

## Lecture 1

# Vector spaces and bases

### 1.1 Abstract vector spaces (review)

A vector space  $V$  is an algebraic structure motivated by the idea of parameterizing all possible displacements of a particular object in space. Here is an incomplete list of intuitive notions about displacements that will lend insight to the formal definition (which is notoriously opaque).

1. There is a “neutral displacement” which consists in not moving at all.
2. Displacements can be composed. If “one mile north” and “two miles east” are two examples of displacements, then “one mile north and two miles east” counts too.
3. In the above example, order of composition doesn’t matter. You can first go north, then east; or you can first go east, then north. Either path leads to the same overall displacement.
4. Given a displacement, you should be able to state *how much* of it to refer to. For example, the displacement from the top of my head to the bottom of my torso is approximately three times that from the top of my head to my chin.
5. There is no obviously meaningful way to “multiply” two displacements to result in another one. Where would I end up if I were displaced by the product of “New York to Los Angeles” and “Chicago to Seattle”?

**Informal Definition 1.** A *field* is a collection of numbers with additive identity 0 and multiplicative identity 1 that you can add, subtract, multiply, and divide (except by 0) in the usual way.

For our purposes, the field is almost always  $\mathbb{C}$ , the complex numbers.<sup>1</sup>

**Definition 1** (Vector space). A *vector space* over a field of scalars is a set on which addition and scalar multiplication are defined satisfying the following:

1. For all vectors  $x$  and  $y$ ,  $x + y = y + x$ . (commutativity of addition)
2. For all vectors  $x$ ,  $y$ , and  $z$ ,  $(x + y) + z = x + (y + z)$ . (associativity of addition)
3. There is a neutral vector  $0$  such that  $x + 0 = x$ . (additive identity)
4. For each vector  $x$  there is a vector  $y$  such that  $x + y = 0$ . (additive inverse, “subtraction”)
5. For each vector  $x$ ,  $1x = x$ . (multiplicative identity)

<sup>1</sup>A fuller definition and theory of fields is not a part of this class, but if you’re curious, there are lots of fields out there! Other examples are  $\mathbb{R}$ , the real numbers;  $\mathbb{Q}$ , the rational numbers;  $\mathbb{Q}[j]$ , the rational numbers with  $\sqrt{-1}$  thrown in; and  $\mathbb{Z}_p$ , the integers modulo a prime  $p$ .

6. For each pair of scalars  $\alpha$  and  $\beta$  and each vector  $x$ ,  $(\alpha\beta)x = \alpha(\beta x)$ . (associativity of scalar multiplication)
7. For each scalar  $\alpha$  and pair of vectors  $x$  and  $y$ ,  $\alpha(x + y) = \alpha x + \alpha y$ . (left distributivity of scalar multiplication)
8. For each pair of scalars  $\alpha$  and  $\beta$  and vector  $x$ ,  $(\alpha + \beta)x = \alpha x + \beta x$ . (right distributivity of scalar multiplication)

There are a lot of definitions and facts about vector spaces from 16A or 54 that are necessary to go further, e.g. that a vector space has exactly one additive identity element. We rely on a lot of commonsense technical results about existence and uniqueness, but we will not interrogate them. The following are some of the most important ones.

**Definition 2** (Linear independence). The vectors  $v_1, v_2, \dots, v_n$  are **linearly independent** if for any scalars  $\alpha_1, \alpha_2, \dots, \alpha_n$ ,<sup>2</sup>  $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = 0$  implies  $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ .

**Definition 3** (Basis). A **basis** of a vector space  $V$  is a maximal list of linearly independent vectors: if you added even one more vector from  $V$ , it would not longer be linearly independent.

**Definition 4** (Dimension). The dimension of a vector space  $V$  is the number of vectors in any basis of  $V$ .

**Theorem 1** (Criteria for vectors to form a basis). A list of vectors  $v_1, v_2, \dots, v_m$  is a basis for  $V$  if any two of the following three conditions hold:

1.  $m = \dim V$ .
2.  $v_1, v_2, \dots, v_m$  generate  $V$ .
3.  $v_1, v_2, \dots, v_m$  are linearly independent.

