Linear Algebra (2LALP/2LA)

Professor Sergey Shpectorov

2023/24 Academic Year

Contents

1 Vector spaces **3**

1.1 Fields and vector spaces 3

1.1.1 Fields 3

1.1.2 Finite fields 4

1.1.3 Vector Spaces 5

1.1.4 Basic properties 6

1.1.5 Examples of vector spaces 7

1.2 Subspaces, intersection and sum of subspaces 10

1.2.1 Subspaces 10

1.2.2 Subspace criterion 11

1.2.3 Intersection of subspaces 12

1.2.4 Span 13

1.2.5 Sum of subspaces, direct sum 14

1.3 Linear combinations and independence 16

1.3.1 Vectors in the span 16

1.3.2 Independent sets 18

1.3.3 Up and down 20

1.4 Bases 22

1.4.1 Bases 22

1.4.2 Standard bases 23

1.4.3 Dimension 24

1.4.4 Dimension of the sum of subspaces 26

1.5 Coordinates 27

1.5.1 Coordinates 27

1.5.2 Coordinate vector 28

1.5.3 Using coordinates 29

1.5.4 Coordinate mapping 30

2 Linear transformations **31**

2.1 Linear mappings 31

2.1.1 Linear mappings 31

2.1.2 Kernel and injectivity 34
2.1.3 Image, rank+nullity, surjectivity 36
2.1.4 Matrix of a linear mapping 37
2.2 Isomorphisms 40
2.2.1 Isomorphisms and inverse mappings 40
2.2.2 Checking for an isomorphism 41
2.2.3 Isomorphic vector spaces 43
2.3 Linear transformations and normal forms 45
2.3.1 Linear transformations 45
2.3.2 Matrix of a linear transformation 46
2.3.3 Eigenvalues and eigenvectors 48
2.3.4 Transition matrix 53
2.3.5 Similarity and normal forms 55

3 Bilinear forms and geometry 60
3.1 Basic definitions 60
3.1.1 Bilinear forms 60
3.1.2 The Gram matrix 61
3.1.3 Symmetric forms 63
3.1.4 Radical and rank 63
3.2 Bilinear forms over the real numbers 65
3.2.1 Inner product spaces 65
3.2.2 Length and distance 66
3.2.3 Orthogonality 67
3.2.4 Cauchy-Schwarz inequality 67
3.2.5 Angles 68
3.2.6 Orthogonal and orthonormal bases 69
3.2.7 Orthogonal projection 70
3.2.8 The Gram-Schmidt orthogonalization process 72

Chapter 1

Vector spaces

1.1 Fields and vector spaces

Definition of field, examples of fields. Definition of vector space, examples, properties

1.1.1 Fields

We first need to briefly discuss the concept of a field. Fields are number systems where we can do all the familiar operations with numbers and these operations satisfy the familiar properties.

Definition 1.1.1 *A* **field** *is a set* $\mathbb{F}$ *on which we have two binary operations:* $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ *and* $\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$*, satisfying the following axioms:*

(F1) (Associativity of addition) For all $a, b, c \in \mathbb{F}$*,* $a + (b + c) = (a + b) + c$*.*

(F2) (Commutativity of addition) For all $a, b \in \mathbb{F}$*,* $a + b = b + a$*.*

(F3) (Existence of zero) There exists $0 \in \mathbb{F}$ *such that, for all* $a \in \mathbb{F}$*,* $a + 0 = a = 0 + a$*.*

(F4) (Existence of additive inverses) For each $a \in \mathbb{F}$*, there exists* $-a \in \mathbb{F}$ *such that* $a + (-a) = 0 = -a + a$*.*

(F5) (Associativity of multiplication) For all $a, b, c \in \mathbb{F}$*,* $a(bc) = (ab)c$*.*

(F6) (Commutativity of multiplication) For all $a, b \in \mathbb{F}$*,* $ab = ba$*.*

(F7) (Existence of identity) There exists $1 \in \mathbb{F}$ *such that , for all* $a \in \mathbb{F}$*,* $a1 = a = 1a$*.*

(F8) (Existence of multiplicative inverses) For each $a \in F$, $a \neq 0$, there exists $a^{-1} \in \mathbb{F}$ such that $aa^{-1} = 1 = a^{-1}a$.

(F9) (Distributivity) For all $a, b, c \in \mathbb{F}$, $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$.

The elements of $\mathbb{F}$ may be called *numbers*, the operations $+$ and $\cdot$ will be called *addition* and *multiplication*, respectively. We can also talk about *subtraction* and *division*. However, please note that these operations are derived from $+$ and $\cdot$. Namely, when we write $a - b$, we mean $a + (-b)$ and, when we write $\frac{a}{b}$, we mean ab^{-1} (and of course, here $b \neq 0$).

The standard examples of fields are $\mathbb{Q}$ (the rationals), $\mathbb{R}$ (the reals), and $\mathbb{C}$ (the complex numbers). The integers $\mathbb{Z}$ do *not* form a field. Can you see why?

1.1.2 Finite fields

In this course we will also need finite fields. By this we mean a field $\mathbb{F}$ which is finite as a set. So there are only finitely many numbers in this number system!

Theorem 1.1.2 *If $\mathbb{F}$ is a finite field then $|\mathbb{F}| = p^f$ for some prime integer p and a positive integer f. Furthermore, for each order p^f, there is exactly one field of that order.*

This theorem tells us how many finite fields exist and what are the possible orders. However, we also need to know what the fields are. We will write $\mathbb{F}_q$ for the only field of order (or size) $q = p^f$.

Prime order

If the order of the field is p (that is, $f = 1$) then the field $\mathbb{F}_p$ is simply $\mathbb{Z}_p$, the integers modulo p. This means that $\mathbb{F}_p = \{0, 1, \ldots, p - 1\}$ and the operations of addition and multiplication on $\mathbb{F}_p$ are the usual addition and multiplication supplemented by subtracting a suitable multiple of p in order to bring the result within the interval $[0, p - 1]$.

Example 1.1.3 *Suppose we need to work with $\mathbb{F}_5$. Then $\mathbb{F}_5 = \{0, 1, 2, 3, 4\}$, and here are examples of addition:*

$$2 + 2 = 4,$$

$$4 + 3 = 7 - 5 = 2.$$

In the first example, the sum, 4 is already in the range, so we do not need to subtract 5. In the second example, we subtract 5 from the sum, 7, in order to bring the result into the range.

Here also are the examples of multiplication:

$$2 \cdot 2 = 4,$$

$$4 \cdot 4 = 16 - 5 - 5 - 5 = 1.$$

In the first example, we are in the range, so we do not subtract any 5s. In the second example, we need to subtract 5 three times in order for the result to be in the target interval.

Non-prime order

In the non-prime case the above method does *not* work! If you need, for example, the field $\mathbb{F}_4$ of order 4 then you *cannot* simply take $\mathbb{Z}_4$, the integers modulo 4. This is not a field! Instead you can take $\mathbb{F}_4 = \{0, 1, a, b\}$, where addition and multiplication is given by the following tables.

$+$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

$\cdot$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

If you are interested to know why these tables are the correct ones and how to construct all finite fields, take the the course in Abstract Algebra.

1.1.3 Vector Spaces

The following is the main definition of the course.

Definition 1.1.4 *Suppose $\mathbb{F}$ is a field. A* **vector space** *over $\mathbb{F}$ is a set V together with two operations: $+ : V \times V \to V$ and $\cdot : \mathbb{F} \times V \to V$, satisfying the following paxioms:*

(VS1) (Associativity of addition) For all $u, v, w \in V$, $u+(v+w) = (u+v)+w$.

(VS2) (Commutativity of addition) For all $u, v \in V$, $u + v = v + u$.

(VS3) (Existence of zero vector) There exists $0 \in V$ *such that* $v+0 = 0+v+0$ *for all.*

(VS4) (Existence of additive inverses) For each $u \in V$, *there exists* $-u \in V$ *such that* $u + (-u) = 0 = -u + u$.

(VS5) (Associativity of multiplication) For all $a, b \in \mathbb{F}$ *and* $u \in V$, $a(bu) = (ab)u$.

(VS6) (Distributivity) For all $a, b \in \mathbb{F}$ *and* $u, v \in V$, $a(u + v) = au + av$ *and* $(a + b)v = av + bv$.

(VS7) (Identity multiplication) For all $u \in V$, $1v = v$.

We refer to the elements of V as *vectors*, as opposed to *scalars*, which are the numbers, that is, the elements of $\mathbb{F}$. Note that the product of a scalar and a vector is always to be written in this order: first the scalar and then the vector; never the other way around.

Also, you may recognize (those who took MSM1B) that the axioms VS1-VS4 mean that $(V, +)$ is an abelian (commutative) group. Thus, a vector space is an abelian group where we also can multiply with scalars.

1.1.4 Basic properties

Here are some basic properties of vector spaces.

Theorem 1.1.5 (Elementary properties) *Suppose* V *is a vector space over a field of scalars* $\mathbb{F}$. *Then the following hold:*

(1) (Cancellation in sums) For $u, v, w \in V$, *if* $u + v = u + w$ *then* $v = w$.

(2) For all $u \in V$, $0u = 0$.

(3) For all $u \in V$, $(-1)u = -u$.

(4) Also, for all $a \in \mathbb{F}$, $a0 = 0$.

Proof. If $u+v = u+w$ then $-u+(u+v) = -u+(u+w)$. Also, $-u+(u+v) = (-u + u) + v = 0 + v = v$ and, similarly, $-u + (u + w) = w$. Hence, $v = w$.

Now, $u + 0 = u = 1u = (1 + 0)u = 1u + 0u = u + 0u$. So $u + 0 = u + 0u$. By cancellation, we now get $0 = 0u$.

The third and fourth claims are left as exercises. □

By commutativity we can also do this: if $u + v = w + u$ then $v = u$; and also this: if $u + v = w + u$ then $v = w$, etc.

Theorem 1.1.6 *Suppose V is a vector space over $\mathbb{F}$. Then:*

(1) For $a \in \mathbb{F}$ and $u \in V$, if $au = 0$ then either $a = 0$ or $u = 0$.

(2) (Cancellation in products)

(a) For $0 \neq a \in \mathbb{F}$ and $u, v \in V$, if $au = av$ then $u = v$.

(b) Also, for $a, b \in \mathbb{F}$ and $0 \neq u \in V$, if $au = bu$ then $a = b$.

Proof. In a field, every nonzero number a has an inverse a^{-1}, which is defined via its property that $aa^{-1} = a^{-1}a = 1$. Hence, assuming that $au = 0$ and that $a \neq 0$, we can write: $a^{-1}(au) = a^{-1}(0)$. On the left we get: $a^{-1}(au) = (a^{-1}a)u = 1u = u$. The right side, on the other hand is equal to 0 by the previous theorem, and so $u = 0$.

For the cancellation property, suppose that $au = av$ and $a \neq 0$. Multiplying both sides with a^{-1} on left, we get $a^{-1}(au) = a^{-1}(av)$. By associativity, $(a^{-1}a)u = (a^{-1}a)v$, and so $1u = 1v$ by the definition of a^{-1}. Finally, this gives $u = v$.

For part (b), $0 = au + (-(au)) = bu + (-(au)) = bu + (-1)(au) = bu + ((-1)a)u = bu + (-a)u = (b - a)u$. Thus, $0 = (b - a)u$. By the above, either $b - a = 0$ or $u = 0$. However, it is given that $u \neq 0$, hence $b - a = 0$. This implies in turn that $a = b$. □

1.1.5 Examples of vector spaces

Here we list some basic examples of vector spaces. More examples can be found later and in the exercises.

Example 1.1.7 (Row space) *Suppose $\mathbb{F}$ is a field. The* row vector space *over $\mathbb{F}$ consists of all rows of length n (that is, matrices of size $1 \times n$), $v = (a_1, a_2, \ldots, a_n)$, where all entries a_i are taken from our field $\mathbb{F}$. The operations of addition and multiplication are performed* component-wise*: Suppose $u = (a_1, a_2, \ldots, a_n)$ and $v = (b_1, b_2, \ldots, b_n)$ are two vectors and suppose $c \in \mathbb{F}$ is a scalar. Then*

$$u + v = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

and

$$cu = (ca_1, ca_2, \ldots, ca_n).$$

We will write $\mathbb{F}^n$ for this vector space. For example, if $\mathbb{F} = \mathbb{R}$, the reals, we get the usual notation $\mathbb{R}^n$.

Example 1.1.8 (Column space) *There is also the* column vector space *over $\mathbb{F}$, where vectors are columns of height n (i.e., matrices of size $n \times 1$) with numbers from $\mathbb{F}$ as entries. The operations are again done componentwise:*

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}$$

and

$$c \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix}$$

We will write ${}^n\mathbb{F}$ for this vector space.

Example 1.1.9 (Function vector spaces) *In modern Analysis one constantly meets vector spaces where vectors are functions. We will just give some basic examples.*

Let X be a subset of R, for example, $X = [0, 1]$ or $X = \mathbb{R}$. The set $C(X)$ of all continuous real-valued functions on X is a vector space over $\mathbb{R}$. The operations are the usual sum of functions and usual multiplication of functions with numbers: If, say, $f(x) = \sin(x)$ and $g(x) = x^2$ then $h := f + g$ is given by the formula $h(x) = \sin(x) + x^2$ and $k := 2f$ is given by the formula $k(x) = 2\sin(x)$.

Similarly, $L^k(X)$ is the vector space of k times continuously differentiable functions on X and $L^\infty(X)$ is the vector space of infinitely differentiable functions on X.

Note that these examples depend on the fact that the sum of two continuous (respectively, differentiable) functions is again continuous (differentiable). If this were not true then we would not have the operation of addition! Similarly, for scalar multiples.

Example 1.1.10 (Space of polynomials) *By P^n we will denote the vector space over $\mathbb{R}$ of all polynomial functions of degree at most n. For example, x^3 is a vector of P^n for all $n \geq 3$. Let P^∞ be the vector space of all polynomial functions, of unlimited degree.*

Finally, let us have a more fancy example of a vector space.

Example 1.1.11 (Subsets of a set) *The set 2^Ω of all subsets of a set Ω is a vector space over $\mathbb{F}_2$. The addition is provided by the operation of* symmetric difference *(often denoted by $\triangle$). Recall, that for two subsets A and B of Ω, the symmetric difference $A \triangle B$ of A and B consists of all elements that are in A, but not in B, and all elements that are in B, but not in A. So*

$$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B).$$

How do we define multiplication with scalars?

1.2 Subspaces, intersection and sum of subspaces

Subspaces, subspace criterion, intersection of subspaces, subspace spanned by a set of vectors, sum of subspaces

1.2.1 Subspaces

Some sets of vectors in a vector space are special.

Definition 1.2.1 *Suppose V is a vector space over $\mathbb{F}$. A subset $U \subseteq V$ is a subspace if it is a vector space over $\mathbb{F}$ in its own right with respect to addition and multiplication inherited from V.*

Let us see some examples.

Example 1.2.2 (Obvious subspaces) *In every vector space V, the subset $U = V$ is obviously a subspace. We say that a subspace $U \subseteq V$ is* proper *if $U \neq V$, that is, U is strictly smaller than V.*

Hence $U = V$ is the only subspace of V that is not proper.

Similarly, in every vector space V, the subset $\{0\}$ is a subspace. This subspace is called the trivial *subspace. We will often write 0 for the trivial subspace.*

Example 1.2.3 ($\mathbb{R}^2$ and $\mathbb{R}^3$) *In the plane $\mathbb{R}^2$, the subspaces are: (1) the trivial subspace containing only the origin 0; (2) the straight lines through the origin; and (3) the improper subspace—the entire plane.*

Similarly, in the 3D space $\mathbb{R}^3$, we have the following subspaces: (1) the trivial subspace; (2) the straight lines through the origin; (3) the planes through the origin; and (4) the improper subspace—the entire space $\mathbb{R}^3$.

Example 1.2.4 (Polynomial subspaces) *The space of polynomials, P^n, contains all polynomial functions up to degree n. If $k \leq n$ then this includes all polynomials of degree at most k. Hence P^k is a subspace of P^n. Note that P^k is a proper subspace of P^n if and only if $k < n$.*

Similarly, all P^n are subspaces of P^∞.

We will see more examples of subspaces later.

1.2.2 Subspace criterion

Suppose V is a vector space over $\mathbb{F}$ and suppose $U \subseteq V$. How can we verify whether U is a subspace? According to the definition, U has to be a vector space in its own right. What does this mean?

In order to be a vector space over $\mathbb{F}$, U must first of all carry two operations: addition of vectors and multiplication of a vector by a scalar. On the one hand, these must be mappings $+ : U \times U \to U$ and $\cdot : \mathbb{F} \times U \to U$. On the other hand, we require above that these operations must be the same ones used in V. This amounts to asking that U be closed with respect to addition and with respect to multiplication with scalars.

More precisely, the following properties must hold:

Definition 1.2.5 *We say that $U \subseteq V$ is* closed for addition *if, for al $u, v \in U$, $u + v \in U$.*

Definition 1.2.6 *A subset $U \subseteq V$ is* closed for multiplication with scalars *if, for all $u \in U$ and all $a \in \mathbb{F}$, $au \in U$.*

In fact these two necessary properties are nearly sufficient.

Theorem 1.2.7 (Subspace criterion) *Suppose V is vector space over $\mathbb{F}$. A* nonempty *subset $U \subseteq V$ is a subspace if and only if U is closed for addition and multiplication with scalars.* □

Note that $U = \emptyset$ contains no elements at all and so, in particular, it contains no elements for which the two closedness properties may fail. Hence the two properties are *trivially* true for $U = \emptyset$—for the lack of counterexamples. However, the empty subset in U is not a subspace, because (VS3) fails!! There is no zero vector in this U!

This why we have to exclude the empty set in the above theorem.

The two conditions in the theorem can be replaced by a single condition.

Theorem 1.2.8 (Subspace criterion, II) *A non-empty subset U of a vector space V over $\mathbb{F}$ is a subspace of V if and only if, for all $u, v \in U$ and all $a \in \mathbb{F}$, we have that $au + v \in U$.*

Proof. Suppose U is a subspace. Then U is closed for multiplication with scalars and so $w := au$ must be in U, since $u \in U$ and $a \in \mathbb{F}$. Also, U is closed for addition. Hence $au + v = w + v \in U$, since $w, v \in U$. Therefore the above condition holds.

Conversely, suppose the condition holds. It is given to us that U is nonempty. We will verify the two closedness conditions from Theorem 1.2.7.

First, suppose $u, v \in U$. Taking $a = 1 \in \mathbb{F}$ gives that $u + v = 1u + v \in U$. Thus, U is closed for addition. Next, suppose $u \in U$ and $a \in \mathbb{F}$. Then also $a - 1 \in \mathbb{F}$ and so, by the condition, $au = ((a-1)+1)u = (a-1)u + 1u = (a-1)u + u$ is in U, proving that U is closed for multiplication with scalars. Thus U is a subspace by Theorem 1.2.7. □

Example 1.2.9 *In $\mathbb{R}^3$, the set of vectors $U = \{(a, b, 0) \mid a, b \in \mathbb{R}\}$ is a subspace. Indeed, this set is clearly nonempty. Also, taking arbitrary $u, v \in U$ and arbitrary $a \in \mathbb{R}$, we have that $u = (b, c, 0)$ and $v = (d, e, 0)$ for some $b, c, d, e \in \mathbb{R}$. Hence $au + v = a(b, c, 0) + (d, e, 0) = (ab + d, ac + e, 0)$ is in U, since it has the required shape!*

On the other hand, the subset $U = \{(a, b, 1) \mid a, b \in \mathbb{R}\}$ in $\mathbb{R}^3$ is not a subspace.

Why? The zero vector is missing! Also, not closed for addition! Also, not closed for scalar multiplication. (Check both!)

Both of these sets are planes in $\mathbb{R}^3$. However, the first one passes through the origin, while the second one does not.

Just one further example before we develop more theory.

Example 1.2.10 (Lines) *Suppose u is a vector in a vector space V over $\mathbb{F}$. Then the set of all* multiples *of u, that is, $U = \{au \mid a \in \mathbb{F}\}$, is a subspace of V. Clearly, U is nonempty as it contains both $0 = 0u$ and $u = 1u$. It is also easy to see U is closed for addition and multiplication with scalars.*

This subspace can be viewed as the straight line in V passing through the origin 0 in the direction of (or parallel to) u.

