gives two different expressions for v as a sum of vectors in V_1 and V_2, contradicting the uniqueness statement in the definition of a direct sum. Therefore $V_1 \cap V_2 = \{0\}$. It follows that

$$\dim V = \dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim V_1 \cap V_2 = \dim V_1 + \dim V_2.$$

Hence (2) holds.

Now we show (2) $\Rightarrow$ (1). Assume that (2) holds. Then

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim V_1 \cap V_2 = \dim V_1 + \dim V_2 = \dim V.$$

Hence $V = V_1 + V_2$. To show uniqueness, suppose $v_1 + v_2 = v_1' + v_2'$ with $v_i, v_i' \in V_i$. Then

$$v_1 - v_1' = v_2' - v_2 \in V_1 \cap V_2.$$

Since $V_1 \cap V_2 = \{0\}$, this implies that $v_1 = v_1', v_2 = v_2'$. Hence $V = V_1 \oplus V_2$. $\square$

The next result shows how to check the direct sum condition (3) for arbitrary values of k.

Proposition 4.2 *The following statements are equivalent:*

(1) $V = V_1 \oplus \cdots \oplus V_k$.

(2) $\dim V = \sum_{i=1}^k \dim V_i$, *and if* B_i *is a basis for* V_i *for* $1 \le i \le k$, *then* $B = B_1 \cup \cdots \cup B_k$ *is a basis of* V.

Proof First we prove (1) $\Rightarrow$ (2). Assume that $V = V_1 \oplus \cdots \oplus V_k$. Let B_i be a basis of V_i for $1 \le i \le k$, and let $B = B_1 \cup \cdots \cup B_k$.

Claim B is a basis of V.

Proof of Claim: Clearly B spans V, since $V = V_1 + \cdots + V_k$. Now we show linear independence. Suppse there is a linear relation on the vectors in B, and write this as

$$\sum_{a \in B_1} \alpha_a a + \cdots + \sum_{z \in B_r} \alpha_z z = 0. \tag{4}$$

Now $V = V_1 \oplus \cdots \oplus V_k$, hence $0 = 0 + \cdots + 0$ is the *unique* expression for the zero vector as a sum of vectors in $V_1, \ldots, V_k$. Hence each sum in the left hand side of (4) is equal to 0, and so all the α's in (4) are 0. This proves that B is linearly independent, hence is a basis, proving the Claim.

As in the proof of 4.1 we see that $V_i \cap V_j = \{0\}$ for $i \neq j$, and hence $B_i \cap B_j = \emptyset$ and B is the disjoint union of the B_i. By the Claim, therefore, we have

$$\dim V = |B| = \sum_{i=1}^k |B_i| = \sum_{i=1}^k \dim V_i,$$

so that (2) holds.

Now we prove that (2) $\Rightarrow$ (1). Assume that (2) holds. For each i let B_i be a basis of V_i, and let $B = \bigcup_1^k B_i$, a basis of V. As $\dim V = \sum_1^k \dim V_i$, we have $|B| = \sum |B_i|$, so the B_i's are disjoint sets. Every vector in V is in the span of B, hence is a sum of vectors in $V_1, \ldots, V_k$, so $V = V_1 + \cdots + V_k$. To prove uniqueness, suppose that

$$v_1 + \cdots + v_k = v_1' + \cdots + v_k'$$

where each $v_i, v_i' \in V_i$. Then

$$0 = (v_1 - v_1') + \cdots + (v_k - v_k').$$

If any term $v_i - v_i'$ is nonzero, this equation will give a nontrivial linear relation on the vectors in the basis B, a contradiction. Hence $v_i = v_i'$ for all i, proving uniqueness, and so $V = V_1 \oplus \cdots \oplus V_k$. $\square$

Example In $\mathbb{R}^4$ let $V_1 = \mathrm{sp}((1,1,0,0),(0,-1,1,0))$, $V_2 = \mathrm{sp}(2,1,2,1)$, $V_3 = \mathrm{sp}(0,0,1,1)$. Is $\mathbb{R}^4 = V_1 \oplus V_2 \oplus V_3$?

Answer: no, as $\{(1,1,0,0),(0,-1,1,0),(2,1,2,1),(0,0,1,1)\}$ is not a basis of $\mathbb{R}^4$. (The simplest way to check this is to write the vectors as the rows of a 4×4 matrix and show that this can be reduced by row operations to a matrix with a zero row.)

To complete this chapter, we demonstrate an important link between direct sums and linear maps. First we need a definition.

Definition Let $T : V \to V$ be a linear map, and W a subspace of V. We say that W is T-*invariant* if $T(W) \subseteq W$, where $T(W) = \{T(w) : w \in W\}$ (in other words, T maps $W \to W$). If W is T-invariant, write $T_W : W \to W$ for the *restriction* of T to W. Thus T_W is the linear map $W \to W$ defined by $T_W(w) = T(w)$ for all $w \in W$.

Proposition 4.3 *Let $T : V \to V$ be a linear map, and suppose that $V = V_1 \oplus \cdots \oplus V_k$, where each subspace V_i is T-invariant. For each i let B_i be a basis of V_i, and let A_i be the matrix of the restriction $[T_{V_i}]_{B_i}$. Then if B is the basis $\bigcup_1^k B_i$ of V, the matrix $[T]_B$ is the block-diagonal matrix*

$$[T]_B = \begin{pmatrix} A_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & A_k \end{pmatrix}. \tag{5}$$

Proof Let $B_1 = \{v_1, \ldots, v_r\}$. Then $T(v_1) = T_{V_1}(v_1)$ is a vector in V_1, say $T(v_1) = a_{11}v_1 + \cdots + a_{r1}v_r$. Similarly for $T(v_2), \ldots$, up to $T(v_r) = T_{V_1}(v_r) = a_{1r}v_1 + \cdots + a_{rr}v_r$. So we see that the top left hand block of $[T]_B$ is the $r \times r$ matrix (a_{ij}), which is $[T_{V_1}]_{B_1}$. Carrying on like this, we see that the next diagonal block is $[T_{V_2}]_{B_2}$, and so on. $\square$

Notation In view of the proposition, and for convenience of notation, we shall denote the block-diagonal matrix in (5) by $A_1 \oplus \cdots \oplus A_k$. Thus for $n_i \times n_i$ matrices A_i $(1 \leq i \leq k)$, we write

$$A_1 \oplus \cdots \oplus A_k = \begin{pmatrix} A_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & A_k \end{pmatrix},$$

an $n \times n$ block-diagonal matrix, where $n = \sum_{i=1}^k n_i$.