### Very important example: $k^n$ as a vector space over $k$

Given a field  $k$  and an integer  $n > 0$ , we can form the vector space  $k^n$  of  $n$ -tuples of scalars from  $k$ . Vectors are expressions of the form  $v = (v_1, v_2, \dots, v_n)$ <sup>3</sup> and are sometimes written as “column vectors”:

$$(v_1, v_2, \dots, v_n) = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Addition is defined by the rule  $(v_1, v_2, \dots, v_n) + (v'_1, v'_2, \dots, v'_n) = (v_1 + v'_1, v_2 + v'_2, \dots, v_n + v'_n)$ . This can be communicated concisely using a variable subscript:  $(v + v')_i = v_i + v'_i$ .<sup>4</sup>

Scalar multiplication is defined by the rule  $(\alpha v)_i = \alpha(v_i)$ .

It can be verified that  $k^n$  with addition and scalar multiplication defined this way is a vector space over  $k$ .

**Theorem 2.** The dimension of  $k^n$  is  $n$ .

<sup>2</sup>A note on notation: it's customary to write the shorthand “ $1, 2, \dots, n$ ” to signify “all of the numbers from 1 to  $n$ ,” even if  $n \leq 2$ . If  $n$  is 0, it's the empty list; if  $n$  is 1 or 2 then the enumeration is understood to end at  $n$ .

<sup>3</sup>Setting off the components with commas distinguishes this expression from a “row vector.”

<sup>4</sup>This equation would be read aloud as “component  $i$  of  $V$  plus  $V$ -prime equals component  $i$  of  $V$  plus component  $i$  of  $V$ -prime.”

### Unimportant, but interesting example: $\mathbb{C}^n$ and its evil twin $\overline{\mathbb{C}^n}$

We will spend most of our time in the vector space  $\mathbb{C}^n$ , defined in the usual way above. In this section we will define a *different* vector space structure (addition and scalar multiplication) on the set  $\mathbb{C}^n$ . The vector space we are about to make is evil and strange! (It is utterly useless except for scaring your friends.)

Let's call it " $\overline{\mathbb{C}^n}$ ." The underlying set is  $n$ -tuples of complex numbers, and addition is defined componentwise as usual. Define scalar multiplication by  $(\alpha v)_i = \bar{\alpha}(v_i)$ . That is, "multiplying" by a scalar actually scales the vector's components by the conjugate of the scalar.

(Is there an isomorphism of vector spaces between normal  $\mathbb{C}^n$  and  $\overline{\mathbb{C}^n}$ ?<sup>5</sup>)

## 1.2 Coordinates on vector spaces, and abuse of notation

There is a difference between *concept* and *representation* of vector spaces.<sup>6</sup> Vectors and linear transformations<sup>7</sup> have a life of their own and do not mind what we think of them. However, in order to work with them hands-on, we need to represent them in a system of coordinates.

A choice of coordinates is like a digital imaging system. If I photograph a cat, the photo is not the cat; it is a representation of the cat. I can choose to make the photo brighter or darker. I can mirror it or rotate it, and even though the image might be very different, the cat is the same; all I have changed is the means of representation.

We represent vectors and linear maps by choosing bases.

**Definition 5** (Coordinates of a vector). *The coordinates of a vector  $v$  in the vector space  $V$  with respect to the basis  $\{v_1, v_2, \dots, v_n\}$  are the unique scalars  $(\alpha_1, \alpha_2, \dots, \alpha_n)$  such that  $v = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ .*

### Example: two bases for $\mathbb{C}^3$

We will represent the vector  $(1, -\sqrt{3}, \sqrt{3})$  in two different bases.

$$\begin{pmatrix} 1 \\ -\sqrt{3} \\ \sqrt{3} \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \sqrt{3} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \sqrt{3} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (1.1)$$

$$= 1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + j \begin{pmatrix} 1 \\ -1/2 + \sqrt{3}/2j \\ -1/2 - \sqrt{3}/2j \end{pmatrix} - j \begin{pmatrix} 1 \\ -1/2 - \sqrt{3}/2j \\ -1/2 + \sqrt{3}/2j \end{pmatrix} \quad (1.2)$$

The coordinates of  $(1, -\sqrt{3}, \sqrt{3})$  relative to the first basis are  $(1, -\sqrt{3}, \sqrt{3})$ . This basis is called the standard basis, and its vectors are usually named  $e_1, e_2, e_3$ .

The coordinates of  $(1, -\sqrt{3}, \sqrt{3})$  relative to the second basis are  $(1, j, -j)$ .