In a sense, the subspace in this example is generated by the vector u, that is, one can recover it from just u. This observation can be made much more general.

1.2.3 Intersection of subspaces

There are two operations, intersection and sum, that start from known subspaces and produce new subspaces. We will begin by discussing intersection.

Theorem 1.2.11 (Intersection of subspaces) *If $\{U_i \mid i \in I\}$ is a collection of subspaces of a vector space V then the intersection $W := \cap_{i \in I} U_i$ is again a subspace of V.*

Proof. Every U_i contains the zero vector, so W also contains the zero vector. This means that $0 \in W$ and so W is nonempty.

Let $u, v \in W$ and $a \in \mathbb{F}$ ($\mathbb{F}$ being the field of scalars of V). Since W is the intersection of all U_i's, we have that u and v lie in each U_i. Since U_i is a subspace, $au + v$ lies in each U_i. Hence $au + v$ lies also in the intersection W, proving that W is a subspace by invoking Theorem 1.2.8. □

For example, if U_1, U_2 and U_3 are subspaces of V then $U_1 \cap U_2$ and $U_1 \cap U_2 \cap U_3$ are subspaces of V, too.

1.2.4 Span

The result that the intersection of subspaces is again a subspace has an interesting corollary.

Theorem 1.2.12 (Minimal subspace) *Suppose V is a vector space and $X \subseteq V$ is an arbitrary subset. Then there is a unique minimal subspace U of V containing X. This U is contained in any other subspace of V containing X.* □

The second claim in this theorem simply explains in which sense U is minimal.

The subspace can be defined as $\cap_{i \in I} U_i$, where $\{U_i\}_{i \in I}$ is the family of *all* subspaces of V containing X. All properties follow at once.

The above theorem allows us to give the following definition.

Definition 1.2.13 (Span) *For a vector space V and a subset $x \subseteq V$, the subspace* generated *(or spanned) by X is the unique minimal subspace of V containing X. This subspace is denoted by $\langle X \rangle$. It is also known as the* span *of X.*

If $X = A \cup B \cup C$, say, we may write $\langle A, B, C \rangle$ instead of $\langle A \cup B \cup C \rangle$. Also, we will write u in place $\{u\}$. Hence, say, $\langle A, u, B, v \rangle = \langle A \cup \{u\} \cup B \cup \{v\} \rangle$, where $A, B \subseteq V$ and $u, v \in V$.

Let us now return to Example 1.2.10.

Example 1.2.14 (Line, II) *If the set X consists of a single vector, u, then every subspace containing u also contains all multiples of u. On the other hand, we already saw that $U = \{au \mid a \in \mathbb{F}\}$ is a subspace. Hence this is the minimal subspace containing x. That is, $\langle u \rangle = \{au \mid a \in \mathbb{F}\}$.*

For example, if $V = P^3$ and $u = x + 1$ then $\langle u \rangle$ contains polynomials 0, $x + 1$, $\frac{1}{2}x + \frac{1}{2}$, $-3x - 3$, and in fact, $\langle u \rangle$ consists exactly of all polynomials of the form $ax + a$ for all $a \in \mathbb{R}$.

Example 1.2.15 *Similarly, if* $X = \{u, v\}$ *then* $\langle u, v\rangle = \{au + bv \mid a, b \in \mathbb{F}\}$.

For example, if $V = 2^{\Omega}$, *where* $\Omega = \{red, blue, yellow, green\}$, *and* $u = \{red, yellow\}$ *and* $v = \{blue, yellow, green\}$ *then* $\langle u, v\rangle$ *consists of the four vectors* $0u + 0v = \emptyset$, $0u + 1v = v = \{blue, yellow, green\}$, $1u + 0v = u = \{red, yellow\}$, *and* $1u + 1v = u + v = \{red, blue, green\}$.

This examples bring us close to the idea of linear combination of vectors, which allows us, among other things, to write explicitly all vectors in the subspace spanned by a set of vectors.

However, before we explore this, we insert another related topic.

1.2.5 Sum of subspaces, direct sum

We now discuss the second operation on subspaces, the sum.

Definition 1.2.16 (Sum of subspaces) *Suppose* U *and* W *are two subspaces of a vector space* V. *Then we define* $U + W = \{u + w \mid u \in U, w \in W\}$.

Theorem 1.2.17 *If* U *and* W *are subspaces of a vector space* V *then* $U + W$ *is also a subspace of* V. □

We leave the proof as an exercise.

The above definition and theorem can be generalized for an arbitrary number of subspaces. Namely, if $U_1, U_2, \ldots, U_k$ are subspaces of V then we define

$$U_1 + \ldots + U_k := \{u_1 + \ldots + u_k \mid u_1 \in U_1, \ldots, u_2 \in U_2\}.$$

This subset, $U_1 + \ldots + U_k$ is always a subspace of V.

To connect with the preceding subsection, about generation, let us state the following result.

Theorem 1.2.18 *If* U *and* W *are subspaces of a vector space* V *then* $U + W = \langle U, W\rangle$. *More generally, if* $U_1, \ldots, U_k$ *are subspaces of* V *then* $U_1 + \ldots + U_k = \langle U_1, \ldots, U_k\rangle$. □

Suppose U and W are subspaces of V and let $D = U + W$. In general, every vector $d \in D$ can be written as a sum $d = u + w$, for $u \in U$ and $w \in W$, in more than one way (that is, there exist different pairs u, w adding up to the same d). So the case where the *components* u and w of d are unique is special.

Definition 1.2.19 (Direct sum) *Suppose U and W are subspaces of a vector space V. We say that $D = U + W$ is the* direct sum *of U and W (denoted $D = U \oplus W$) if for every $d \in D$ there exists only one choice of $u \in U$ and $w \in W$ such that $d = u + w$.*

We state the following result.

Theorem 1.2.20 (Direct sum criterion) *For subspace U and W of a vector space V, we have that $D = U + W$ is the direct sum of U and W if and only if $U \cap W = 0$.*

Proof. First suppose that $U \cap W \neq 0$. Let $0 \neq v \in U \cap W$. Then $v \in U$ and $-v \in W$. Taking now $d = 0 \in D$ we see that $d = 0 = 0 + 0$ (i.e., we can take $u = 0$ and $w = 0$) and $d = 0 = v + (-v)$ (i.e., we can also take $u = v$ and $w = -v$). Hence the choice of u and w for d is not unique and so the sum is not direct.

Conversely, suppose that $U \cap W = 0$. We need to show that $D = U \oplus W$. By contradiction, suppose that for some $d \in D$ we have $d = u + w = u' + w'$, where $u, u' \in U$, $w, w' \in W$ and either $u \neq u'$ or $w \neq w'$. Since $u + w = u' + w'$, we obtain $u - u' = w' - w$. Setting $v := u - u' = w' - w$, we note that $v \in U \cap W$. Since $U \cap W = 0$, we conclude that $v = 0$, that is, $u - u' = 0$, yielding $u = u'$, and $w' - w = 0$, yielding $w = w'$. This is a contradiction. □

All of this can be generalized to any number of summands.

Definition 1.2.21 (Direct sum, II) *Suppose $U_1, \ldots, U_k$ are subspaces of a vector space V. We say that $D = U_1 + \ldots + U_k$ is the* direct sum *of $U_1, \ldots, U_k$ (and write $D = U_1 \oplus \ldots \oplus U_k$) if every $d \in D$ can be uniquely written as $d = u_1 + \ldots + u_k$ for $u_1 \in U_1, \ldots, u_k \in U_k$.*

Here is how this property can be checked.

Theorem 1.2.22 (Direct sum criterion, II) *For subspaces $U_1, \ldots, U_k$ of a vector space V, we have that $D = U_1 + \ldots + U_k$ is the direct sum of the subspaces U_i if and only if $(U_1 + \ldots + U_{i-1}) \cap U_i = 0$ for $i = 2, 3, \ldots, k$.* □

For example, if we have three subspaces, U, W, and T, then to claim that $D = U + W + T$ is the direct sum it is not enough to verify that $U \cap W = 0 = U \cap T = W \cap T$. However, it suffices to check that $U \cap W = 0$ and $(U + W) \cap T = 0$.

1.3 Linear combinations and independence

Vectors in the span, linear combinations of vectors, minimal spanning sets, independent sets of vectors.

1.3.1 Vectors in the span

Suppose V is a vector space over a field $\mathbb{F}$ and let $X \subseteq V$.

Definition 1.3.1 (Linear combination) *An expression of the form*

$$\sum_{x \in X} a_x x$$

is called a linear combination of the vectors from X. Here for each vector $x \in X$, a_x is a scalar from $\mathbb{F}$. The numbers a_x are called the coefficients of the linear combination. If X is an infinite *set, we additionally assume that only* finitely *many coefficients a_x are non-zero, so that a linear combination is always a finite sum and so it can be evaluated.*

In the case where $X = \{x_1, x_2, \ldots, x_k\}$ is a finite set a linear combination of X looks simply as an expression of the form

$$a_1 x_1 + a_2 x_2 + \ldots + a_k x_k$$

for arbitrary coefficients $a_1, a_2, \ldots, a_k \in \mathbb{F}$.

Definition 1.3.2 *We also say, by a slight abuse of terminology, that a vector $v \in V$ is a linear combination of X if*

$$v = \sum_{x \in X} a_x x$$

for some coefficients a_x. That is, if v is the result of evaluation of some linear combination of X.

Note that, in general, there may be more than one linear combination evaluating to v.

Example 1.3.3 (Trivial linear combination) *The trivial linear combination of X is the one where all coefficients $a_x = 0$. Clearly, the trivial linear combination sums up to the zero vector. Hence the zero vector is a linear combination of every set X of vectors, including the empty set X!*

Example 1.3.4 (Vector from X) *Each vector $y \in X$ is a linear combination of X. Indeed, if we choose $a_x = 1$, if $x = y$, and $a_x = 0$, otherwise, then the resulting linear combination evaluates to y.*

Theorem 1.3.5 (The span) *Suppose V is a vector space over $\mathbb{F}$ and suppose $X \subseteq V$. Then $\langle X \rangle$ consists of all vectors that are linear combinations of X. That is,*

$$\langle X \rangle = \{\sum_{x \in X} a_x x \mid a_x \in \mathbb{F} \text{ for all } x \in X\}.$$

Proof. We first show that the set U of vectors that are linear combinations of X is a subspace.

Since the zero vector is a linear combination of X (see Example 1.3.3), U is non-empty. Suppose $u, v \in U$ and $d \in \mathbb{F}$. Since $u \in U$, we have that $u = \sum_{x \in X} a_x x$ for some coefficients a_x with only finitely many of them non-zero. Similarly, $v = \sum_{x \in X} b_x x$, where only finitely many coefficients b_x are non-zero.

Setting $c_x = da_x + b_x$, we see that only finitely many of these new coefficients are non-zero. Hence $\sum_{x \in X} c_x x$ is a *bona fide* linear combination of X.

On the other hand, $du + v = d(\sum_{x \in X} a_x x) + \sum_{x \in X} b_x x = \sum_{x \in X} c_x x$. Since we found a linear combination that evaluates to $du + v$, we conclude that $du + v \in U$ and so U is a subspace of V by the Subspace Criterion.

By Example 1.3.4 above, U contains every vector from X, that is, $X \subseteq U$. In particular, this means that $\langle X \rangle \subseteq U$.

On the other hand, since $\langle X \rangle$ is a subspace, it is closed for addition and multiplication with scalars. Also, $\langle X \rangle$ by definition contains X. Hence $\langle X \rangle$ contains all vectors that are linear combinations of X. Therefore, $U \subseteq \langle X \rangle$, proving the equality $\langle X \rangle = U$. □

Because of this theorem, if we want to check whether a given vector is contained in the subspace $W = \langle X \rangle$, we just need to decide whether w is a linear combination of the vectors u_i from X.

Example 1.3.6 *In $\mathbb{R}^3$, is $u = (1, 2, 3)$ contained in $W = \langle u_1, u_2 \rangle$, where $u_1 = (1, 1, 1)$ and $u_2 = (-1, 0, 1)$? To answer this we need to see whether u is a linear combination of u_1 and u_2. Setting the equality $u = a_1 u_1 + a_2 u_2$, where the unknowns a_1 and a_2 represent the coefficients of the linear combination that we are trying to find, we obtain:*

$$(1, 2, 3) = a_1(1, 1, 1) + a_2(-1, 0, 1) = (a_1 - a_2, a_1, a_1 + a_2).$$

This gives us three linear equations (one for each coordinate): $a_1 - a_2 = 1$, $a_1 = 2$, *and* $a_1 + a_2 = 3$. *The equations can be solved by the general methods as in MSM1A. In this particular case, the system of three equations has a solution* $a_1 = 2$ *and* $a_2 = 1$.

Hence $u = 2u_1 + u_2$, *meaning that* u *is a linear combination of* u_1 *and* u_2, *and so* $u \in W = \langle u_1, u_2 \rangle$.

We will now look at the example where we try to list all vectors in the subspace spanned by a set of vectors.

Example 1.3.7 *Suppose* $V = \mathbb{F}_5^3$ *and* $X = \{(1,1,1), (4,2,1), (0,3,2)\}$. *Then* $\langle X \rangle$ *consists of all vectors* $a(1,1,1) + b(4,2,1) + c(0,3,2)$, *where* $a, b, c \in \mathbb{F}_5$. *This gives us* $5^3 = 125$ *linear combinations to compute, which is probably not something we wont to do. However, if we indeed started doing the sums, we would soon discovered that many linear combinations produce the same vectors, and that the final count of vectors in* $\langle X \rangle$ *is much smaller that* 125. *In fact,* $\langle X \rangle$ *consists of only* 25 *vectors.*

1.3.2 Independent sets

Now that we know how to find all vectors in the subspace $U = \langle X \rangle$, we will take a different point of view. Note that *every* subspace U of a vector space V coincides with $\langle X \rangle$ for a suitable set of vectors X. In fact, for a given U there are zillions of different sets X generating U. So we can ask: Are there better and worse generating (spanning) sets oe are all generating sets about the same?

The answer is that definitely some generating sets are better than the others. In a sense, the smaller the generating set for U, the better this set is. Consider again Example 1.3.7. The given generating set of three vectors is not too good, as we get a lot of repetitions when we compute linear combinations. This generating set is excessive and we can get a better, more economical generating set by removing one of the three vectors.

Say, the set $X' = \{(1,1,1), (4,2,1)\}$ spans the same subspace U. However, this time different linear combinations eveluate to different vectors! We have $5^2 = 25$ different ways to select the coefficients for the two vectors in X', and this leads to 25 different vectors in U.

Definition 1.3.8 (Linear dependence) *Suppose* X *is a set of vectors of a vector space* V. *We say that* X *is linearly dependent if a non-trivial linear combination of* X *evaluates to the zero vector. We will call such a linear combination a non-trivial linear dependence or a non-trivial linear relation on* X.

We say that X is linear independent if X is not linearly dependent.

Example 1.3.9 *The set $X = \{(1,1,1),(4,2,1),(0,3,2)\}$ in $V = \mathbb{F}_5^3$ is linearly dependent. Indeed, $1(1,1,1)+1(4,2,1)+4(0,3,2) = (0,0,0)$. This dependence is the reason why this set produces the same vectors multiple times. Indeed, $(0,0,0) = 0(1,1,1)+0(4,2,1)+0(0,3,2) = 1(1,1,1)+1(4,2,1)+4(0,3,2) = 2(1,1,1)+2(4,2,1)+3(0,3,2) = 3(1,1,1)+3(4,2,1)+2(0,3,2) = 4(1,1,1)+4(4,2,1)+1(0,3,2)$. Similarly, every non-zero vector in U appears five times as a linear combination.*

On the contrary, the set $X' = \{(1,1,1),(4,2,1)\}$ is linearly independent and hence no repetitions arise.

To summarize this example, let us prove the following result.

Theorem 1.3.10 (Unique coefficients) *Suppose X is a set of vectors in a vector space V over $\mathbb{F}$. Let $U = \langle X \rangle$ and $u \in U$. Then, for u, there exists a unique choice of coefficients $a_x \in \mathbb{F}$ such that $u = \sum_{x\in X} a_x x$ if and only if X is independent.*

Proof. We will prove contrapositive of this statement: The coefficients for u are not unique if and only if X is linearly dependent.

First suppose X is dependent. Then $0 = \sum_{x\in X} a_x x$ for some coefficients a_x not all equal to zero. Since $u \in U$, we can also write $u = \sum_{x\in X} b_x x$ for some $b_x \in \mathbb{F}$. Note that $u = 0+u = \sum_{x\in X} a_x x + \sum_{x\in X} b_x x = \sum_{x\in X}(a_x+b_x)x$. Note that these new coefficients $c_x = a_x + b_x$ are different from the b_x, since some a_x are nonzero. Hence the choice of coefficients for u is not unique.

Conversely, suppose that $u = \sum_{x\in X} b_x x = \sum_{x\in X} c_x x$ for some coefficients $b_x, c_x \in \mathbb{F}$, where $b_x \neq c_x$ for at least one $x \in X$. Then $0 = u - u = \sum_{x\in X} b_x x - \sum_{x\in X} c_x x = \sum_{x\in X}(b_x - c_x)x$. Setting $a_x = b_x - c_x$ for all x, we now see that not all a_x are zero while $0 = \sum_{x\in X} a_x x$. Hence X is linearly dependent. □

Example 1.3.11 *The vectors $(1,1,1)$, $(4,2,1)$, and $(0,3,2)$ were shown to be dependent in $\mathbb{R}^3$.*

On the other hand, the polynomials $x^2 - x$, $x^2 + x$ and $x - 2$ are linearly independent in P^2. Indeed, if we try to write the zero polynomial as a linear combination: $0 = y_1(x^2 - x) + y_2(x^2 + x) + y_3(x - 2)$, then this leads to three equations: $y_1 + y_2 = 0$ (for x^2), $-y_1 + y_2 + y_3 = 0$ (for x) and $-2y_3 = 0$ (the constant term). Solving this system we readily find that $y_3 = 0$ and then also that $y_1 = y_2 = 0$. This means that only the zero linear combination of these three polynomials produces the zero polynomial, and so they are linearly independent.

1.3.3 Up and down

Can we also find an independent generating set? Yes, we simply need to remove unnecessary vectors from a known generating set, as we already did in Example 1.3.9.

Theorem 1.3.12 (Unnecessary vector) *Suppose a set X of vectors of a vector space V is linearly dependent, namely,*

$$\sum_{x \in X} a_x x = 0,$$

where not all a_x are zero. Let $x \in X$ be such that $a_x \neq 0$. Then $X' = X \setminus \{x\}$ spans the same subspace $U = \langle X \rangle$.

Proof. Since $X' \subseteq X \subseteq U$, we have that $\langle X' \rangle \leq U$. So we just need to show that every vector from U is contained in $\langle X' \rangle$. Take $u \in U$. Then $u = \sum_{y \in X} b_y y$ for some coefficients b_y. Let $d = b_x / a_x$. (We can divide since $a_x \neq 0$.)

Note that $u = u - d0 = \sum_{y \in X} b_y y - d \sum_{y \in X} a_y y = \sum_{y \in X} (b_y - d a_y) y$. Setting $c_y = b_y - d a_y$, we see that $u = \sum_{y \in X} c_y y$. Also, $c_x = b_x - d a_x = 0$, which means that x is not involved in this linear combination, and so u is in fact a linear combination of the set X'. Therefore $u \in \langle X' \rangle$, proving that $U = \langle X' \rangle$. □

Example 1.3.13 *Suppose X consists of four vectors: $X = \{u_1, u_2, u_3, u_4\}$ which satisfy a linear dependency $2u_1 - u_3 + 5u_4 = 0$. Hence the set X is linearly dependent. Due to the relation we have, we can claim that the subsets $\{u_2, u_3, u_4\}$ (removing u_1), $\{u_1, u_2, u_4\}$ (removing u_3), and $\{u_1, u_2, u_3\}$ (removing u_4) are generating for $U = \langle X \rangle$.*

We cannot remove u_2 without losing the generation property, unless we find another non-trivial dependency involving u_2.

Note that we can remove any one vector involved in a non-trivial relation, but we cannot remove more than one, not without finding more relations.

Does it mean that one of the smaller sets is necessary independent? No, although the above relation is not valid for this smaller set, there may exist further non-trivial relations, and so we may need to continue the process of eliminating unnecessary vectors.

Just like we can remove unnecessary vectors from generating sets, we can sometimes add new vectors to independent sets.

Theorem 1.3.14 (Adding a vector) *Suppose X is a linearly independent set in a vector space V and $u \in V$. If u is a not a linear combination of the vectors from X (that is, $u \notin \langle X \rangle$) then $X' = X \cup \{u\}$ is linearly independent.*

Proof. By contradiction, suppose we have a non-trivial linear dependency

$$\sum_{x \in X'} a_x x$$

on X'. If $a_u = 0$ then this linear dependency is in fact a non-trivial linear dependency on X, which is a contradiction since X is linearly independent. Therefore, $a_u \neq 0$. This allows us to solve the above equation for u as follows:

$$u = \sum_{x \in X} (-\frac{a_x}{a_u})x,$$

which shows that u is in fact a linear combination of X. The contradiction proves the claim. □

Example 1.3.15 *Suppose $V = {}^3\mathbb{R}$. The vectors $u_1 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$ and $u_2 = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix}$ are linearly independent. Take $u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. If we try to solve $u = au_1 + bu_2$ for a and b we soon arrive at a contradiction, which means that u is independent of u_1 and u_2. Therefore, $\{u_1, u_2, u\}$ is a linearly independent set.*

Can we extend this set further? In this case, no! The reason is that $\{u_1, u_2, u\}$ already spans the whole $V = {}^3\mathbb{R}$.

1.4 Bases

Bases, standard bases, all bases have the same size, dimension, bases and direct sums, dimension formula.

1.4.1 Bases

Definition 1.4.1 (Basis) *A basis in a vector space V is a linearly independent generating set.*

Theorem 1.4.2 (Existence of bases) *Every vector space has a basis.*

We will show two ways to prove this.

Proposition 1.4.3 (Down to a basis) *Every generating set X in a vector space V contains a basis.*

Proof. If X is linearly independent then it is a basis, and there is nothing to prove. Otherwise, X is linearly dependent and so there is a non-trivial linear relation. It allows us to remove one element from X, as in Theorem 1.3.12, leaving a smaller set that is still generating. If X is finite then iterating this process, we will surely end up with a generating set for which no non-trivial linear dependence exists, and this set is a basis.

If the set X is infinite the same elimination process can be made rigorous using transfinite induction. □

Every vector space V has a generating set X of vectors. For example, we can take all (!) vectors, that is, $X = V$. Hence the above proposition implies Theorem 1.4.2

There is also a second way to prove the theorem.

Proposition 1.4.4 (Up to a basis) *Every linearly independent set X in a vector space V is contained in a basis.*

Proof. If X spans V then X is itself a basis. Otherwise, $\langle X \rangle < V$ and so we can find $u \in V$ such that $u \notin \langle X \rangle$. By Theorem 1.3.14, adding u to X produces a larger independent set that span a larger subspace of V. We now repeat this process for the new set X and do it as many times as necessary. Clearly, the process stops only if X already spans the whole V and then X is a basis.

Again, the completely rigorous way of proving this involves transfinite induction. □

Every vector space contains independent sets of vectors, for example, the empty set! Hence this proposition also implies Theorem 1.4.2.

1.4.2 Standard bases

Some of the vector spaces we consider have a particular basis that is easy to find and that is useful in many applications. Such bases are called standard (also, natural or canonical).

Example 1.4.5 *In the row space $V = \mathbb{F}^n$, where $\mathbb{F}$ is a field, the set $B = \{e_1, \ldots, e_n\}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the ith position, is a basis of V. We will give here a complete proof of this statement.*

To see that B spans all of V we need to show that every vector $u \in V$ is a linear combination of B. Consider $u = (a_1, \ldots, a_n)$. Then $u = a_1u_1 + \ldots + a_ne_n$ and so indeed u is a linear combination of B.

Secondly, we need to check that B is linearly independent. Suppose that $a_1e_1 + \ldots + a_ne_n = 0$ for some coefficients $a_i \in \mathbb{F}$. Note that $a_1e_1 + \ldots + a_ne_n = (a_1, \ldots, a_n)$ and so we get $(a_1, \ldots, a_n) = 0 = (0, \ldots, 0)$. Clearly this means that all a_i are zero. Hence there are no non-trivial dependencies on B. Thus, B is a basis.

This basis B it the standard basis of $\mathbb{F}^n$.

Example 1.4.6 *Similarly, in the column space ${}^n\mathbb{F}$ we choose the vectors*

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, f_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, f_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

to form its standard basis. The proof that this is a basis is similar to the above.

The space of polynomials also have a standard basis.

Example 1.4.7 *In the vector space P^n of all polynomial functions of degree up to n, the monomial functions $p_i = x^i$, $0 \leq i \leq n$, form the standard basis.*

When we write an arbitrary polynomial f as $f = a_nx^n + \ldots + a_1x + a_0$ we implicitly acknowledge the fact that f is a linear combination of the monomials p_i: $f = a_np_n + \ldots + a_1p_1 + a_0p_0$. So the monomials span P^n. The fact that the p_i are linearly independent is left as an exercise.

The vector space P^∞ of polynomial functions of unbounded degree has the set of all monomials, of all degrees, as its standard basis. This is an example of an infinite basis.

Here is a slightly trickier example of a standard basis.

Example 1.4.8 *Let Ω be a set and let $V = 2^\Omega$ be the vector space of subsets of Ω (with symmetric difference as addition). In the case where Ω is finite, say $\Omega = \{red, blue, yellow, green\}$ the one-element sets $\{red\}$, $\{blue\}$, $\{yellow\}$, and $\{green\}$ form the standard basis of V.*

If Ω is infinite then V has no standard basis! In fact it is difficult or impossible to write down any explicit basis at all!

1.4.3 Dimension

Definition 1.4.9 (Finite-dimensional spaces) *A vector space is called finite-dimensional if it contains a finite generating (spanning) set of vectors. Otherwise, we call the vector space infinite dimensional.*

Proposition 1.4.10 *Suppose V is a vector space, X is an independent set of vectors in V, and Y is a generating set of vectors, that is, $\langle Y \rangle = V$. Then $|X| \leq |Y|$.*

Proof. We only deal with the case where Y is finite, say, of size m. In this case, V is finite-dimensional. By contradiction, suppose $|X| > |Y|$. Since every subset of an independent set is independent, we can assume that X is finite, of size $n > m$.

Let $X = \{x_1, \ldots, x_n\}$. Since X is linearly independent, the equation $\sum_{i=1}^n a_i x_i = 0$ has only one solution: $a_1 = 0 = a_2 = \ldots = a_n$.

On the other hand, since Y is generating, every x_i is a linear combination of $Y = \{y_1, \ldots, y_m\}$. So we can write

$$x_i = c_{i1}y_1 + c_{i2}y_2 + \ldots + c_{im}y_m$$

for each i and for some coefficients c_{ij}.

Substituting these into the above equation, we get $\sum_{i=1}^n a_i(\sum_{j=1}^m c_{ij}y_j) = 0$, which can be rearranged to give: $\sum_{j=1}^m(\sum_{i=1}^n c_{ij}a_i)y_j = 0$, that is,

$$(\sum_{i=1}^n c_{i1}a_i)y_1 + \ldots + (\sum_{i=1}^n c_{im}a_i)y_m = 0.$$

If each of the m coefficient sums is equal to zero then the above is a true equality. This gives us m homogeneous equations in n variables $a_1, \ldots, a_n$:

$$\sum_{i=1}^n c_{ij}a_i = 0$$

for $j = 1, \ldots, m$. Since the number of variables, n, is greater than the number of equations, m, the system of equations has a non-zero solution.

This, however, contradicts the given fact that X is an independent set of vectors. $\square$

As a consequence of this, we state the following result.

Theorem 1.4.11 (Basis size) *Any two bases in a vector space have the same size.*

Proof. Suppose B and B' are two bases. Then B is independent while B' is a generating set. Hence by Proposition 1.4.10, $|B| \leq |B'|$. Symmetrically, B' is independent while B is generating. Hence by Proposition 1.4.10, $|B'| \leq |B|$. Clearly, this means that $|B| = |B'|$. $\square$

Hence all bases in a vector space have the same size.

Definition 1.4.12 (Dimension) *The dimension of a vector space V is the size of an arbitrary basis of V.*

We will write $\dim V$ for the dimension of V. Clearly, a finite-dimensional vector spaces has a finite non-negative integer as its dimension, while for an infinite-dimensional vector space V we just write $\dim V = \infty$.

Example 1.4.13 *If a vector has a standard basis (see the preceding subsection) then it's very easy to tell the dimension. For example, the dimension of the row space $\mathbb{F}^n$ is n, because its standard basis consists of n vectors.*

Similarly, P^n has dimension $n+1$, because its standard basis consists of $n+1$ vectors $1, x, \ldots, x^n$. The space P^∞ of all polynomial functions is infinitely-dimensional.

We conclude this section with the following observation.

Theorem 1.4.14 *Suppose V is a vector space and $U \leq V$ is a subspace. Then $\dim U \leq \dim V$. In particular, if V is finite-dimensional then so also is U. Furthermore, if both U and V are finite-dimensional and $\dim U = \dim V$ then $U = V$.*

Proof. Let B be a basis of U. Since B is linearly independent, by Proposition 1.4.4, B is contained in a basis B' of V. Clearly, $|B| \leq |B'$ and so $\dim U \leq \dim V$. The second claim is clear.

For the final claim, if B and B' are finite and have the same size then clearly $B = B'$ (since $B \subseteq B'$). Therefore, $U = \langle B \rangle = \langle B' \rangle = V$. $\square$

Example 1.4.15 *In the plane $\mathbb{R}^2$ we have the trivial subspace 0 (containing only the zero vector, the origin), which has dimension 0. We also have 1-dimensional subspaces, which are the straight lines through the origin. By the above theorem, the only 2-dimensional subspace in $\mathbb{R}^2$ is $\mathbb{R}^2$ itself, and this completes the list of all subspaces of $\mathbb{R}^2$!*

As an exercise, try listing all subspaces in the 3-dimensional space $\mathbb{R}^3$.

1.4.4 Dimension of the sum of subspaces

Here we discuss an important and very useful formula.

Theorem 1.4.16 (Dimension of sum) *Suppose V is a vector space and suppose U and W are finite-dimensional subspaces of V. Then* $\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$.

Proof. Let $T = U \cap W$. Pick a basis B in T, then extend B to a basis X of U, and also extend B to a basis Y of W.

Note, first of all, that $X \cap T$ contains B, and it is an independent set in T, hence its size cannot be larger than that of B. It follows that $X \cap T = B$. Similarly, $Y \cap T = B$. This implies in particular that $X \cap Y = B$, since $X \cap Y \subseteq U \cap W = T$.

Since $X \cap Y = B$, we have that $|X \cup Y| = |X| + |Y| - |B|$, which translates into the dimension formula in the theorem, once we show that $X \cup Y$ is a basis in $U + W$.

Every vector in $U + W$ looks like $u + w$, where $u \in U$ and $w \in W$. Since u is a linear combination of X and w is a linear combination of Y, we conclude that $u + w$ is a linear combination of $X \cup Y$, which shows that the latter is a spanning set in $U + W$.

It remains to show that $X \cup Y$ is linearly independent. Consider a linear combination of $X \cup Y$ that sums up to the zero vector:

$$\sum_{z \in X \cup Y} a_z z = 0.$$

Let $R = Y \setminus B$ and set $t = \sum_{z \in R} a_z z$. Clearly, $t \in W$ since $R \subseteq Y$. On the other hand, $t = -\sum_{z \in X} a_z z$, and so t lies in U. Therefore, $t \in U \cap W = T$. In particular, we can write

$$t = \sum_{z \in B} b_z z$$

for some coefficients b_z.

Finally, $0 = t - t = \sum_{z \in B} b_z z + \sum_{z \in R} (-a_z) z$, which means that all coefficients here (the b_z and well as the $-a_z$) are zero, as $Y = B \cup R$ is linearly independent. In particular, $a_z = 0$ for all $z \in R$. This means that our initial linear combination is just a linear combination of X. However, X is also independent, and so all a_z are zero. Hence $X \cup Y$ is linearly independent, as claimed. □

Example 1.4.17 (Nontrivial intersection) *Suppose V has dimension 3 while its two subspaces U and W both have dimension 2. This gives us*

enough information to claim that U and W intersect nontrivially, that is, they have a non-zero vector in common.

Indeed, let k be the dimension of $U + W$. Then clearly $k \leq 3$, because $U + W$ is a subspace of V. By the above formula, $k = 2 + 2 - \dim(U \cap W)$. Hence $\dim(U \cap W) = 4 - k \geq 1$. So $U \cap W$ cannot be trivial.

Example 1.4.18 *Similarly, let $\dim V = 7$, $\dim U = 4$ and $\dim W = 5$. What are the possible dimensions of $U \cap W$? As in the preceding example, we get that $\dim(U \cap W) \geq 4+5-7 = 2$. So $U \cap W$ is at least 2-dimensional. On the other hand, $U \cap W$ is contained in both U and W, and so $\dim(U \cap W) \leq \min(\dim U, \dim W) = \min(4, 5) = 4$. Thus, $2 \leq \dim U \leq 4$. Note that all these dimensions are in fact possible.*

Also note that $\dim(U \cap W) = 2$ if and only if $\dim(U + W) = 7$, that is, if and only if $U + W = V$. Also, $\dim(U \cap W) = 4$ if and only if $U \cap W = U$, that is, $U \leq W$.

Theorem 1.4.16 has the following implications.

Theorem 1.4.19 *Suppose U and W are finite-dimensional subspaces of a vector space V. Let $S = U + W$. Then $S = U \oplus W$ if and only if $\dim S = \dim U + \dim W$.* □

Another way to state the same is as follows.

Theorem 1.4.20 *Suppose U and W are finite-dimensional subspaces of a vector space V. Let $S = U + W$. Let X be a basis in U and Y be a basis in W. Then $S = U \oplus W$ if and only if $X \cap Y = \emptyset$ and $X \cup Y$ is linearly independent (in which case it is automatically a basis of S).* □

1.5 Coordinates

Coordinates, coordinate vector, standard coordinates, use of coordinates: echelon form and determinant, coordinate mapping.

1.5.1 Coordinates

Definition 1.5.1 (Coordinates) *Suppose V is a vector space over a field $\mathbb{F}$ and suppose B is a basis of V. For a vector $u \in V$, the coordinates of u with respect to B (aka B-coordinates) are the coefficients c_b, for $b \in B$, of the linear combination of B that sums up to u. That is, $u = \sum_{b \in B} c_b b$.*

Why is this a valid definition? First of all, B is a spanning set for V and so certainly a linear combination summing up to u exists. Also, B is linearly independent and so, by Theorem 1.3.10, the coefficients c_b as above are in fact unique. Hence indeed B-coordinates of u are well defined.

In most cases we will be considering finite dimensional vector spaces, and so B is a finite set in such a case. However, for the sake of generality, let us mention that B happens to be infinite then almost all coordinates (all but finitely many, to be precise!) are zero. This is our usual proviso for the coefficients of a linear combination—so that the summation could actually be performed.

Let now B be finite and let us order B in some way, $B = \{b_1, \ldots, b_n\}$. Hence $n = \dim V$.

1.5.2 Coordinate vector

Definition 1.5.2 (Coordinate vector) *For $u \in V$, its B-coordinate (column) vector is $[u]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$, where $c_1, c_2, \ldots, c_n$ are the corresponding B-coordinates of u.*

In simple terms, we took the coordinates of u and we arranged them in a column in the same order in which the corresponding basis vectors appear in B.

Note that the coordinate vector u_B is not a vector of U in most cases! It is a vector in the column space ${}^n\mathbb{F}$.

Example 1.5.3 (Standard coordinates) *Let us see some examples for spaces admitting a standard basis, and let us choose this standard basis as our B. In this case, we will talk about the standard coordinates of the vector.*

(a) For example, suppose $V = \mathbb{R}^3$ and $u = (1, 2, 3)$. Recall that the standard basis of $\mathbb{R}^3$ is $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and let us assume that this is our B and that it is ordered in exactly this way. Then $(1, 2, 3) = 1(1, 0, 0) + 2(0, 1, 0) + 3(0, 0, 1)$ and so the B coordinates are 1, 2 and 3. $u_B = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Now suppose that take a basis B' that is the same as B as a set, but where we use a different (reverse) order of vectors. Can you see in which way $[u]_{B'}$ is different from $[u]_B$?

(b) Suppose $V = P^3$, the space of polynomial functions of degree up to 3. Recall that its standard basis is $\{1, x, x^2, x^3\}$, and we take exactly this order. Suppose $u = 2x^2 + 13$. Then $u = 13 \cdot 1 + 0x + 2x^2 + 0x^3$, and so its standard coordinates are 13, 0, 2, and 0. Hence $[u]_B = \begin{pmatrix} 13 \\ 0 \\ 2 \\ 0 \end{pmatrix}$

There are no xs in the coordinate vector!

(c) Suppose $V = 2^{\{red,blue,yellow,green\}}$. We defined the standard basis of this V as $B = \{\{red\}, \{blue\}, \{yellow\}, \{green\}\}$, in this order. Suppose that $u = \{green, red, yellow\}$. Then $u = 1\{red\}+0\{blue\}+1\{yellow\}+1\{green\}$, and so the coordinates are 1, 0, 1, and 1, and so the standard coordinate vector u_B is $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$.

Recall that in this last example $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$ and the addition and multiplication operations are "modulo 2".

Let us now see an example with a non-standard basis.

Example 1.5.4 (Non-standard basis) *Suppose $V = P^2$, but instead of the standard basis $\{1, x, x^2\}$ let us take $B = \{1 + x, x^2 + 1, (x - 1)(x + 2)\}$. Which vector u satisfies* $u_B = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$*? The B-coordinates of this vector u are -1, -1, and 2 and so $u = -1(1+x)-1(x^2+1)+2(x-1)(x+2) = x^2+x-6$.*

This shows that for a non-standard basis B, given just the vector $u = x^2 + x - 6$, we would hardly be able to guess the B-coordinates directly from u, but rather we would have to solve a system of linear equations to find the coordinates.

1.5.3 Using coordinates

Suppose V is an n-dimensional vector space over $\mathbb{F}$. Then any question about vectors of V can be reformulated as a similar question about column vectors from ${}^n\mathbb{F}$.

Theorem 1.5.5 (Re-writing) *Suppose V is a vector space with a basis B and let $u_1, \ldots, u_k, u \in V$. Then*

(1) for $a_1, \ldots, a_k \in \mathbb{F}$, $u = \sum_{i=1}^{k} a_i u_i$ if and only if $[u]_B = \sum_{i=1}^{k} a_i [u_i]_B$; and therefore,

(2) $u_1, \ldots, u_k$ span V if and only if $[u_1]_B, \ldots, [u_k]_B$ span ${}^n\mathbb{F}$;

(3) $u_1, \ldots, u_k$ are linearly independent in V if and only if $[u_1]_B, \ldots, [u_k]_B$ are linearly independent in ${}^n\mathbb{F}$; and

(4) $u_1, \ldots, u_k$ form a basis of V if and only if $([u_1]_B, \ldots, [u_k]_B$ form a basis of ${}^n\mathbb{F}$. □

Let us illustrate this principle.

Example 1.5.6 *Suppose $V = P^2$ and we need to decide whether $u = x^2 - x - 1$ lies in $\langle x^2, 2x - x + 1 \rangle$. By part (1) of the above theorem, the answer is yes if and only if $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ lies in $\langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \rangle$. (Here we used the standard basis for re-writing.)*

So the theory allows us to deal with familiar column vectors instead of less familiar function vectors.

1.5.4 Coordinate mapping

Note that for a fixed (ordered) basis B of V we obtain a mapping $\phi : V \to {}^n\mathbb{F}$, sending each $u \in V$ to its B-coordinate column vector $[u]_B$. We call this mapping the coordinate mapping with respect to B.