**Definition 6** (Coordinates of a matrix). *The representation of the linear function  $f : V \rightarrow U$  relative to the basis  $\{v_1, v_2, \dots, v_n\}$  for  $V$  and the basis  $\{u_1, u_2, \dots, u_m\}$  for  $U$  are the  $m \times n$  matrix  $A$ . Column  $j$  of  $A$  is the coordinate vector of  $f(v_j)$  relative to  $\{u_1, u_2, \dots, u_m\}$ .*

**Theorem 3.** *Coordinate representations of vector spaces are faithful: matrix-vector multiplication represents function application, and matrix-matrix multiplication represents function composition.*

<sup>5</sup>Yes, because all equidimensional vector spaces over the same field are isomorphic. If we use the standard bases for  $\mathbb{C}^n$  and  $\overline{\mathbb{C}^n}$ , the matrix of an isomorphism is  $I$ . In my opinion the proof is somewhat confusing.

<sup>6</sup>For this wording I am indebted to Hegel's theory of knowledge as a synthesis of *Begriff* (tr. concept/idea) and *Vorstellung* (tr. representation).

<sup>7</sup>A linear transformation from a vector space to itself is called an operator.

In applied mathematical subjects (such this class), we will often use  $\mathbb{C}^n$  with the standard basis. As a result, we can, and sometimes will delude ourselves with notation that appears to identify the representation of a vector with the vector itself. Likewise, sometimes, but not always, we will identify linear functions with their matrix representations.

### 1.3 Diagonalization

Throughout this section, let  $T : V \rightarrow V$  be a linear map from vector space  $V$  to itself.<sup>8</sup>

**Definition 7.** (*Eigenvalues and eigenvectors*) If a scalar  $\lambda$  and a nonzero vector  $v$  satisfy the relationship  $T(v) = \lambda v$ , then  $v$  is called an **eigenvector** of  $T$  for **eigenvalue**  $\lambda$ .

Note that the eigenvector-eigenvalue condition is equivalent to saying that the linear map  $(T - \lambda I) : V \rightarrow V$  has a nontrivial nullspace, which can be called the eigenspace of  $T$  for eigenvalue  $\lambda$ .

**Theorem 4.** (*Diagonalization*) If

$$\begin{aligned} T(v_1) &= \lambda_1 v_1, \\ T(v_2) &= \lambda_2 v_2, \\ &\vdots = \vdots \\ T(v_n) &= \lambda_n v_n \end{aligned}$$

are eigenvalue-eigenvector pairs such that  $\{v_1, v_2, \dots, v_n\}$  is a basis of  $V$ , then the representation of  $T$  in this basis is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

where the empty spaces are zero.

*Proof.* To make column  $i$  of the matrix representation of  $T$  relative to this basis, we must represent  $T(v_i)$  in the basis  $\{v_1, v_2, \dots, v_n\}$ . As  $v_i$  was chosen to be an eigenvector,  $T(v_i) = \lambda_i v_i$ . The coordinates of  $\lambda_i v_i$  in this basis are  $\lambda_i$  in place  $i$  and 0 everywhere else.  $\square$

#### On determinants

For computing and discussing eigenvalues we humans<sup>9</sup> need to use the scalar-valued matrix function  $\det : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}$ , characterized by the following:

1.  $\det$  is linear in every column.
2.  $\det A = 0$  if the columns of  $A$  are linearly dependent.
3.  $\det I = 1$ .

$\det$  is a polynomial function with a number of important properties and interpretations, many of which are covered in a prerequisite course. Our primary use will be in a function called the characteristic polynomial  $\chi_A(s)$  that reveals in its roots the eigenvalues of  $A$ . Given a matrix  $A$ ,  $\chi_A(s) = \det(sI - A)$ , or sometimes  $\det(A - sI)$  (which works equally well).

<sup>8</sup>It is safe to think of  $V$  as  $\mathbb{C}^n$  and  $T$  as a matrix, but you should nevertheless prefer the toolkit to the laws of vector spaces and linear maps to that of coordinate manipulation.

<sup>9</sup>Generally computers have more efficient approximations.

### Trivial diagonalization example

Let  $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$  be dilation by factor of  $\rho$ ; that is,  $T(v) = \rho v$ . Choose any basis  $\{v_1, v_2, \dots, v_n\}$ . The image of  $v_i$  under  $T$  can be constituted in this basis as  $\rho$  of  $v_i$  and 0 of every other basis vector. Therefore the matrix of  $T$  in any basis is  $\rho I$ .