The reason why the above principle works lie in the properties of this mapping. What are those important properties?

Theorem 1.5.7 (Coordinate mapping) *Suppose V is an n-dimensional vector space over $\mathbb{F}$, B an ordered basis of V, and ϕ is the corresponding coordinate mapping. Then*

(1) $\phi(u + v) = \phi(u) + \phi(v)$ for all $u, v \in V$;

(2) $\phi(au) = a\phi(u)$ for all $u \in V$ and all $a \in \mathbb{F}$;

(3) ϕ is a bijection from V onto ${}^n\mathbb{F}$.

This leads us directly to the topic of Chapter 2.

Chapter 2

Linear transformations

2.1 Linear mappings

Linear mappings, elementary properties of linear mappings, kernel and injectivity, rank plus nullity, surjectivity, isomorphisms.

2.1.1 Linear mappings

Definition 2.1.1 (Linear mapping) *Suppose U and V are vector spaces over the same field of scalars $\mathbb{F}$. A mapping $\phi : U \to V$ is linear if the following two properties hold:*

(L1) for all $u, v \in U$, we have $\phi(u + v) = \phi(u) + \phi(v)$;

(L2) for all $u \in U$ and $a \in \mathbb{F}$, we have $\phi(au) = a\phi(u)$.

Just like with the subgroup test, there is an easier, one-condition check of linearity.

Theorem 2.1.2 (Linearity check) *A mapping $\phi : U \to V$ is linear if and only if for all $u, v \in U$ and all $a \in \mathbb{F}$, we have $\phi(au + v) = a\phi(u) + \phi(v)$.*

Proof. Clearly, (L1) and (L2) imply the condition in this theorem. So it suffices to check our new condition implies both (L1) and (L2).

So suppose that for all $u, v \in U$ and all $a \in \mathbb{F}$, we have $\phi(au + v) = a\phi(u)+\phi(v)$. Taking $a = 1$, we see that $\phi(u+v) = \phi(1u+v) = 1\phi(u)+\phi(v) = \phi(u) + \phi(v)$, and so (L1) holds for ϕ. (Here we twice used one of the axioms of vector spaces. Which one?)

Now we show that (L2) holds too: $\phi(au) = \phi((a - 1 + 1)u) = \phi((a - 1)u + 1u) = \phi((a - 1)u + u)$. By our assumption, taking $v = u$ and the

scalar $a-1$, we obtain: $\phi(au) = (a-1)\phi(u) + \phi(u) = (a-1)\phi(u) + 1\phi(u) = (a-1+1)\phi(u) = a\phi(u)$. So (L2) holds, as claimed. □

Example 2.1.3 (Linear function) *In Calculus, we call a function $f : \mathbb{R} \to \mathbb{R}$ linear if f is given by the formula $y = ax + b$. Is there a relation between these functions and the linear mappings that we introduced?*

There is a relation, but only partial. First of all, a function $f : \mathbb{R} \to \mathbb{R}$ can be understood as a mapping between two vector spaces since $\mathbb{R}$ can be identified with $\mathbb{R}^1 = {}^1\mathbb{R}$. In this sense, our question now becomes more precise: is the mapping given by $x \mapsto y = ax + b$ a linear mapping?

The answer is: this mapping is linear if and only if $b = 0$. We first check that f is linear if $b = 0$. Indeed, take two real numbers-vectors, u and v, and a real number-scalar, say c (since a is already taken). Then $f(cu + v) = a(cu + v) = c(au) + (av) = cf(u) + f(v)$. So by Theorem 2.1.2 we have that f is a linear mapping.

Note that in this example both vectors and scalars are real numbers, so we need to be clear which real numbers are viewed as vectors and which are viewed as scalars.

Since $f(0) = a0 + b = b$, if $b \neq 0$ then f is not linear by part (1) of the theorem following this example.

Theorem 2.1.4 (Properties of linear mappings) *Suppose $\phi : U \to V$ is a linear mapping. Then*

(1) $\phi(0) = 0$; and

(2) for all $u \in U$, $\phi(-u) = -\phi(u)$.

Proof. Take an arbitrary $u \in U$. Then $0u = 0$ by Theorem 1.1.5(2). Therefore, by (L2), $\phi(0) = \phi(0u) = 0\phi(u) = 0$, giving us the claim (1).

Similarly, by Theorem 1.1.5(3), $-u = (-1)u$. So $\phi(-u) = \phi((-1)u) = (-1)\phi(u) = -\phi(u)$. Hence (2) holds as well. □

Let us also record the following important property of linear maps. (It can also be viewed as an exercise in applying Theorem 2.1.2 within a proof.)

Theorem 2.1.5 (Composition) *Suppose that $\phi : U \to V$ and $\psi : V \to W$ are linear mappings. Then the composition map $\psi \circ \phi : U \to W$ is also linear.*

Proof. Let $\mathbb{F}$ be the field of scalars of U, V, and W. Take arbitrary $u, v \in U$ and $a \in \mathbb{F}$. Since ϕ is linear, Theorem 2.1.2 implies that $\phi(au + v) = a\phi(u) + \phi(v)$. Note that $u' := \phi(u) \in V$ and, similarly, $v' := \phi(v) \in V$. The

same Theorem 2.1.2 implies now, since ψ is also linear, that $\psi(au' + v') = a\psi(u') + \psi(v')$.

Putting these two statements together, $\psi \circ \phi(au + v) = \psi(\phi(au + v)) = \psi(a\phi(u) + \phi(v)) = \psi(au' + v') = a\psi(u') + \psi(v') = a\psi(\phi(u)) + \psi(\phi(v)) = a\psi \circ \phi(u) + \psi \circ \phi(v)$.

We have shown that $\psi \circ \phi(au + v) = a\psi \circ \phi(u) + \psi \circ \phi(v)$ for all $u, v \in U$ and $a \in \mathbb{F}$. Thus, by Theorem 2.1.2, $\psi \circ \phi$ is linear, as claimed. □

Note that we could also use the definition of linear maps in this proof, instead of Theorem 2.1.2, and the proof would only be slightly longer in this case.

Here are some examples of linear mappings.

Example 2.1.6 (Trivial and identity mappings) *(a) For arbitrary U and V defined over the same field of scalars $\mathbb{F}$, we can define $\phi : U \to V$ via $\phi(u) = 0$ for all $u \in U$. Since $\phi(au+v) = 0 = a0+0 = a\phi(u)+\phi(v)$, this mapping is linear. We call this the trivial linear mapping from U to V, and we write $\phi = 0$.*

(b) If $V = U$, we can define $\phi : U \to U$ via $\phi(u) = u$ for each $u \in U$. This is also linear. Indeed, $\phi(au + v) = au + v = a\phi(u) + \phi(v)$. This linear mapping is called the identity mapping, and we write $\phi = id$.

Here are some further examples of linear mappings.

Example 2.1.7 (Transposing) *Suppose $U = \mathbb{F}^n$, the row space, and $V = {}^n\mathbb{F}$, the column space. Then transposing a row vector $u = (a_1, \ldots, a_n) \in U$ gives us a column vector $u^T = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$. This operation is a linear mapping from U to V. Obviously, we can also use transposing to define a mapping from V to U, which is also linear and which is the inverse of the above.*

More generally, all matrices of a particular shape $n \times m$ and with entries from $\mathbb{F}$ form a vector space $M_{n\times m}(\mathbb{F})$, where addition of matrices and multiplication of a matrix by a scalar from $\mathbb{F}$ are done in the usual way, that is, entry-wise.

The row and column spaces are just particular cases of this general construction. Namely, $\mathbb{F}^n = M_{1\times n}(\mathbb{F})$ and ${}^n\mathbb{F} = M_{n\times 1}(\mathbb{F})$. Transposing defines mutually inverse linear mappings between $U = M_{n\times m}(\mathbb{F})$ and $V = M_{m\times n}(\mathbb{F})$ for all n and m.

The following example generalizes our linear function example.

Example 2.1.8 (Multiplication with a matrix) *Suppose $U = {}^m\mathbb{F}$ and $V = {}^n\mathbb{F}$ are column spaces and let $A \in M_{n\times m}(\mathbb{F})$ be an arbitrary $n\times m$ matrix. Let $\phi : U \to V$ be defined by $\phi(u) = Au$. Then this is a linear mapping, as follows immediately from the familiar properties of matrix multiplication. Indeed, $\phi(au + v) = A(au + v) = A(au) + Av = a(Au) + Av = a\phi(u) + \phi(v)$.*

Similarly, a matrix $A \in M_{n\times m}(\mathbb{F})$ defines a linear mapping between the row spaces $U = \mathbb{F}^n$ and $V = \mathbb{F}^m$. However, this time the matrix should be on the right: $\psi(u) = uA$.

This example will play an important role in the course. Also, the following example is important to us.

Example 2.1.9 (Differentiation) *Suppose X is an open subdomain in $\mathbb{R}$ (such as, say, an open interval (a, b)) and let $U = L^1(X)$ and $V = C(X)$ be the vector spaces of continuously differentiable and continuous functions on X,respectively. Then differentiation $D : f(x) \mapsto f'(x) = \frac{d}{dx}f(x)$ is a linear mapping. Indeed, the familiar rules of differentiation give: $D(af(x)+g(x)) = (af(x) + g(x))' = af'(x) + g'(x) = aD(f(x)) + D(g(x))$.*

This example is central in Analysis.

Example 2.1.10 (Integration) *Naturally, since integration is un-does differentiation, it also defines a linear mapping. This time the source space is $V = C(X)$ and the target space is $U = L^1(X)$.*

The indefinite indegral (antiderivative) is not good for defining a linear mapping, as it outputs not a single function but rather a family of functions differing by a constant. Hence we will have to use a definite integral.

For $f(x) \in C(\bar{X})$, define $I(f(x)) = \int_0^x f(t)dt$. This outputs a continuously differential function on X, hence a vector of U. This mapping is linear, as one immediately gets from the rules of integration.

Finally, here we have a combinatorial example.

Example 2.1.11 (Intersection is linear) *Suppose $U = 2^\Omega$ and $V = 2^\Delta$ where Δ is a subset of Ω. Define a mapping $\phi : U \to V$ via $\phi(A) = A \cap \Delta$ for each vector-subset A. This is a linear mapping.*

2.1.2 Kernel and injectivity

Definition 2.1.12 (Kernel) *For a linear transformation $\phi : U \to V$, its kernel* $\ker \phi$ *is the following subset of the source space U:*

$$\ker \phi = \{u \in U \mid \phi(u) = 0\}.$$

Example 2.1.13 *For the trivial linear mapping $\phi = 0 : U \to V$ the kernel* $\ker \phi$ *coincides with the entire U, that is,* $\ker \phi = U$*. For the identity mapping $id : U \to U$,* $\ker id = \{0\} = 0$*.*

Note that in both cases $\ker \phi$ is a subspace of U. This in fact is true in general.

Theorem 2.1.14 (Kernel is a subspace) *Suppose $\phi : U \to V$ is a linear mapping. Then* $\ker \phi$ *is a subspace of U.*

Proof. Since $\phi(0) = 0$ by Theorem 2.1.4(1), we see that $\ker \phi$ is nonempty. By the subspace criterion, it remains to see that $au + v \in \ker \phi$ for all $u, v \in \ker \phi$ and all $a \in \mathbb{F}$.

Note that $\phi(au + v) = a\phi(u) + \phi(v) = a0 + 0$, since $u, v \in \ker \phi$ and so $\phi(u) = \phi(v) = 0$. It follows that $\phi(au + v) = 0$, and hence $au + v \in \ker \phi$. □

Let us see more examples of kernels.

Example 2.1.15 *Consider differentiation mapping D as a mapping from $U = P^n$ to itself (and so $V = U = P^n$. Then* $\ker D = \{f \in P^n \mid f' = 0\}$ *coincides with the subspace of constant polynomials. Each such polynomial is a multiple of the constant-1 function, so* $\ker D = \langle 1 \rangle$ *is 1-dimensional.*

How about the combinatorial example, where the linear function comes from the intersection?

Example 2.1.16 *Suppose $U = 2^\Omega$, $V = 2^\Delta$, for some $\Delta \subseteq \Omega$, and let ϕ be defined by $\phi(A) = A \cap \Delta$. We already claimed that ϕ is linear. What is its kernel? Clearly, $\phi(A) = 0 = \emptyset$ if and only if $A \cap \Delta = \emptyset$, that is, A is contained in the complement $\Sigma = \Omega \setminus \Delta$. It follows that* $\ker \phi = 2^\Sigma$ *and its dimension is $|\Sigma|$.*

The kernel is important to us because of the following fact.

Theorem 2.1.17 (Injectivity) *Suppose $\phi : U \to V$ is a linear mapping. Then ϕ is injective if and only if* $\ker \phi = 0$*.*

Proof. First suppose that ϕ is injective. Then there is only one vector, 0, of U that maps to 0. Therefore, $\ker \phi = \{0\} = 0$.

Conversely, suppose $\ker \phi = 0$. If $u, v \in U$ map to the same vector of V, that is, if $\phi(u) = \phi(v)$, then $\phi(u - v) = \phi(u) - \phi(v) = 0$, which means that $u - v \in \ker \phi$. Since $\ker \phi = 0$, we conclude that $u - v = 0$, that is, $u = v$. It follows that ϕ is injective. □

Example 2.1.18 *The trivial mapping $0 : U \to V$ is injective only if $U = 0$ is 0-dimensional. The identity mapping $id : U \to U$ is always injective.*

Example 2.1.19 *The differentiation mapping $D : P^n \to P^n$ is never injective, as $\dim(\ker D) = 1$ for all b. The intersection mapping $\phi : 2^\Omega \to 2^\Delta$ is injective if and only if $\Delta = \Omega$ (in which case ϕ is the identity mapping, id).*

2.1.3 Image, rank+nullity, surjectivity

Definition 2.1.20 *Suppose $\phi : U \to V$ is a linear mapping. The image of ϕ is the set*

$$\mathrm{im}\phi = \{v \in V \mid v = \phi(u) \text{ for some } u \in U\} = \{\phi(u) \mid u \in U\}.$$

This follows the standard terminology whereby, whenever we have $v = \phi(u)$, we say that v is the image of u and u is the preimage of v. Thus, $\mathrm{im}\phi$ is the set of all images of individual elements of U.

Theorem 2.1.21 *For a linear mapping $\phi : U \to V$, $\mathrm{im}\phi$ is a subspace of V.*

Proof. Clearly, $\phi(0) = 0$ means that $\mathrm{im}\phi$ is non-empty. Suppose $v, v' \in \mathrm{im}\phi$ and $a \in \mathbb{F}$. Since $v, v' \in \mathrm{im}\phi$, $v = \phi(u)$ and $v' = \phi(u')$ for some $u, u' \in U$. Therefore, $\phi(au + u') = a\phi(u) + \phi(u') = av + v'$, and so $av + v' \in \mathrm{im}\phi$. Hence, $\mathrm{im}\phi$ is a subspace by the subspace criterion. □

We now focus on how to determine whether ϕ is surjective, that is, whether $\mathrm{im}\phi = V$.

Theorem 2.1.22 *Suppose $\phi : U \to V$ is a linear mapping and suppose $u_1, \ldots, u_k$ is a spanning set of vectors of U. Then setting $v_1 = \phi(u_1), \ldots, v_k = \phi(u_i)$, we have that $v_1, \ldots, v_k$ span $\mathrm{im}\phi$.* □

We also have the following.

Theorem 2.1.23 *Suppose $\phi : U \to V$ is a linear mapping and suppose $v_1, \ldots, v_k$ is an independent set of vectors of V. Let $u_1, \ldots, u_k \in U$ be such that $\phi(u_i) = v_i$ for all i. Then $u_1, \ldots, u_k$ are independent in U.* □

Both these theorems are left as exercise.

The following is a partial reverse of the latter theorem.

Theorem 2.1.24 *Suppose $u_1, \ldots, u_k$ are linearly independent vectors of U. If $\langle u_1, \ldots, u_k \rangle \cap \ker \phi = 0$ then $v_1 = \phi(u_1), \ldots, v_k = \phi(u_k)$ are linearly independent in V.*

Proof. Suppose $\sum_{i=1}^{k} a_i v_i = 0$. Note that $\phi(\sum_{i=1}^{k} a_i u_i) = \sum_{i=1}^{k} a_i \phi(u_i) = \sum_{i=1}^{k} = 0$. Hence $w = \sum_{i=1}^{k} a_i u_i \in \ker \phi$. Clearly, $w \in \langle u_1, \ldots, u_k \rangle$. Since the latter meets $\ker \phi$ trivially, we have that $w = 0$, that is, $\sum_{i=1}^{k} a_i u_1 = 0$. Since $u_1, \ldots, u_k$ are linearly independent, we now get that all a_i are equal zero. □

This theorem together with Theorem 2.1.22 imply the main result of the current subsection.

Theorem 2.1.25 (Rank+nullity) *Suppose $\phi : U \to V$ is a linear mapping and suppose that $\dim U < \infty$. Then $\dim \mathrm{im}\phi + \dim \ker \phi = \dim U$.*

Proof. Let $w_1, \ldots, w_s$ be a basis of $\ker \phi$. Hence $\dim \ker \phi = s$. Extend this independent set to a basis $w_1, \ldots, w_s, u_1, \ldots, u_r$ of U. So $\dim U = s + r$. The claim follows if we establish that $v_1 = \phi(u_1), \ldots, v_r = \phi(u_r)$ form a basis of $\mathrm{im}\phi$.

We have already seen that $\langle u_1, \ldots, u_s \rangle \cap \langle w_1, \ldots, w_r \rangle = 0$. (This was in the homework.) Since $\langle u_1, \ldots, u_s \rangle = \ker \phi$, we conclude that $\langle w_1, \ldots, w_r \rangle \cap \ker \phi = 0$. By the preceding theorem, we get that $v_1, \ldots, v_r$ are linearly independent.

It remains to show that they span V. However, this is clear. Since $w_1, \ldots, w_s, u_1, \ldots, u_r$ span U, their images span $\mathrm{im}\phi$ by Theorem 2.1.22. Since all w_i map to the zero vector, their images do not contribute to spanning, that is, $v_1, \ldots, v_r$ span V. This shows that indeed $v_1, \ldots, v_r$ form a basis of $\mathrm{im}\phi$ and so $\dim \mathrm{im}\phi = r$. □

2.1.4 Matrix of a linear mapping

We start by considering the case where $U = \mathbb{F}^m$ and $V = {}^n\mathbb{F}$ are column spaces.

Theorem 2.1.26 *Suppose $\phi : {}^m\mathbb{F} \to {}^n\mathbb{F}$ is a linear mapping. Then there exists a unique $n \times m$ matrix $A \in M_{n \times m}(\mathbb{F})$ such that $\phi(u) = Au$ for all $u \in {}^m\mathbb{F}$.*

Proof. We start with the uniqueness of A. That is, suppose such an A exists. If we can find a way to compute A then it is certainly unique.

Consider the ith vector $u_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ from the standard basis of $U = {}^m\mathbb{F}$.

(Here the column is of height m and the only nonzero entry 1 is in the ith position in the column). Note that Au_i coincides with the ith column of A (check it!) On the other hand, Au_i must be equal to $\phi(u_i)$. Hence, for each i, the ith column of A must coincide with $\phi(u_i)$. This tells us exactly what A must look like and so A is unique.

The above also gives us an idea for the existence proof. Namely, let us set A to be the matrix whose ith column v_i coincides with $\phi(u_i)$ (u_i as above). We just need to verify that this A works, that is, it satisfies the claim from our theorem.

Consider an arbitrary vector $u = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$ of $U = {}^m\mathbb{F}$. We need to show that $Au = \phi(u)$.

Note that $u = c_1u_1 + \ldots c_mu_m$. On the other hand, from the way we multiply matrices we deduce that $Au = c_1v_1 + \ldots + c_mv_m$. By linearity of ϕ, we get $\phi(u) = \phi(c_1u_1 + \ldots + c_mu_m) = c_1\phi(u_1) + \ldots + c_m\phi(u_m) = c_1v_1 + \ldots + c_mv_m = Au$, as claimed. □

Let us stress that this proof gives us a method for computing A. Namely, this matrix consists of the columns $v_i = \phi(u_i)$, where $\{u_1, \ldots, u_m\}$ is the standard basis of $U = {}^m\mathbb{F}$.

We now turn to the case of an arbitrary linear mapping. Suppose U and V are vector spaces over $\mathbb{F}$ of finite dimensions m and n respectively. Suppose we have some bases $B = \{u_1, \ldots, u_m\}$ and $B' = \{v_1, \ldots, v_n\}$ chosen in U and V. Recall that for every $u \in U$ we denote by $[u]_B$ the B-coordinate column vector of U, that is, the column made of coordinates of u with respect to the basis B. Similarly, $[v]_{B'}$ is the B'-coordinate vector of $v \in V$.

Definition 2.1.27 (Matrix of a linear mapping) *Under the setup above, suppose $\phi : U \to V$ is a linear mapping. A matrix $A \in M_{n\times m}(\mathbb{F})$ is the* matrix of the linear mapping ϕ with respect to bases B and B' *if $[\phi(u)]_{B'} = A[u]_B$ for all $u \in U$.*

Note that $\phi(u)_{B'}$ is a column of height n and u_B is a column of height m. Hence A must indeed be of size $n \times m$ for the equation $\phi(u)_{B'} = Au_B$ to make sense.

We pair up the above definition with the following result.

Theorem 2.1.28 *Suppose $\phi : U \to V$ is a linear mapping, where U and V are vector spaces over $\mathbb{F}$ with bases B and B' respectively. Then the matrix of ϕ with respect to B and B' exists and is unique.*

Proof. Note that the mapping $u_B \mapsto \phi(u)_{B'}$ is a linear mapping from ${}^m\mathbb{F}$ to ${}^n\mathbb{F}$. (We leave this claim as an exercise.) By Theorem 2.1.26, there exists a unique matrix $A \in M_{n,m}(\mathbb{F})$ such that $[\phi(u)]_{B'} = A[u]_B$ for all $u \in U$. Clearly, this A is the matrix of ϕ with respect to B and B'. □

How can we compute this A? The method described after Theorem 2.1.26 suggests that A has to do with images of the column vectors from the standard basis of ${}^m\mathbb{F}$. Note that the ith vector from this standard basis is simply $(u_i)_B$, the B-coordinate vector of the ith vector u_i from B. Hence the ith column of A coincides with $[\phi(u_i)]_{B'}$, the B'-coordinate vector of $\phi(u_i)$.

Let us see an example.

Example 2.1.29 *Suppose $U = 2^\Omega$, where $\Omega = \{a, b, c\}$, and $V = 2^{\Omega'}$, where $\Omega' = \{a, c, d, e\}$. Suppose that the linear mapping $\phi : U \to V$ is given by $\phi(A) = A \cap \Omega'$. Let us compute the matrix A of ϕ with respect to the standard bases $B = \{\{a\}, \{b\}, \{c\}\}$ and $B' = \{\{a\}, \{c\}, \{d\}, \{e\}\}$ of U and V.*

First of all, $\dim U = 3$ and $\dim V = 4$, hence A is going to be of size 4×3. Secondly, our field in this example is $\mathbb{F}_2$, and so all entries of A are going to be 0 or 1.

According to our recipe, we need to compute the B'-coordinate vectors of the images of the vectors from B.

We compute: $\phi(\{a\}) = \{a\} \cap \Omega' = \{a\}$. Since this is the first vector of B', we get the B'-coordinate vector $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, which must therefore be the first column of A. Similarly, since $\phi(\{b\}) = \{b\} \cap \Omega = \emptyset$, we have that the second column of A is all zero. Finally, since $\phi(\{c\}) = \{c\} \cap \Omega' = \{c\}$, which is the second vector of B', we have that the third column of A is $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. Thus we

conclude that

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

The matrix of the linear mapping can be used for computations.

Example 2.1.30 *Let U, V, ϕ, B and B' be as above. Take $u = \{b, c\} \in U$. The the B-coordinate vector of u is $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Hence*

$$[\phi(u)]_{B'} = A[u]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore $\phi(u) = \{c\}$, the second vector from B'.

Of course, here it was easier to just compute $\phi(u)$ using the definition of ϕ. In applications, however, ϕ might be quite complicated and so the matrix multiplication using the coordinate vectors and the matrix of ϕ might be a handy alternative.

2.2 Isomorphisms

Isomorphisms, inverse mappings, checking for isomorphism, isomorphic vector spaces, coordinate isomorphism.

2.2.1 Isomorphisms and inverse mappings

Definition 2.2.1 (Isomorphism) *A linear mapping $\phi : U \to V$ that is bijective is called an isomorphism.*

In the last homework we discussed, for a general ϕ, that $\phi^{-1}(v)$ is a set, not a single vector. When ϕ is bijective, $\phi^{-1}(v)$ consists of a single vector, that is, $\phi^{-1}(v) = \{u\}$ for some u and so we can define a proper inverse mapping $\phi^{-1} : V \to U$ by setting $\phi^{-1}(v) = u$.

Theorem 2.2.2 (Inverse isomorphism) *If $\phi : U \to V$ is an isomorphism then $\phi^{-1} : V \to U$ is linear, and so it is also an isomorphism.*

Proof. Suppose $v, v' \in V$ and $a \in \mathbb{F}$. We need to show that $\phi^{-1}(av + v') = a\phi^{-1}(v) + \phi^{-1}(v')$. Let $u = \phi^{-1}(v)$ and $u' = \phi^{-1}(v')$, that is, $u, u' \in U$ are such that $\phi(u) = v$ and $\phi(u') = v'$. Then, since ϕ is linear, we have that $\phi(au + u') = a\phi(u) + \phi(u') = av + v'$. Thus, $\phi(au + u') = av + v'$, which means that $\phi^{-1}(av + v') = au + u' = a\phi^{-1}(v) + \phi^{-1}(v')$, as claimed.

Clearly, ϕ^{-1} is also bijective, hence an isomorphism. □

Definition 2.2.3 (Inverse linear mapping) *Given $\phi : U \to V$, we define its* inverse mapping *as a linear mapping $\psi : V \to U$ such that $\phi \circ \psi = id_V$ and $\psi \circ \phi = id_U$, that is, $\phi(\psi(v)) = v$ for all $v \in V$ and $\psi(\phi(u)) = u$ for all $u \in U$.*

If ϕ is an isomorphism then $\psi = \phi^{-1}$ (as defined above) satisfies the above property, that is, an inverse for ϕ exists. In fact, inverses exist only for isomorphisms.

Theorem 2.2.4 (Inverse) *If $\phi : U \to V$ is a linear mapping that has an inverse $\psi : V \to U$ then ϕ is an isomorphism and $\psi = \phi^{-1}$.* □

Thus, the inverse, when it exists, is unique.

Why are we interested in isomorphisms? Because they are the nicest possible linear mappings!

Theorem 2.2.5 *Suppose $\phi : U \to V$ is an isomorphism. Suppose further that X is a subset of U and let $Y = \phi(X)$. Then*

(1) X spans U if and only if Y spans V;

(2) X is independent in U if and only if Y is independent in V; and hence,

(3) X is a basis of U if and only if Y is a basis of V. □

In particular, if W is a subspace of U then $T = \phi(W)$ is a subspace of V of the same dimension as W.

2.2.2 Checking for an isomorphism

Theorem 2.2.6 (Isomorphism criterion) *A linear mapping $\phi : U \to V$, where U is finite-dimensional, is an isomorphism if and only if* $\ker \phi = 0$ *and* $\dim V = \dim U$.

Proof. If ϕ is an isomorphism then it is bijective, and so it is injective and surjective. Since ϕ is injective, $\ker\phi = 0$ by Theorem 2.1.17. Also, since ϕ is surjective, $\mathrm{im}\phi = V$ and so $\dim \mathrm{im}\phi = \dim V$. Hence by the Rank+Nullity Theorem, $\dim U = \dim\ker\phi + \dim\mathrm{im}\phi = 0 + \dim V = \dim V$. Thus, $\dim V = \dim U$.

Conversely, suppose $\ker\phi = 0$ and $\dim V = \dim U$. The first condition gives us that ϕ is injective by Theorem 2.1.17 while the Rank+Nullity Theorem gives us that $\dim\mathrm{im}\phi = \dim V$, since $\dim V = \dim U$. It follows that $\mathrm{im}\phi = V$, that is ϕ is surjective. Thus ϕ is an isomorphism. □

In practical terms, how can we check whether a given linear transformation $\phi : U \to V$ is an isomorphism? Let us pick bases X and Y in U and write the matrix A of ϕ with respect to X and Y. Note, first of all, that ϕ cannot be an isomorphism unless $\dim U = \dim V$. (We assume of course that U and V are finite dimensional—or else there is no matrix!) Thus, the matrix A has to be square.

Theorem 2.2.7 *For U and V of the same finite dimension, a linear mapping $\phi : U \to V$ is an isomorphism if and only $\det A \neq 0$, where A is the matrix of ϕ written with respect to some (arbitrary) bases of U and V.* □

Example 2.2.8 *Suppose $\phi : \mathbb{R}^4 \to \mathbb{R}^4$ is given by $\phi((a,b,c,d)) = (d,a,c,b)$ for all $a,b,c,d \in \mathbb{R}$. In the homework it was checked that this is a linear mapping. Is ϕ an isomorphism? If U and V did not have the same dimension, we would know right away that ϕ could not be an isomorphism. However, $U = V = \mathbb{R}^4$, in particular, U and V have the same dimension, and so ϕ may be an isomorphism. Is it?*

Let us write the matrix of ϕ with respect to the standard basis both in U and in V. Since $\phi((1,0,0,0)) = (0,1,0,0)$, the first column of A is $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. *Similarly, $\phi((0,1,0,0)) = (0,0,0,1)$, $\phi((0,0,1,0)) = (0,0,1,0)$, and $\phi((0,0,0,1)) = (1,0,0,0)$ give us the second, third and fourth columns, and so* $A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. *It is easy to compute that $\det A = 1 \neq 0$, and so ϕ is an isomorphism.*

Another way to show that ϕ is an isomorphism is as follows. Check that $\psi : \mathbb{R}^4 \to \mathbb{R}^4$ given by $\psi((a,b,c,d)) = (b,d,c,a)$ is the inverse mapping to ϕ! By Theorem 2.2.4, this implies, once checked, that ϕ is an isomorphism.

We now want to investigate when, for given U and V, we can find an isomorphism from U to V.

Is differentiation D ever an isomorphism? We are tempted to say no, because we know that the constant functions are in the kernel of D and so D should not be injective. However, what if we make sure that the constant functions are not contained in U. This idea leads to the following example.

Example 2.2.9 *Let $V = P^n$ and let U be the subset of P^{n+1} consisting of all those polynomial functions of degree up to $n+1$, whose constant term is zero. Check that U is a subspace!*

Differentiation D sends all of P^{n+1} to $P^n = V$. In particular, if we restrict D to U, we get a linear mapping $D : U \to V$.

Clearly, $X = \{x, x^2, \ldots, x^{n+1}\}$ is a basis for U. For V, we take as Y the standard basis $\{1, x, \ldots, x^n\}$. With respect to these X and Y, the matrix of D is $\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & n+1 \end{pmatrix}$. The determinant of this matrix is equal to $(n+1)!$, and so it is non-zero. This shows that D is an isomorphism from U to V.

2.2.3 Isomorphic vector spaces

Definition 2.2.10 (Isomorphic vector spaces) *Suppose that U and V are vector spaces over the same field of scalars $\mathbb{F}$. We say that U and V are isomorphic if and only if there exists an isomorphism $\phi : U \to V$.*

Clearly, U and V must have the same dimension if they are isomorphic. We will see here that this condition is also sufficient.

To achieve this, we must be able to construct linear mappings between vector spaces. The following theorem is very useful for this purpose.

Theorem 2.2.11 (Unique linear mapping) *Suppose U and V are vector spaces over $\mathbb{F}$ and suppose $B = \{u_i\}_{i \in I}$ is a basis of U. For each $i \in I$, pick a vector $v_i \in V$. Then there exists a unique linear mapping $\phi : U \to V$ such that $\phi(u_i) = v_i$ for all $i \in I$.* □

Note that $i \neq j$ then $u_i \neq u_j$. However, we allow the possibility that $v_i = v_j$—this is no problem.

We omitted the proof, but let us mention that the unique ϕ as in the theorem is simply the following mapping

$$\phi(u) = \sum_{i \in I} c_i v_i,$$

where c_i are the coordinates of u with respect to the basis B. (And so $u = \sum_{i \in I} c_i u_i$.)

Theorem 2.2.12 (Isomorphic spaces) *Two vector spaces U and V over $\mathbb{F}$ are isomorphic if and only if they have the same dimension.* □

Again we omit the detailed proof, but the idea is simple. Pick a basis $X = \{u_i\}_{i \in I}$ in X and a basis Y in V. Since U and V have the same dimension, X and Y have the same cardinality, which means we have a bijection $\alpha : X \to Y$. For each $i \in I$, select $v_i = \alpha(u_i)$ and apply Theorem 2.2.11. The linear mapping that we get is the desired isomorphism.

Let us state part of this "proof" as a separate theorem, and additional criterion of when a linear mapping is an isomorphism.

Theorem 2.2.13 *Suppose $\phi : U \to V$ is a linear mapping. Then ϕ is an isomorphism if and only if it maps a basis of U bijectively onto a basis of V.* □

Example 2.2.14 *Let $U = 2^{\{\spadesuit,\clubsuit,\heartsuit,\diamondsuit\}}$ and $V = (\mathbb{F}_2)^4$. Do we have a linear mapping sending $\{\heartsuit\}$ to, say, $(1,0,0,0)$? Yes, and many of them! We can still select the images of $\{\spadesuit\}$, $\{\clubsuit\}$, and $\{\diamondsuit\}$ arbitrarily in V and Theorem 2.2.11 guarantees us such a linear mapping.*

For example, suppose we choose $\phi(\{\spadesuit\}) = (0,0,1,0)$, $\phi(\{\clubsuit\}) = (0,0,0,1)$ and $\phi(\{\diamondsuit\}) = (0,1,0,0)$. Then the resulting linear mapping $\phi : U \to V$ is an isomorphism, because the standard basis of U is mapped bijectively onto the standard basis of V, and so above theorem applies.

On the other hand, the choice $\phi(\{\spadesuit\}) = \phi(\{\clubsuit\}) = \phi(\{\diamondsuit\}) = (0,1,0,1)$ leads to a mapping ϕ that is clearly not an isomorphism because it is not injective.

Each choice of images gives a different ϕ and so we have many such mappings.

In this example, can you see that there are in total 16^3 linear mappings sending $\{\heartsuit\}$ to $(1,0,0,0)$? Of course, not all of them are isomorphisms, but a large proportion of them are!

Once two vector spaces are isomorphic, there are great many different isomorphisms between them!

Let us finish this section with the following example.

Example 2.2.15 *Suppose U is a vector space with a basis B of size n. Let $V = {}^n\mathbb{F}$. What if we send the vectors from B to the vectors of the standard basis of V? That is, if $B = \{u_1, \ldots, u_n\}$ then we take the linear map ϕ sending u_1 to $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, u_2 to $\begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$, and so on. What is this linear mapping?*

If $u = c_1u_1 + c_2u_2 + \ldots + c_nu_n$ then $\phi(u) = c_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + c_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$. So $\phi(u) = [u]_B$, the coordinate column of u with respect to B! Hence ϕ is the coordinate mapping we introduced in Chapter 1. Note that ϕ maps B to the standard basis of ${}^n\mathbb{F}$, and so ϕ is an isomorphism. This explains why our coordinate methods work so well! There is an isomorphism behind them.

2.3 Linear transformations and normal forms

2.3.1 Linear transformations

Definition 2.3.1 *A linear transformation of a vector V is a linear mapping $\phi : V \to V$.*

That is, instead of two different vector spaces, there is only one vector space, V, and we view the linear mapping ϕ as an operation transforming V.

Example 2.3.2 *Let $V = \mathbb{R}^2$, the plane. Many familiar operations in the plane are linear transformations. First note that every linear mapping takes the zero vector to itself. So the zero vector cannot move.*

(a) Rotations of the plane around the origin (the zero vector) are linear transformations.

(b) If L is a straight line through the origin then the reflection in L is also a linear mapping.

(c) More generally, given any basis $\{u, v\}$ of $\mathbb{R}^2$ and two scalars $a, b \in \mathbb{R}$ there is a linear transformation that sends u to au and v to bv. This linear transformation "stretches" the plane by a factor of a in the direction of u and by a factor of b in the direction of v.

In the particular case where $a = 1$, $b = -1$ and the vectors u and v are perpendicular, we obtain the reflection in $L = \langle u \rangle$.

Example 2.3.3 *Similarly, in $\mathbb{R}^3$, all rotations around the origin are linear transformations, as are various "stretchings" of the space in up to three different independent directions.*

All of these also generalize to spaces of arbitrary dimension.

2.3.2 Matrix of a linear transformation

In this subsection we are dealing only with finite-dimensional vector spaces.

When we talked about the matrix of a linear mapping $\phi : U \to V$, we had to choose bases in U and V, say B in U and B' in V. When we are dealing with a linear transformation, we have $U = V$. In principle, we still can have separate bases B and B' in V, one in the source space V and the other in the target copy of V. However, it is more natural to use the same basis $B = B'$ for both copies of V, and this is how we will do it in most cases.

Thus, we give the following definition.

Definition 2.3.4 *Suppose that $\phi : V \to V$ is a linear transformation of V and suppose that $B = \{v_1, \ldots, v_n\}$ is a basis of V. The matrix of ϕ with respect to V is the matrix A, whose ith column, for each i, coincides with the coordinate column vector $[\phi(v_i)]_B$, that is, the ith column is made of the coordinates of $\phi(v_i)$ with respect to the basis B.*

This matrix A has the property that $[\phi(v)]_B = A[v]_B$ for all $v \in V$. That is, multiplication with A transforms the coordinates of v into the coordinates of $\phi(v)$.

Let us consider some examples.

Example 2.3.5 *Consider the differentiation D as a linear transformation of the space of polynomial functions P^n. Let us write the matrix of this linear transformation with respect to the standard basis of P^n. We have $D(1) = 0$, so the first column is $\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. Similarly, $D(x) = 1$, so the second column is*

$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. *In general,* $D(x^i) = ix^{i-1}$*, so the* $(i+1)$*st column has* i *is the* i*th position (*i*th row) and zeros elsewhere.*

So the matrix of D *is as follows:* $\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 2 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & n \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$.

The second example involves the space of matrices.

Example 2.3.6 *Suppose that* $V = M_{2\times 2}(\mathbb{R})$*, the spaces of* 2×2 *matrices with real entries. Suppose that* $\phi : V \to V$ *is defined by* $\phi(A) = A - A^T$*, the difference between* A *and its transpose,* A^T*. The standard basis here consists of the matrices* $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, *and* $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$*. We compute* $\phi(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\phi(\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $\phi(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, *and* $\phi(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$*. So the matrix of this linear transformation with respect to the standard basis is* $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

Before we switch to the topic of eigenvalues and eigenvectors, let us record the following observation following from the Rank+Nullity Theorem.

Theorem 2.3.7 *Suppose* V *is finite-dimensional. Then a linear transformation* $\phi : V \to V$ *is injective if and only if it is surjective if and only if it is bijective (i.e., an isomorphism).*

Furthermore, each and all of these conditions can be checked numerically as follows: the above conditions hold if and only if the determinant of the matrix of ϕ *is nonzero.* □

2.3.3 Eigenvalues and eigenvectors

Definition 2.3.8 (Eigenvector) *Suppose $\phi : V \to V$ is a linear transformation of V. A vector $v \in V$ is an eigenvector of ϕ with respect to $a \in \mathbb{F}$ if $\phi(v) = av$.*

Note that $v = 0$ is an eigenvector of ϕ with respect to any $a \in \mathbb{F}$. Indeed, $\phi(0) = 0 = a0$, so the condition holds, and 0 is an eigenvector with respect to a. Interestingly, in all other cases the scalar a that works for an eigenvector v is unique.