### Easy diagonalization example

The vector space is  $\mathbb{C}^2$ , and the operator is left multiplication by the matrix

$$A = \begin{pmatrix} 1 & j \\ -j & 1 \end{pmatrix} \quad (1.3)$$

This matrix has characteristic polynomial

$$\chi_A(s) = s^2 - 2s = s(s - 2) \quad (1.4)$$

and therefore eigenvalues 0 and 2, which we will keep in that order. The eigenvector corresponding to 0 comes from the null space of  $A - 0I$ .  $(1, j)$  will do. The eigenvector corresponding to 2 comes from the null space of  $A - 2I$ . I'll take  $(j, 1)$ . To convert eigenbasis coordinates to standard basis coordinates, line up these eigenvectors side by side.

$$V = \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix} \quad (1.5)$$

The matrix that converts standard basis coordinates to eigenbasis coordinates is  $V^{-1}$ :

$$V^{-1} = \begin{pmatrix} 1/2 & -j/2 \\ -j/2 & 1/2 \end{pmatrix} \quad (1.6)$$

Therefore  $A$  can be factored as follows.

$$A = \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/2 & -j/2 \\ -j/2 & 1/2 \end{pmatrix} \quad (1.7)$$

### Hard diagonalization example

Let  $T$  be the operator on  $\mathbb{C}^3$  that circularly shifts the coordinates one place to the right. It maps  $e_1$  to  $e_2$  and  $e_2$  to  $e_3$ , and wraps  $e_3$  around to  $e_1$ ; so its coordinates in the standard basis are the following matrix.

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (1.8)$$

This matrix can be diagonalized with the help of the characteristic polynomial as above, but we'll take a different, more scenic path. Parameterize an eigenvalue-eigenvector pair  $(\lambda, v)$  and solve the relation directly.

$$A \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_3 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \lambda v_3 \end{pmatrix} \quad (1.9)$$

This relationship says that  $v_1 = \lambda v_2$ ,  $v_2 = \lambda v_3$ , and  $v_3 = \lambda v_1$ . Substituting,

$$v_3 = \lambda v_1 \quad (1.10)$$

$$= \lambda^2 v_2 \quad (1.11)$$

$$= \lambda^3 v_3 \quad (1.12)$$

$$v_3(1 - \lambda^3) = 0 \quad (1.13)$$

According to the zero product property,  $v_3 = 0$  or  $\lambda^3 = 1$ . But  $v_3 = 0$  would imply  $v = 0$ , which is not an eigenvector. Therefore  $\lambda^3 = 1$ :  $\lambda$  is a number that you can cube and get 1 back as result. Parameterize  $\lambda$  in polar form as  $re^{j\theta}$ .

$$(re^{j\theta})^3 = 1 \quad (1.14)$$

$$r^3 e^{3j\theta} = 1 e^{j(0+2\pi n)}, \quad (1.15)$$

where  $n$  is any integer. Matching magnitude and phase, we have

$$r = 1 \text{ and} \quad (1.16)$$

$$3\theta = 0, 2\pi, 4\pi, 6\pi, \dots \quad (1.17)$$

$$\theta = 0, \frac{2\pi}{3}, \frac{4\pi}{3} \quad (1.18)$$

### Non-diagonalizable example

The following matrix  $A$  has eigenvalue 0, but  $(A - 0I)$  has a null space of dimension only 1.

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$$

Therefore  $A$  does not have a basis of eigenvectors and is not diagonal in any basis.

## 1.4 Solving vector differential equations

Diagonalization is able to factor most square matrices  $A = V\Lambda V^{-1}$ , where  $\Lambda$  is a diagonal matrix of eigenvalues and  $V$  has eigenvectors of  $A$  in the same order. This fact makes it possible to solve vector differential equations  $\dot{x} = Ax + bu$ :

$$\frac{d}{dt} x = Ax + bu \quad (1.19)$$

$$= V\Lambda V^{-1}x + bu \quad (1.20)$$

$$V^{-1} \frac{d}{dt} x = \frac{d}{dt} (V^{-1}x) = \Lambda V^{-1}x + V^{-1}bu \quad (1.21)$$

Choose a new dependent variable  $z = V^{-1}x$ , and write its differential equation using the Chain Rule.

$$\frac{d}{dt} z = \Lambda z + (V^{-1}b)u \quad (1.22)$$

Each row is a scalar differential equation which we can solve, given initial conditions, using a formula. A solution for  $x$  can be reconstructed by  $x = Vz$ .