Theorem 2.3.9 *Suppose $v \neq 0$ is an eigenvector for a linear transformation $\phi : V \to V$. Then there is only one $a \in \mathbb{F}$ such that $\phi(v) = av$.*

Proof. Suppose $av = \phi(v) = bv$. Since $av = bv$ and $v \neq 0$, Theorem 1.1.6 (2b) implies that $a = b$. □

This motivates the following definition.

Definition 2.3.10 (Eigenvalue) *A scalar $a \in \mathbb{F}$ is an eigenvalue of a linear transformation $\phi : V \to V$ if $\phi(v) = av$ for a nonzero eigenvector $v \in V$.*

Now let us group all eigenvectors corresponding to a particular $a \in \mathbb{F}$ into one set.

Definition 2.3.11 (Eigenspace) *Suppose $a \in \mathbb{F}$. The eigenspace of $\phi : V \to V$ corresponding to a is the following set:*

$$V_a = \{v \in V \mid \phi(v) = av\}.$$

Theorem 2.3.12 *The subset V_a is a subspace of V for all $a \in \mathbb{F}$. Furthermore, $V_a = \ker(\phi - a\mathrm{id})$.*

Proof. Recall that $\mathrm{id} : V \to V$ is the identity map sending every $v \in V$ to itself, i.e., $\mathrm{id}(v) = v$.

We start by proving the second claim. First of all, note that $\phi - a\mathrm{id}$ is a linear transformation of V. Indeed, it is clearly a map from V to V, since $(\phi - a\mathrm{id})(v) = \phi(v) - a\mathrm{id}(v) = \phi(v) - av \in V$. Also, it is linear, since for all $u, v \in V$ and all $c \in \mathbb{F}$, we have that $(\phi - a\mathrm{id})(cu + v) = \phi(cu+v) - a(cu+v) = c\phi(u) + \phi(v) - acu - av = c(\phi(u) - av) + (\phi(v) - av) = c(\phi - a\mathrm{id})(u) + (\phi - a\mathrm{id})(v)$. (We used here that ϕ is linear.) Thus, $\phi - a\mathrm{id}$ is a linear transformation of V, as claimed.

If $v \in V_a$ then $\phi(v) = av$. Therefore, $(\phi - a\mathrm{id})(v) = \phi(v) - av = 0$, which means that $v \in \ker(\phi - a\mathrm{id})$. Conversely, if $v \in \ker(\phi - a\mathrm{id})$ then $(\phi - a\mathrm{id})(v) =$

0, which gives $\phi(v) - av = 0$, which yields $\phi(v) = av$. Therefore, $v \in V_a$. We have shown that V_a is the kernel of the linear map $\phi - a\mathrm{id}$.

Since the kernel of every linear map is a subspace, we can now conclude that V_a is a subspace. □

We note that $V_a \neq 0$ if and only if a is an eigenvalue of ϕ. This leads to a practical method of computing eigenvalues and eigenvectors of ϕ. Note that a is an eigenvalue of ϕ if and only if $V_a \neq 0$, that is, if V_a contains nonzero vectors. Hence we have the following result.

Theorem 2.3.13 *A scalar $a \in \mathbb{F}$ is an eigenvalue of ϕ if and only if* $\ker(\phi - a\mathrm{id}) \neq 0$. □

To turn this observation into a method, let us choose a basis B in V. We will only consider the case where V is finite dimensional, say, $\dim V = n$. Let A be the matrix of ϕ with respect to B. The matrix of $a\mathrm{id}$ with respect to any basis at all is aI_n, where I_n is the $n \times n$ matrix with 1s on the main diagonal and 0s elsewhere. Then the matrix of $\phi - a\mathrm{id}$ with respect to the basis B is $A - aI_n$.

Theorem 2.3.14 *Suppose ϕ is a linear transformation of an n-dimensional vector space V over $\mathbb{F}$ and suppose $a \in \mathbb{F}$. Then a is an eigenvalue of ϕ if and only if a satisfies the equation* $\det(A - aI_n) = 0$*, where A is the matrix of ϕ written with respect to an arbitrary basis of V.*

Proof. Since $A - aI_n$ is the matrix of $\phi - a\mathrm{id}$, written with respect to a basis B of V, we have that the nullity of $\phi - a\mathrm{id}$ is not zero if and only if the matrix $aI_n - P$ does not have n pivots in its echelon form, that is, if and only if $\det(A - aI_n) = 0$. □

Let us substitute a in $\det(A - aI_n) = 0$ by an indeterminate λ

Definition 2.3.15 (Characteristic polynomial) *The determinant of $A - \lambda I_n$ is a polynomial of degree n in indeterminate λ. It is called the* characteristic polynomial *of ϕ.*

It follows from the above that the eigenvalues of ϕ are exactly the roots (zeroes) of its characteristic polynomial.

Once the eigenvalues are found, we can deal with each one in turn. The eigenspace V_a for an eigenvalue a can be found from Theorem 2.3.12, as $\ker(\phi - a\mathrm{id})$.

We will now give a complete example how the eigenvalues and eigenvectors can be computed.

Example 2.3.16 *Suppose ϕ is a linear transformation of a vector space V over $\mathbb{R}$ of dimension 3. It is given to us that the matrix of ϕ with respect to some basis B is* $A = \begin{pmatrix} 1 & 2 & -4 \\ -10 & -11 & 20 \\ -6 & -6 & 11 \end{pmatrix}$. *This information saves us a couple of steps. We don't need to choose the basis and don't need to determine ourselves the matrix of ϕ with respect to that basis.*

Our task is to find all eigenvalues and all eigenvectors of ϕ. We start by finding the eigenvalues. We know that the eigenvalues are the roots of the characteristic polynomial f of ϕ, so we need to compute this polynomial. By definition, f is the determinant of the matrix

$$A - \lambda I_n = \begin{pmatrix} 1-\lambda & 2 & -4 \\ -10 & -11-\lambda & 20 \\ -6 & -6 & 11-\lambda \end{pmatrix}.$$

This leads to $f = -\lambda^3 + \lambda^2 + 5\lambda + 3$.

The roots of this polynomial (if integer) must divide the constant term 3. Trying various divisors of 3, we soon find that 3 and -1 are roots of f. The product of all three roots must be equal to 3. As $3 = 3(-1)(-1)$, we conclude that the roots are 3 and -1, and that -1 is a double root.

Therefore, the eigenvalues of ϕ are 3 and -1. We will find the eigenvectors for each of the two eigenvalues in turn. It is impossible to list all eigenvectors, as there are infinitely many of them, so instead we describe the eigenspace (the set of all eigenvectors for a given eigenvalue) by finding a basis in it. Then the eigenvectors are simply all linear combinations of that basis.

We start with the eigenvalue $\lambda = 3$. By Theorem 2.3.12, the eigenspace V_3 coincides with $\ker(\phi - 3\mathrm{id})$. *To find this kernel we switch to matrices. With respect to the same basis B, the matrix of $\phi - 3\mathrm{id}$ is $A - 3I_3$, where A is as before (the matrix of ϕ). It follows that the coordinate vectors (x_1, x_2, x_3) of eigenvectors for the eigenvalue 3 must satisfy the equation* $(A - 3I_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. *It is this equation that we need to solve.*

In the matrix form this is

$$\begin{pmatrix} -2 & 2 & -4 \\ -10 & -14 & 20 \\ -6 & -6 & 8 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

To solve this, we need to reduce the 3×3 matrix in the left side to its reduced echelon form *using the row operations.*

We first reduce to the echelon form getting:

$$\begin{pmatrix} -2 & 2 & -4 \\ 0 & -12 & 20 \\ 0 & 0 & 0 \end{pmatrix}$$

Note that we got a row of zeroes, which means that the system has nonzero solutions. This is an essential check: while solving a similar problem, if you don't get a row of zeroes then either you made an error in the row operations, or else the eigenvalue you are trying (in our case, it's 3) is not in fact an eigenvalue, and so you must have made an error with the characteristic polynomial.

The reduced echelon form is defined as an echelon form where (a) all pivots are equal to 1, and (b) the entries directly above pivots are zero.

Recall that the pivots in an echelon matrix are the first nonzero entries in the nonzero rows.

So in our echelon matrix the pivots are the entries -2 and -12.

To satisfy condition (a) we divide row 1 with -2 and row 2 with -12. This gives

$$\begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 \end{pmatrix}.$$

It remains to attend to condition (b). In general, it is more convenient to work with pivots from right to left (backward phase). However, in our particular case, we only need to clear one entry, over the second pivot. We add row 2 to row 1.

$$\begin{pmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 \end{pmatrix}$$

This is the reduced echelon form.

The reduced echelon form allows us to write immediately all solutions of the system.

Here is what the transformed system looks like:

$$\begin{array}{rcrcrcl} 1x_1 & + & 0x_2 & + & \frac{1}{3}x_3 & = & 0 \\ 0x_1 & + & 1x_2 & - & \frac{5}{3}x_3 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & = & 0 \end{array}$$

This gives us $x_1 = -\frac{1}{3}x_3$ and $x_2 = \frac{5}{3}x_3$, so we can express everything via x_3, which is the free variable *for this system. The free variables correspond to the columns that don't contain pivots.*

Finally, we have

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3}x_3 \\ \frac{5}{3}x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{3} \\ 1 \end{pmatrix}.$$

This means that the eigenvectors of A corresponding to the eigenvalue 3 are all multiples of the vector $u = \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{3} \\ 1 \end{pmatrix}$. Hence the eigenspace of A for eigenvalue $a = 3$ is 1-dimensional, spanned by the vector u. Note that in place of u we could take any its nonzero multiple. For example, $3u = \begin{pmatrix} -1 \\ 5 \\ 3 \end{pmatrix}$ looks a bit better.

(We need to stress that u is not itself an eigenvector of ϕ, but rather the B-coordinate vector of the actual eigenvector!)

After finishing the case of eigenvalue $a = 3$, we switch to the second eigenvalue, $\lambda = -1$. Here $A - (-1)I_3$ is as follows:

$$\begin{pmatrix} 2 & 2 & -4 \\ -10 & -10 & 20 \\ -6 & -6 & 8 \end{pmatrix}.$$

Its echelon form and reduced echelon forms are

$$\begin{pmatrix} 2 & 2 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence we need to solve

$$\begin{array}{rcrcrcl} 1x_1 & + & 1x_2 & - & 2x_3 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & = & 0 \\ 0x_1 & + & 0x_2 & + & 0x_3 & = & 0 \end{array}$$

This gives us $x_1 = -x_2 + 2x_3$, where x_2 and x_3 are the free variables.

Finally, we have

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_2 + 2x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}.$$

This means that the eigenvectors corresponding to the eigenvalue -1 are all linear combinations (x_2 and x_3 play the role of the arbitrary coefficients)

of the vectors $v = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ *and* $w = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$. *Hence the eigenspace for eigenvalue* $a = -1$ *is 2-dimensional, spanned by* v *and* w. *Note that the vectors we obtain in this way are automatically linearly independent and so they form a basis of the eigenspace. In particular, the dimension of the eigenspace is always equal to the number of free variables, that is, the number of columns without pivots.*

(Again we stress that while v *and* w *can be called eigenvectors of the matrix* A*, with respect to our initial linear transformation* ϕ *they are the* B*-coordinate vectors of eigenvectors.)*

2.3.4 Transition matrix

The purpose of this subsection is to discuss the relation between the coordinates of a vector $v \in V$ with respect to two different bases.

Theorem 2.3.17 *Suppose* V *is a vector space over* $\mathbb{F}$ *of a finite dimension* n *and suppose that* $B = \{v_1, \ldots, v_n\}$ *and* $B' = \{v'_1, \ldots, v'_n\}$ *are two bases of* V. *Then there exists a unique matrix* $C \in M_{n\times n}(\mathbb{F})$ *such that for every* $v \in V$ *we have* $[v]_{B'} = C[v]_B$. *Furthermore,* C *is invertible, that is,* C^{-1} *exists (and so* $\det C \neq 0$*).*

Recall that $[v]_B$ is the column vector consisting of the coordinates of v with respect to B (and so, similarly, $[v]_{B'}$ is the column of coordinates of v with respect to B'). Hence Theorem 2.3.17 provides a simple relation between the coordinates of vectors with respect to B and B'.

Definition 2.3.18 *The unique matrix* C *from Theorem* 2.3.17 *is known as the transition matrix from the basis* B *to the basis* B'.

We will now turn to the reasons why Theorem 2.3.17 holds, that is, in essence, we provide its proof.

Recall that for a linear mapping $\phi : U \to V$, where U and V are finite-dimensional, and for a choice of bases, X in U and Y in V, there is a unique matrix $A \in M_{n\times m}(\mathbb{F})$ such that $A[u]_X = [\phi(u)]_Y$. Here $m = \dim U$ and $n = \dim V$. This matrix A is constructed as follows: if x_i is the ith vector from the basis X then the ith column of A coincides with $[\phi(x_i)]_Y$, the column of coordinates of $\phi(x_i)$ with respect to Y.

We can apply this as follows. Let $U = V$ and so $m = n$. Let $\phi : U \to V$ be the identity mapping, $\phi = \mathrm{id}$. That is, $\phi(u) = u$ for all $u \in U = V$. Set $X = B$ and $Y = B'$. Then the above result gives us a matrix A such that

$Au_B = Au_X = \phi(u)_Y = u_Y = u_{B'}$, which is exactly what we need and so we simply take $C = A$.

Note that $\phi = \mathrm{id}$ is trivially an isomorphism and so $C = A$ is invertible, as claimed. This completes the discussion of the proof of Theorem 2.3.17.

Next, we record some properties of transition matrices.

Theorem 2.3.19 *Suppose B, B' and B'' are three bases of a vector space V. Suppose C is the transition matrix from B to B' and C' is the transition matrix from B' to B''. Then the transition matrix from B' to B is $D = C^{-1}$ and the transition matrix from B to B'' is $F = C'C$.*

Proof. Note that if $[v]_{B'} = C[v]_B$ then $C^{-1}[v]_{B'} = C^{-1}C[v]_B = I_n[v]_B = [v]_B$, and so $[v]_B = C^{-1}[v]_{B'}$. Hence $D = C^{-1}$ is the transition matrix from B' to B. Similarly, since $[v]_{B'} = C[v]_B$ and $[v]_{B''} = C'[v]_{B'}$, we get that $[v]_{B''} = C'[v]_{B'} = C'C[v]_B = F[v]_B$. So $F = C'C$ is the transition matrix from B to B''. □

Let us now see how this works for particular spaces and bases. Suppose $V = {}^n\mathbb{F}$, $B = \{u_1, \ldots, u_n\}$ is an arbitrary basis of V and B' is the standard basis of V. We claim that the matrix C is made simply of all the columns u_i. Indeed, $C = A$ has as its ith column $[\phi(u_i)]_{B'} = [u_i]_{B'}$. (Recall that $\phi = \mathrm{id}$.) Since B' is the standard basis of ${}^n\mathbb{F}$, we have $[u_i]_{B'} = u_i$. So u_i is the ith column of C, as claimed.

Example 2.3.20 *Suppose $V = {}^3\mathbb{R}$. Let $B = \left\{ \begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 7 \\ 10 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} \right\}$. Then the transition matrix from the basis B to the standard basis $E = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is simply the matrix $C = \begin{pmatrix} -1 & 0 & 0 \\ 3 & 7 & 2 \\ -2 & 10 & 3 \end{pmatrix}$. For example, if a vector u has coordinates $1, -1, 2$ with respect to B then its E-coordinates must be $\begin{pmatrix} -1 & 0 & 0 \\ 3 & 7 & 2 \\ -2 & 10 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -6 \end{pmatrix}$. Furthermore, since E is the standard basis, $[u]_E = u$ and so $u = \begin{pmatrix} -1 \\ 0 \\ -6 \end{pmatrix}$.*

Once we know how to find the transition matrix C from an arbitrary B to the standard basis, we also know how to find the transition matrix from the standard basis to B. Indeed, the latter is simply $D = C^{-1}$.

Example 2.3.21 *Let again* $V = {}^3\mathbb{R}$ *and* $B' = \{\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}\}$. *Then the transition matrix from* B' *to the standard basis* E *is* $C' = \begin{pmatrix} 3 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$. *Hence the transition matrix from the standard basis to* B' *is* $D = (C')^{-1} = \begin{pmatrix} 3 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -1 & -3 \\ -3 & 2 & 5 \\ -1 & 1 & 1 \end{pmatrix}$. *For example, if* $v = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ *then* $v = [v]_E$, *its own standard coordinate vector, and so the* B'*-coordinate vector of* v *must be* $[v]_{B'} = Dv = \begin{pmatrix} 3 \\ -4 \\ 0 \end{pmatrix}$.

Finally, in the case where we want to pass from one arbitrary basis to another arbitrary basis, we can do it via the standard basis E.

Example 2.3.22 *Suppose* B *and* B' *are as in the previous two examples. We know that* $C = \begin{pmatrix} -1 & 0 & 0 \\ 3 & 7 & 2 \\ -2 & 10 & 3 \end{pmatrix}$ *is the transition matrix from* B *to the standard basis* E, *and that* $D = (C')^{-1} = \begin{pmatrix} 2 & -1 & -3 \\ -3 & 2 & 5 \\ -1 & 1 & 1 \end{pmatrix}$ *is the transition matrix from* E *to* B'. *Hence the transition matrix from* B *to* B' *is* $F = DC = \begin{pmatrix} 2 & -1 & -3 \\ -3 & 2 & 5 \\ -1 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 \\ 3 & 7 & 2 \\ -2 & 10 & 3 \end{pmatrix} = \begin{pmatrix} 1 & -37 & -11 \\ -1 & 64 & 19 \\ 2 & 17 & 5 \end{pmatrix}$. *Therefore, if* w *has* B*-coodinates* $0, -1, 2$ *then* $[w]_{B'} = F[w]_B = \begin{pmatrix} 1 & -37 & -11 \\ -1 & 64 & 19 \\ 2 & 17 & 5 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 15 \\ -26 \\ -7 \end{pmatrix}$. *We conclude that that the coordinates of* w *with respect to* B' *are* $15, -26, -7$. *Can you see which vector is* w*?*

2.3.5 Similarity and normal forms

Suppose ϕ is a linear transformation of a vector space V. Given a basis B, we can write the matrix A of ϕ with respect to B. If we take a second basis,

B', then, for this basis, ϕ is represented by another matrix A'. What is the relation between A and A'?

Theorem 2.3.23 *Suppose ϕ is a linear transformation of V and suppose B and B' are two bases of V. Then for the matrices A and A' of ϕ with respect to B and B' we can write*

$$A' = CAC^{-1},$$

where C is the transition matrix from B to B'.

Proof. We know that A' is the unique matrix such that $\phi(v)_{B'} = A'v_{B'}$ for all $v \in V$. Since $v_{B'} = Cv_B$, $\phi(v)_{B'} = C\phi(v)_B$, and $\phi(v)_B = Av_B$, we obtain $\phi(v)_{B'} = C\phi(v)_B = CAb_B = CAC^{-1}v_{B'}$, proving that $A' = CAC^{-1}$. □

Note that instead of C we can use the transition matrix $D = C^{-1}$ from B' to B. In terms of this matrix we have equality $A' = D^{-1}AD$.

Definition 2.3.24 *Suppose two $n \times n$ matrices A and A' are related via $A' = CAC^{-1}$, where C is an invertible matrix of the same size. In this case we say that A and A' are similar (another terminology: conjugate).*

Thus, according to Theorem 2.3.23, similar (or conjugate) matrices represent the same linear transformation ϕ with respect to different bases. Hence the following question is valid. Can we find a basis where the matrix of ϕ looks particularly simple? Or, given a square matrix A, what is the simplest looking matrix similar to A?

This of depends of what we view as simple. Let us first consider the following definition.

Definition 2.3.25 *We say that a square matrix is* diagonal *if all its off-diagonal entries are zero. The entries on the diagonal (and of course, we mean the main diagonal, from the top left corner to the bottom right one) may or may not be zero.*

We say that a matrix is diagonalizable *if it is similar to a diagonal matrix.*

Definition 2.3.26 *A linear transformation $\phi : V \to V$ is* diagonalizable *if the matrix of ϕ for an arbitrary basis of V is diagonalizable. Equivalently, there exists a (special) basis of V, for which the matrix of ϕ is diagonal.*

Not every square matrix is diagonalizable, and so not every linear mapping is diagonalizable.

Theorem 2.3.27 *The matrix of ϕ with respect to a basis B of V is diagonal if and only if every vector from B is an eigenvector of ϕ.*

The proof is left as an exercise.

Hence, if we want to find a basis for which the matrix of ϕ is diagonal, we need to find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ of ϕ. Then for each λ_i we find the corresponding eigenspace V_{λ_i}, which gives us a basis B_i of V_{λ_i}. Let us take the union $B = B_1 \cup B_2 \cup \ldots \cup B_k$. We will show that this set of vectors is automatically linearly independent. Hence it is a basis if and only if its size coincides with the dimension of V.

To prove the claim we just made, we first need the following result.

Theorem 2.3.28 *Suppose that $u_1, u_2, \ldots, u_m \in U$ are non-zero eigenvectors of a linear transformation $\phi : U \to U$, corresponding to pairwise distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Then $u_1, u_2, \ldots, u_m$ are linearly independent.*

Proof. We prove this by induction on m, which we will organize as a proof *ad absurdum*. Namely, suppose that the claim of the theorem is not true. Then it should fail for some values of m and some non-zero eigenvectors u_i and so we can assume that in fact our given $u_1, u_2, \ldots, u_m$ are linearly dependent and m is the *smallest* for which such a dependent set of eigenvectors exists.

Since the eigenvectors are dependent, we can find scalars $c_i \in \mathbb{F}$, not all equal to zero, such that

$$c_1 u_1 + c_2 u_2 + \ldots + c_m u_m = 0. \tag{2.1}$$

Since ϕ is linear, we also have that

$$\begin{aligned} 0 =& \phi(0) \\ =& \phi(c_1 u_1 + c_2 u_2 + \ldots + c_m u_m) \\ =& c_1 \phi(u_1) + c_2 \phi(u_2) + \ldots + c_m \phi(u_m) \\ =& c_1 \lambda_1 u_1 + c_2 \lambda_2 u_2 + \ldots + c_m \lambda_m u_m. \end{aligned}$$

(We used that each u_i is an eigenvector of ϕ corresponding to λ_i; that is, $\phi(u_i) = \lambda_i u_i$.) Thus we also have the equality

$$c_1 \lambda_1 u_1 + c_2 \lambda_2 u_2 + \ldots + c_m \lambda_m u_m = 0. \tag{2.2}$$

Without of loss of generality, $c_m \neq 0$. Multiplying equation (2.1) with λ_m and subtracting the result from equation (2.2), we obtain

$$c_1(\lambda_1 - \lambda_m)u_1 + c_2(\lambda_2 - \lambda_m)u_2 + \ldots + c_m(\lambda_m - \lambda_m)u_m = 0.$$

Obviously, the last term on the left is zero, so we can also write

$$c_1(\lambda_1 - \lambda_m)u_1 + c_2(\lambda_2 - \lambda_m)u_2 + \ldots + c_{m-1}(\lambda_{m-1} - \lambda_m)u_{m-1} = 0.$$

By minimality of m, the set $\{u_1, u_2, \ldots, u_{m-1}\}$ cannot be linearly dependent. So each coefficient $c_i(\lambda_i - \lambda_m)$ here is zero. Furthermore, since the eigenvalues are pairwise distinct, $\lambda_i - \lambda_m \neq 0$, which means that $c_i = 0$ for $i = 1, 2, \ldots, m-1$.

So, in fact, the initial linear dependence relation (2.1) reduces to simply the equality $c_m u_m = 0$, since all other terms on the left are zero. However, $c_m u_m = 0$ implies that either $c_m = 0$ or $u_m = 0$. Now recall that it is given to us that none of the eigenvectors is zero, so $u_m \neq 0$, and hence $c_m = 0$, which is a contradiction. □

Now we can show the main claim.

Theorem 2.3.29 *Suppose $\lambda_1, \lambda_2, \lambda_k$ are the eigenvalues of a linear transformation $\phi : U \to U$. Suppose further that, for each i, the set B_{λ_i} is a basis of the eigenspace V_{λ_i}. Then $B = \cup_{i=1}^{k} B_i$ is linearly independent.*

Proof. Let $n_i = \dim V_{\lambda_i}$, so that $B_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,n_i}\}$ has cardinality n_i. Suppose that we have coefficients $c_{i,j} \in \mathbb{F}$ such that $\sum_{i=1}^{k} \sum_{j=1}^{n_i} c_{i,j} u_{i,j} = 0$. For $1 \leq i \leq k$, set $v_i := \sum_{j=1}^{n_i} c_{i,j} u_{i,j}$. Then, clearly, $v_i \in V_{\lambda_i}$.

We have that $\sum_{i=1}^{k} v_i = 0$. According to Theorem 2.3.28, this means that all v_i are zero. Indeed, if some v_i are non-zero, drop from the sum all zero terms. The remainder is a linear dependence relation between non-zero vectors corresponding to distinct eigenvalues, which is impossible by Theorem 2.3.28.

Thus all v_i are zero. However then $\sum_{j=1}^{n_i} c_{i,j} u_{i,j} = 0$. Since $B_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,n_i}\}$ is a basis of B_{λ_i}, these vectors are linearly independent, and so $c_{i,j} = 0$ for all i and j. This means that B is a linearly independent set. □

Clearly, we cannot have a larger linearly independent set consisting of eigenvectors. Therefore, we also have the following result.

Theorem 2.3.30 *A linear transformation ϕ of a vector space V is diagonalizable if and only if the dimensions of its eigenspaces add up to* $\dim V$.

Let us a consider an example.

Example 2.3.31 *Let us return to Example 2.3.16. For simplicity, we assume that $V = {}^3\mathbb{R}$ and that $A = \begin{pmatrix} 1 & 2 & -4 \\ -10 & -11 & 20 \\ -6 & -6 & 11 \end{pmatrix}$ is the matrix of ϕ with respect to the standard basis of V. In this case the vectors $u = \begin{pmatrix} -\frac{1}{3} \\ \frac{5}{3} \\ 1 \end{pmatrix}$,*

$v = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ *and* $w = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ *are the actual eigenvectors of* ϕ*. We see that the dimensions of the eigenspaces* V_3 *and* V_{-1}*, being* 1 *and* 2*, add up to the dimension of* V*. Hence* ϕ *is diagonalizable.*

Furthermore, $\{u, v, w\}$ *is the basis, for which the matrix of* ϕ *is diagonal. Namely, it is* $A' = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$*. Note that the diagonal entries are simply the eigenvalues for the corresponding eigenvectors!*

If ϕ were a linear transformation of ${}^3\mathbb{R}$ having two eigenvectors, with both eigenspaces 1-dimensional, then such a linear transformation ϕ wouldn't have been diagonalizable, as the dimensions of the eigenspaces add up to 2, which is less than $\dim V$.

Chapter 3

Bilinear forms and geometry

3.1 Basic definitions

Bilinear forms, Gram matrix, symmetric bilinear forms, radical, rank

3.1.1 Bilinear forms

Definition 3.1.1 *Suppose V is a vector space over $\mathbb{F}$. A bilinear form on V is a mapping $\Phi : V \times V \to \mathbb{F}$ (hence Φ takes two vector inputs and its output is a scalar) satisfying linearity in both arguments. That is, for all $u, v, w \in V$ and all $a \in \mathbb{F}$ we have:*

(B1) $\Phi(u + v, w) = \Phi(u, w) + \Phi(v, w)$ *and* $\Phi(au, v) = a\Phi(u, v)$*; and also*

(B2) $\Phi(u, v + w) = \Phi(u, v) + \Phi(u, w)$ *and* $\Phi(u, av) = a\Phi(u, v)$.

As usual, instead of (B1) we can verify that $\Phi(au + v, w) = a\Phi(u, w) + \Phi(v, w)$, and similarly, instead of (B2) we can verify that $\Phi(u, av + w) = a\Phi(u, v) + \Phi(u, w)$.

Example 3.1.2 (Standard bilinear form) *Suppose $V = \mathbb{F}^n$. For $u = (x_1, \ldots, x_n)$ and $v = (y_1, \ldots, y_n)$ define $(u, v) = \sum_{i=1}^{n} x_i y_i$. This is a bilinear form. If $\mathbb{F} = \mathbb{R}$ then this is the standard inner product (also known as the scalar product) that is used for computations in the Euclidean space.*

For an arbitrary $\mathbb{F}$, we will call this form the standard bilinear form on $V = \mathbb{F}^n$.

Example 3.1.3 (Function inner product) *Let $V = C([a, b])$ be the space of continuous functions on a closed interval $[a, b]$. For $f, g \in C([a, b])$, define $I(f, g) = \int_a^b f(x)g(x)dx$. This is a bilinear form on V. Let us check*

this. Take three vectors $f, g, h \in V = C([a,b])$, that is, three functions defined and continuous on the closed interval $[a,b]$. Let $c \in \mathbb{R}$. Then $I(cf + g, h) = \int_a^b (cf + g)(x)h(x)dx = \int_a^b (cf(x) + g(x))h(x)dx = \int_a^b [cf(x)h(x) + g(x)h(x)]dx = c\int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx = cI(f,h) + I(g,h)$, as claimed. Similarly, one can check that $I(f, cg + h) = cI(f,g) + I(f,h)$.

Let us record the following observation.

Theorem 3.1.4 *Suppose that Φ is a bilinear form on V and let $U \leq V$ be a subspace. Then the restriction of Φ to U is a bilinear form on U.* □

No proof is required here, because if the axioms (B1) and (B2) on all of V, they certainly hold on a subspace.

The above theorem allows us to view the form I from Example 3.1.3 as a bilinear form on any of the subspaces $C([a,b])$ that we considered: for example, the polynomial function spaces P^n and P^∞.

Here is a concrete computation:

Example 3.1.5 *Suppose $[a,b] = [0,1]$. Then $I(1,x) = \int_0^1 1 \cdot x dx = \int_0^1 x dx = \frac{1}{2}x^2|_0^1 = \frac{1}{2}$. In general, $I(x^s, x^t) = \int_0^1 x^{s+t} dx = \frac{1}{s+t+1}x^{s+t+1}|_0^1 = \frac{1}{s+t+1}$.*

On the other hand, if $[a,b] = [-1,1]$, say, then $I(1,x) = \int_{-1}^1 x dx = \frac{1}{2}x^2|_{-1}^1 = 0$.

Thus, the value of the form certainly depend on the choice of a and b.

3.1.2 The Gram matrix

The first order of business for us is how to represent the bilinear form in the coordinates.

Definition 3.1.6 (Gram matrix) *Suppose that Φ is a bilinear form on V and suppose that $B = \{v_1, \ldots, v_n\}$ is a basis of V. The matrix G with entries $(g_{ij}) = \Phi(v_i, v_j)$ is called the Gram matrix of Φ with respect to the basis B.*

This concept allows us to show that every bilinear form can be expressed in terms of coordinates.

Theorem 3.1.7 (Form in coordinates) *Suppose that G is the Gram matrix of the form Φ with respect to a basis $B = \{v_1, \ldots, v_n\}$ of V. Then, for $u = \sum_{i=1}^n x_i v_i$ and $v = \sum_{i=1}^n y_i v_i$, we have*

$$\Phi(u,v) = \sum_{i,j=1}^n g_{ij} x_i y_j.$$

Proof. Using bilinearity, we get $\Phi(u,v) = \Phi(\sum_{i=1}^n x_i v_i, v) = \sum_{i=1}^n x_i \Phi(v_i, v)$. Similarly, $\Phi(v_i, v) = \Phi(v_i, \sum_{j=1}^n y_j v_j) = \sum_{j=1}^n y_j \Phi(v_i, v_j)$. Substituting this into the first sum, we get the double sum as claimed. □

Note that we also can write this as

$$\Phi(u,v) = u_B^T G v_B,$$

where, as usual, u_B and v_B are the coordinate column vectors of u and v with respect to B. (Also, T indicates transposing.) Note that the right side of the equality is a 1×1 matrix, which we identify with a number, namely, the only entry of this matrix. Hence the equality makes sense.

Let us see what the Gram matrix looks like in a concrete situation.

Example 3.1.8 *Consider the standard bilinear form on $\mathbb{F}^n$, as in Example 3.1.2. Note that the entries x_i of u (and similarly, the entries y_i of v) in this example are the standard coordinates of u (and similarly, v). Comparing the expression $\sum_{i=1}^n x_i y_i$ with the expression from Theorem 3.1.7, we obtain*

$$g_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases}$$

So G is the identity matrix, $G = I_n$!

This can be confirmed directly by computation, using that the vector v_i from the standard basis has 1 in the ith position and zeroes everywhere else.

Clearly, if we choose a different basis then the Gram matrix may be completely different.

Example 3.1.9 *Let us now consider the function inner product. We cannot do a Gram matrix on $C([a,b])$ or even P^∞, because those spaces are infinite-dimensional, but we can do $V = P^n$ for a small n.*

Let us take $V = P^2$ and $[a,b] = [0,1]$. Then according to the computation carried out in Example 3.1.5, the Gram matrix of the function form I with respect to the standard basis $\{1, x, x^2\}$ is as follows:

$$G = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

3.1.3 Symmetric forms

Definition 3.1.10 (Symmetric form) *A bilinear form Φ on V is called symmetric if $\Phi(u,v) = \Phi(v,u)$ for all $u, v \in V$.*

This is the class of forms we will be focussing on. In particular, both examples of forms above are symmetric forms. We note that there are also other interesting classes of bilinear forms that we could discuss if we had more time. For example, a form Φ is anti-symmetric if $\Phi(u,v) = -\Phi(v,u)$ for all $u, v \in V$. There is also the class of symplectic forms, which are defined by the condition that $\Phi(u,u) = 0$ for each $u \in V$. Every symplectic form is anti-symmetric, but the converse is true only if $1 + 1 \neq 0$ in $\mathbb{F}$.

Having mentioned that, let us now concentrate on symmetric forms. First of all, how can we check that the form is symmetric? This can of course be checked directly, using the definition.

Example 3.1.11 *The standard bilinear form $(u,v) = \sum_{i=1}^{n} x_n y_n$ on $\mathbb{F}^n$ is symmetric. Indeed, $(v,u) = \sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} x_i y_i = (u,v)$.*

Similarly, the function inner product $I(u,v)$ is also symmetric: $I(g,f) = \int_a^b g(x)f(x)dx = \int_a^b f(x)g(x)dx = I(f,g)$ for all $f, g \in C([a,b])$.

We can also use the Gram matrix, if it is known.

Theorem 3.1.12 *Suppose that V is finite-dimensional. A bilinear form Φ on V is symmetric if and only if its Gram matrix G written with respect to an arbitrary basis of V is symmetric. (The latter means that $g_{ij} = g_{ji}$ for all i and j.)*

Proof. One way it is obvious. Indeed, if Φ is symmetric then $g_{ij} = \Phi(v_i, v_j) = \Phi(v_j, v_i) = g_{ji}$. Conversely, suppose $g_{ij} = g_{ji}$ for all i and j. This means that $G^T = G$. Therefore, $\Phi(v,u) = \Phi(v,u)^T = ((v_B)^T G u_B)^T = (u_B)^T G^T v_B = (u_B)^T G v_B = \Phi(u,v)$. Here we used that the number $\Phi(v,u)$ is viewed as a 1×1 matrix and, in this sense, it certainly is equal to its transpose. □

Checking the Gram matrices from the examples above, we see that they are symmetric and so those forms are indeed symmetric.

3.1.4 Radical and rank

Definition 3.1.13 (Radical) *Suppose that Φ is a symmetric bilinear form on V. The radical of Φ is the subset*

$$R(\Phi) = \{u \in V \mid \Phi(u,v) = 0 \text{ for all } v \in V\}$$

of V.

Theorem 3.1.14 *The radical $R(\Phi)$ of a symmetric bilinear form Φ on V is a subspace of V.*

Since $R(\Phi)$ is a subspace, it has a dimension. This dimension measures the degree of degeneracy of the form Φ. If $\dim R(\Phi) = n = \dim V$ then $R(\Phi) = V$, which means that Φ is the trivial form: $\Phi(u, v) = 0$ for all $u, v \in V$.

Definition 3.1.15 *The symmetric bilinear form Φ is called non-degenerate if $R(\Phi) = 0$.*

It can be checked that both the standard form on $\mathbb{F}^n$ and the function form I on $C([a, b])$ are non-degenerate.

Definition 3.1.16 *The rank of the symmetric bilinear form is the number $\dim V - \dim R(\Phi)$, the so-called co-dimension of the radical.*

Just like $\dim R(\Phi)$ measures the degree of degeneracy of Φ, the rank measures the degree of non-degeneracy. We say that Φ has full rank if the rank is $n = \dim V$. In this case $R(\Phi) = 0$ and so the form Φ is non-degenerate.

Note that we can also write

$$R(\Phi) = \{u \in V \mid \Phi(v, u) = 0 \text{ for all } v \in V\},$$

since Φ is symmetric. If Φ were not symmetric then instead of a single radical we would have to talk about the left radical and the right radical of Φ. However, these two happen to always have the same dimension, and so the rank of the form is well-defined even in the non-symmetric case.

Let us state without proof some practical results about the radical and rank.

Theorem 3.1.17 *Suppose Φ is a symmetric bilinear form on V. If G is the Gram matrix of Φ with respect to some basis B of V then $R(\Phi)$ consists of all vectors $u \in V$ satisfying $Gu_B = 0$.* □

Theorem 3.1.18 *Suppose Φ is symmetric with a Gram matrix G. Then the rank of Φ coincides with the number of pivots in an echelon form of G. (Also, $\dim R(\Phi)$ coincides with the number of free variables, i.e. the number of columns in the echelon form, not containing a pivot.)* □

Let us now see how this works in an example.

Example 3.1.19 *Suppose V is a vector space over $\mathbb{R}$ and $\dim V = 3$. Suppose that Φ is a bilinear form on V and its Gram matrix for some basis $B = \{v_1, v_2, v_3\}$ is as follows:*

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

First of all, this matrix is symmetric and so Φ is a symmetric bilinear form. In the coordinate form Φ is given by the following expression: for $u = x_1v + x_2v_2 + x_3v_3$ and $v = y_1v_1 + y_2v_2 + y_3v_3$, we have

$$\Phi(u, v) = x_1y_2 + x_2y_1 + x_2y_3 + x_3y_2.$$

It is easy to see that the (reduced) echelon form of G is

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

So the from Φ is degenerate of rank 2. The radical $R(\Phi)$ is 1-dimensional, and in fact, $R(\Phi)$ consists of all vectors u with B-coordinates satisfying $x_1 = -x_3$ and $x_2 = 0$. Hence $R(\Phi) = \langle v_1 - v_3 \rangle$.

3.2 Bilinear forms over the real numbers

Positive-definite forms, inner products, length, orthogonality, Cauchy-Schwarz inequality, angles, orthogonal projection, Gram-Schmidt orthogonalization.

3.2.1 Inner product spaces

In this section we consider vector spaces over $\mathbb{R}$.

Definition 3.2.1 (Positive-definite forms) *A symmetric bilinear form Φ on a vector space V over $\mathbb{R}$ is positive-definite if $\Phi(u, u) > 0$ for all $u \neq 0$.*

Positive-definite forms are also sometimes called inner products, in view of the following definition. We record the following fact.

Theorem 3.2.2 *Every positive-definite form is non-degenerate.*

Proof. Suppose Φ is a positive-definite form on V and suppose $u \in R(\Phi)$. Then $\Phi(u, v) = 0$ for all $v \in V$. In particular, taking $v = u$, we get $\Phi(u, u) = 0$. Since Φ is positive-definite, we conclude that $u = 0$. This shows that $R(\Phi) = 0$, and so Φ is non-degenerate. $\square$

We now switch to a single "unnamed" positive-definite form.

Definition 3.2.3 (Inner product space) *An inner product space is a pair $(V, (\cdot, \cdot))$ consisting of a vector space and a positive-definite bilinear form on it.*

Example 3.2.4 *Both the standard bilinear form on $V = \mathbb{R}^n$ and the function form I on $V = C([a, b])$ are positive-definite, and so these give us examples of inner product spaces.*

3.2.2 Length and distance

From this point and till the end of the section, $(V, (\cdot, \cdot))$ is an inner product space.

Definition 3.2.5 (Length) *For a vector $u \in V$, we call $\|u\| = \sqrt{(u, u)}$ the length (or norm) of u.*

By the above, the length is a non-negative number and $u = 0$ is the only vector of zero length. Without the square roots, we can write $\|u\|^2 = (u, u)$. Note that $\|au\| = \sqrt{(au, au)} = \sqrt{a^2(u, u)} = |a| \cdot \|u\|$.

Example 3.2.6 *Let $V = C([0, 1])$ and $(f, g) = I(f, g) = \int_0^1 f(x)g(x)dx$. What is the length of the vector x^s in this space? We have $\|x^s\|^2 = (x^s, x^s) = \frac{1}{2s+1}$. Hence, $\|x^s\| = \frac{1}{\sqrt{2s+1}}$.*

For example, $\|1\| = 1$, $\|x\| = \frac{1}{\sqrt{3}}$, etc.

Please note that the results of our computation crucially depend on the interval $[0, 1]$ that we chose. For a different interval we would have gotten different lengths.

Definition 3.2.7 (Distance) *The distance between two vectors $u, v \in V$ is $\|u - v\|$.*

Example 3.2.8 *What is the distance in $\mathbb{R}^4$ between $u = (1, 1, -2, 0)$ and $v = (-1, 2, -1, 1)$? By the above formula, we get $\|u - v\| = \|(2, -1, -1, -1)\| = \sqrt{4 + 1 + 1 + 1} = \sqrt{7}$.*

3.2.3 Orthogonality

We now turn to orthogonality.

Definition 3.2.9 (Orthogonality) *Two vectors $u, v \in V$ are orthogonal if $(u, v) = 0$.*

Note that if u and v are orthogonal then also au and bv are orthogonal for all $a, b \in \mathbb{R}$.

Theorem 3.2.10 (Pythagoras) *For $u, v \in V$, we have $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ if and only if $(u, v) = 0$.*

Proof. Indeed, $\|u + v\|^2 = (u + v, u + v) = (u, u) + (u, v) + (v, u) + (v, v) = \|u\|^2 + \|v\|^2 + 2(u, v)$. (Here we used that $(u, v) = (v, u)$, since the inner product is symmetric.) The claim now follows. □

3.2.4 Cauchy-Schwarz inequality

From the Pythagorean Theorem we can derive a very important inequality.

Theorem 3.2.11 (Cauchy-Schwarz inequality) *For $u, v \in V$, we have $(u, v)^2 \leq (u, u)(v, v) = \|u\|^2\|v\|^2$. The equality holds if and only if u and v are linearly dependent.*

Proof. If $u = 0$ then the claim is clear. So assume that $u \neq 0$.

Note that $z = v - \frac{(u,v)}{(u,u)}u$ is orthogonal to u. Indeed, $(u, z) = (u, v - \frac{(u,v)}{(u,u)}u) = (u, v) - \frac{(u,v)}{(u,u)}(u, u) = 0$. By Pythagoras, $\|v\|^2 = \|z + \frac{(u,v)}{(u,u)}u\|^2 = \|z\|^2 + \|\frac{(u,v)}{(u,u)}u\|^2 \geq \|\frac{(u,v)}{(u,u)}u\|^2 = \frac{(u,v)^2}{(u,u)^2}\|u\|^2 = \frac{(u,v)^2}{\|u\|^2}$. Here we used that $\|z\|^2 \geq 0$ and that $(u, u) = |u|^2$.

Thus, we have $\|v\|^2 \geq \frac{(u,v)^2}{\|u\|^2}$, yielding the inequality.

Clearly, the equality holds if and only if $z = 0$, and that is equivalent to $v = \frac{(u,v)}{(u,u)}u$, which holds if and only if u and v are linearly dependent. □

Example 3.2.12 *Let us see what this inequality is saying for our two examples.*

Let first $(V, (\cdot, \cdot))$ be $\mathbb{R}^n$ with the standard bilinear form $(u, v) = \sum_{i=1}^n x_i y_i$. In this case the Cauchy-Schwarz inequality reads:

$$(\sum_{i=1}^n x_i y_i)^2 \leq (\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2).$$

For the function space $V = C([a,b])$ and the integral form $I(f,g) = \int_a^b f(x)g(x)dx$, we get a different statement:

$$(\int_a^b f(x)g(x)dx)^2 \leq (\int_a^b f(x)^2dx)(\int_a^b g(x)^2dx).$$

Imagine proving these inequalities individually from scratch!

3.2.5 Angles

Now we turn to the implications of the Cauchy-Schwarz inequality.

Corollary 3.2.13 *For non-zero vectors $u, v \in V$, we have $-1 \leq \frac{(u,v)}{\|u\|\,\|v\|} \leq 1$.* □

The above corollary allows us to give the following definition.

Definition 3.2.14 (Angle) *Given two non-zero vectors $u, v \in V$, the angle α between u and v is defined via $\cos\alpha = \frac{(u,v)}{\|u\|\,\|v\|}$. This identifies α uniquely within the interval $[0,\pi]$.*

Immediately from this definition we obtain the following useful formula:

$$(u,v) = \|u\|\|v\|\cos\alpha.$$

In particular, $\alpha = 0$ means that $(u,v) = \|u\|\,\|v\|$ and this yields that $v = au$ for a non-negative $a \in \mathbb{R}$. Similarly, $\alpha = \pi$ means that $v = au$ for a non-positive $a \in \mathbb{R}$. Finally, $\alpha = \frac{\pi}{2}$ if and only if $(u,v) = 0$, that is, if and only if u and v are orthogonal. Because of this, we will also call orthogonal vectors perpendicular.

Example 3.2.15 *For $V = C([0,1])$, let us compute the angle between functions x^s and x^t. We get $\cos\alpha = \frac{(x^s,x^t)}{\|x^s\|\,\|x^t\|} = \frac{\frac{1}{s+t+1}}{\sqrt{\frac{1}{2s+1}\frac{1}{2t+1}}} = \frac{\sqrt{(2s+1)(2t+1)}}{s+t+1}$.*

For example, for 1 (i.e., $s = 0$) and x (i.e., $t = 1$), we get $\cos\alpha = \frac{\sqrt{3}}{2}$, and so the angle between 1 and x in $C([0,1])$ is $\frac{\pi}{6}$.

Again, note that the angle critically depends on the interval $[a,b]$ that we choose. For example, in $C([-1,1])$ the angle between 1 and x will be different.

As a consequence of the above definition, we get the famous result.

Theorem 3.2.16 (Law o cosines) *For (non-zero) $u, v \in V$, we have $\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\cos\alpha\|u\|\,\|v\|$, where α is the angle between u and v.*

Proof. Indeed, $\|u+v\|^2 = (u+v, u+v) = (u,u) + (v,v) + 2(u,v) = \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|\cos\alpha$. □

3.2.6 Orthogonal and orthonormal bases

Definition 3.2.17 *A set of vectors* $\{e_1, \ldots, e_k\}$ *in* V *is called* orthogonal *if* e_i *and* e_j *are orthogonal for all* $i \neq j$*. That is,* $(e_i, e_j) = 0$ *for* $i \neq j$*. Furthermore, this set is called* orthonormal *if additionally* $\|e_i\| = 1$ *for all* $1 \leq i \leq k$*.*

We will mostly deal with orthogonal bases. It is clear from the above definition that a basis is orthogonal if and only if the Gram matrix of the inner product with respect to this basis is diagonal. The basis is orthonormal if and only if the Gram matrix for this basis is the identity matrix I_n.

Let us record, without proof, how to construct an orthonormal set starting from an orthogonal one.

Theorem 3.2.18 *Suppose* $B = \{u_1, \ldots, u_k\}$ *is an orthogonal set in an inner product space* V*. Then* $B' = \{e_1, \ldots, e_k\}$*, where* $e_i = \frac{1}{\|u_i\|} u_i$*, is an orthonormal set, provided that none of the vectors* u_i *is zero.*

Hence we can switch from an orthogonal basis to an orthonormal one, by scaling the vectors.

Orthogonal and orthonormal bases are the most convenient ones to use. First of all, as we show now, orthogonal sets are automatically independent.

Theorem 3.2.19 *Suppose* $\{u_1, u_2, \ldots, u_k\}$ *is an orthogonal set of non-zero vectors. Then this set is linearly independent.*

Proof. Consider a linear combination equal to zero:

$$c_1 u_1 + c_2 u_2 + \ldots c_k u_k = 0,$$

for some $c_i \in \mathbb{F}$. We have that

$$\begin{aligned} 0 =& (u_i, 0) \\ =& (u_i, c_1 u_1 + c_2 u_2 + \ldots c_k u_k) \\ =& c_1 (u_i, u_1) + c_2 (u_i, u_2) + \ldots + c_k (u_i, u_k) \\ =& c_i (u_i, u_i), \end{aligned}$$

since $(u_i, u_j) = 0$ for all $j \neq i$ by orthogonality of the set. Thus, $c_i (u_i, u_i) = 0$. Note that the inner product is positive definite by definition and $u_i \neq 0$ by the assumption of the theorem. Hence $(u_i, u_i) > 0$, i.e., it is non-zero. We now deduce that $c_i = 0$. Since this is true for each i, the set is linearly independent. $\square$

This means that an orthogonal set of non-zero vectors is a basis if and only if it has the correct number of vectors, $n = \dim V$. Needless to say, an orthonormal set cannot contain the zero, as every vector in such a set has length 1. Thus, we can simply state that all orthonormal sets are linearly independent.

Let us now see how to find the coordinates of a vector with respect to an orthogonal basis. The idea of the method is in fact the same as in the above proof.

Theorem 3.2.20 *Suppose that $B = \{u_1, u_2, \ldots, u_n\}$ is an orthogonal basis in V. Then, for each $v \in V$, its coordinates c_i with respect to B can be computed as follows: $c_i = \frac{(v,u_i)}{(u_i,u_i)}$.*

Proof. By the definition of coordinates, we have that $v = c_1u_1 + c_2u_2 + \ldots + c_nu_n$. Hence,

$$\begin{aligned}(v, u_i) =&(c_1u_1 + c_2u_2 + \ldots + c_nu_n, u_i)\\ =&c_1(u_1, u_i) + c_2(u_2, u_i) + \ldots + c_n(u_n, u_i)\\ =&c_i(u_i, u_i),\end{aligned}$$

since all other terms are zero by orthogonality.

Thus, $(v, u_i) = c_i(u_i, u_i)$, and since $(u_i, u_i) \neq 0$ (since no vector of a basis is zero), we can also write that $c_i = \frac{(v,u_i)}{(u_i,u_i)}$ for each i. □

This takes an especially simple form when the basis is orthonormal, i.e., when $(u_i, u_i) = 1$ for all i.

Theorem 3.2.21 *If $B = \{u_1, u_2, \ldots, u_n\}$ is an orthonormal basis of V then the coordinates c_i of a vector $v \in V$ are given by the formula $c_i = (v, u_i)$.*

We finish this section with a comment that every subspace in an inner product space has orthogonal and orthonormal bases. We will develop a method of constructing orthogonal bases, known as the Gram-Schmidt process. However, first we need to discuss orthogonal projections.

3.2.7 Orthogonal projection

Definition 3.2.22 *Suppose U is a subspace of V. We say that $v \in V$ is* orthogonal *(or perpendicular) to U if $(v, u) = 0$ for all $u \in U$; that is, v is orthogonal to every $u \in U$.*

In reality, we only need to check this property for finitely many vectors u.

Theorem 3.2.23 *The vector v is orthogonal to a subspace U if and only if v is orthogonal to every vector in an arbitrary spanning set of U.*

For example, we can verify that v is orthogonal to the vectors from an arbitrary basis of U.

Now we turn to the concept of orthogonal projection.

Theorem 3.2.24 *Suppose U is a subspace of an inner product space V and let $v \in V$. Then there exists a unique vector $p \in U$ such that $w = v - p$ is orthogonal to the whole subspace U.*

Proof. Here we only verify uniqueness of p, leaving its existence until later (see Theorem 3.2.26). Suppose there exist two vectors, $p, p' \in U$, such that both $w = v - p$ and $w' = v - p'$ are orthogonal to U. Consider the vector $t = w - w'$. First note that that t is orthogonal to U. Indeed, for $u \in U$, we have that $(t, u) = (w - w', u) = (w, t) - (w', u) = 0 - 0 = 0$, since both w and w' are orthogonal to U. On the other hand, $t = w - w' = (v - p) - (v - p') = -p + p' \in U$. Therefore, t must be orthogonal to itself, i.e., $(t, t) = 0$. However, since the inner product is positive definite, this is only possible when $t = 0$.

Thus, $t = 0$, or equivalently, $w = w'$ and $p = p'$, proving uniqueness. □

This results motivates the following definition.

Definition 3.2.25 *For a subspace U of an inner product space V and a vector $v \in V$, the unique vector $u \in U$ such that $v - u$ is orthogonal to U is called the* orthogonal projection *of v to U. We will write $u = \mathrm{proj}_U(v)$.*

We will only discuss how to find the orthogonal projection in the particular case, where we have already selected an orthogonal basis in the subspace U.

Theorem 3.2.26 *Suppose $B = \{u_1, \ldots, u_k\}$ is an orthogonal basis in a subspace U of V. Then, for every $v \in V$, we have $\mathrm{proj}_U(v) = c_1u_1 + c_2u_2 + \ldots + c_ku_k$, where $c_i = \frac{(v,u_i)}{(u_i,u_i)}$ for all i.*

Proof. Let $p := c_1u_1 + c_2u_2 + \ldots + c_ku_k$, where, as above, $c_i = \frac{(v,u_i)}{(u_i,u_i)}$ for all i. (Note that this is exactly the same formula that used for the coordinates with respect to an orthogonal basis.) To show that p is the projection, we

need to check that $w = v - p$ is orthogonal to U. By Theorem 3.2.23, it suffices to check that w is orthogonal to every u_i. We have that

$$\begin{aligned}(w, u_i) =& (v - p, u_i)\\ =& (v - c_1u_1 + c_2u_2 + \ldots + c_ku_k, u_i)\\ =& (v, u_i) - c_1(u_1, u_i) - c_2(u_2, u_i) - \ldots - c_k(u_k, u_i)\\ =& (v, u_i) - c_i(u_i, u_i)\\ =& (v, u_i) - \frac{(v, u_i)}{(u_i, u_i)}(u_i, u_i)\\ =& (v, u_i) - (v, u_i) = 0.\end{aligned}$$

So indeed, w is orthogonal to U and p is the projection of v to U. □

In particular, this shows that the projection exists, as we found an explicit vector p satisfying the definition.

Example 3.2.27 *Suppose $V = \mathbb{R}^3$ with the standard inner product $(u, v) = x_1y_1+x_2y_2+x_3y_3$, where as usual $u = (x_1, x_2, x_3)$ and $v = (y_1, y_2, y_3)$. Suppose $U = \langle e_1, e_2\rangle$, where $e_1 = (1, 1, -2)$ and $e_2 = (1, 1, 1)$. Clearly e_1 and e_2 are orthogonal, so they form an orthogonal basis of U.*

Take $v = (3, 1, -1)$. Then $\mathrm{proj}_U(v) = \frac{(v,e_1)}{(e_1,e_1)}e_1 + \frac{(v,e_2)}{(e_2,e_2)}e_2 = \frac{6}{6}(1, 1, -2) + \frac{3}{3}(1, 1, 1) = (2, 2, -1)$. It is easy to check that the difference between v and this vector is indeed perpendicular to both e_1 and e_2 and hence to the whole of U.

3.2.8 The Gram-Schmidt orthogonalization process

As an application of the orthogonal projection, we describe how to transform an arbitrary basis of V into an orthogonal one. This is known as the Gram-Schmidt orthogonalization.

Suppose $B = \{e_1, \ldots, e_n\}$ is a basis of an inner product space V. Define $U = \langle e_1, \ldots, e_i\rangle$. This includes $U_0 = 0$ and $U_n = V$.

We will inductively construct an orthogonal basis $B' = \{f_1, \ldots, f_n\}$ in such a way that $\langle f_1, \ldots, f_i\rangle = U_i = \langle e_1, \ldots, e_i\rangle$. Namely, for each i in turn we set $f_i = e_i - \mathrm{proj}_{U_{i-1}}(e_i)$. By the above, f_i is orthogonal to the subspace U_{i-1} and hence f_i is orthogonal to all vectors f_j, $j < i$, since by the inductive assumption $U_{i-1} = \langle f_1, \ldots, f_{i-1}\rangle$.

To complete the induction we need to show that we also have $U_i = \rangle f_1, \ldots, f_i\rangle$, and this is now easy to check.

Since every f_i is orthogonal to all f_j with $j < i$, we conclude that B' is inndeed an orthogonal basis of V.

Let us write down the explicit expressions for the vectors f_i. We have $f_1 = e_1 - 0 = e_1$, $f_2 = e_2 - \frac{(e_2,f_1)}{(f_1,f_1)}$, $f_3 = e_3 - \frac{(e_3,f_1)}{(f_1,f_1)} - \frac{(e_3,f_2)}{(f_2,f_2)}$, and so on.

We conclude this discussion with an example.

Example 3.2.28 (Gram-Schmidt process) *Let $V = \mathbb{R}^3$ and suppose that the Gram matrix of our bilinear form for the standard basis $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, and $e_3 = (0,0,1)$ looks as follows:*

$$G = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 1 \end{pmatrix}.$$

Note that the matrix is symmetric and so the bilinear form is symmetric. Let us use the Gram-Schmidt to construct an orthogonal basis starting from e_1, e_2, and e_3.

Hence $f_1 = e_1 = (1,0,0)$. We set $f_2 = e_2 - \alpha_2^1 e_1$, where $\alpha_2^1 = \frac{(e_2,f_1)}{(f_1,f_1)} = \frac{(e_2,e_1)}{(e_1,e_1)} = g_{21}/g_{11} = 1/1 = 1$. Therefore, $f_2 = e_2 - 1e_1 = (0,1,0) - (1,0,0) = (-1,1,0)$.

We will need $(f_2,f_2) = (e_2 - e_1, e_2 - e_1) = (e_2,e_2) - (e_1,e_2) - (e_2,e_1) + (e_1,e_1) = 2 - 1 - 1 + 1 = 1$. Thus, $(f_2,f_2) = 1$. It remains to find f_3. We set it to be $f_3 = e_3 - \alpha_3^1 e_1 - \alpha_3^2 e_2$ and we compute $\alpha_3^1 = \frac{(e_3,f_1)}{(f_1,f_1)} = g_{31}/g_{11} = 2/1 = 2$ and $\alpha_3^2 = \frac{(e_3,f_2)}{(f_2,f_2)} = \frac{(e_3,e_2-e_1)}{1} = (e_3,e_2) - (e_3,e_1) = g_{32} - g_{31} = (-1) - 2 = -3$. Thus, $\alpha_3^1 = 2$ and $\alpha_3^2 = -3$. This gives us $f_3 = e_3 - 2f_1 + 3f_2 = (0,0,1) - (2,0,0) + (-3,3,0) = (-5,3,1)$.

Our new orthogonal basis is $f_1 = (1,0,0)$, $f_2 = (-1,1,0)$, and $f_3 = (-5,3,1)$. Note that $f_3 = e_3 - 2e_1 + 3(e_2 - e_1) = -5e_1 + 3e_2 + e_3$ and so $(f_3,f_3) = (-5e_1 + 3e_2 + e_3, -5e_1 + 3e_2 + e_3) = (-5,3,1)G(-5,3,1)^T = (0,0,-12)(-5,3,1)^T = -12$. Thus, the Gram matrix with respect to the new orthogonal basis is as follows

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -12 \end{pmatrix}.$$

The Gram-Schmidt process can also be used to verify that the symmetric bilinear form with a given Gram matrix is positive definite (i.e., it is an inner product). Start computing the new basis and if for some i, you see that (f_i,f_i) is zero or negative then the bilinear form is not positive definite. If $(f_i,f_i) > 0$ for all i then it is an inner product. In the above example $(f_1,f_1) = 1 > 0$, $(f_2,f_2) = 1 > 0$, but $(f_3,f_3) = -12 < 0$, so in fact our bilinear form is not an inner product.