Linear algebra

— Incomplete Notes —

DAVID GROSS
Institute for Theoretical Physics
University of Cologne

Illustration by TS, 10. Originally presented to me as the cover of a self-made booklet with fun division exercises, intended to cheer me up when I was suffering from an annoying flu.
May the Rechenspaß get a hold of you, too.

JUNE 27, 2024

Contents

Comments, discussions, and corrections are kindly acknowledged from: Lionel Dmello, Albert Marks, Kai Meinerz.

Chapter 0

Introduction

Historically, linear algebra started with the study of the properties of physical space. As an introduction, we'll have a brief look at this approach. If you find this chapter confusing, just skip ahead to Chap. 1, where we will restart from zero and introduce the theory in more general (but also more abstract) terms.

0.1 Vectors as translations in real space

In physical space, vectors can be defined geometrically, as *translations* (Fig. 0.1).

Definition 1. *A* geometric vector *is an operation that translates each point in space by a fixed distance along a fixed direction.*

Geometric vectors can be visualized by an arrow from some point to its translate.

Two geometric vectors can be combined, by performing one translation after the other. The result is again a translation. If v, w are geometric vectors, we write $v + w$ for the geometric vector that corresponds to "first translate by w, then by v". It is an empirical fact that in physical space, the order does not matter: I.e. addition of vectors is *commutative*

$$v + w = w + v \tag{1}$$

in the sense that both sides define the same translation.

> ***Remark.*** OK, strictly speaking, commutativity fails in *curved* spaces, like on the surface of earth or in space-time as described by general relativity in the presence of gravitation. But "reasonably small" translations commute to great approximation. In the mathematical idealization we are concerned with here, we may assume this property to hold exactly.

There is a "trivial translation", which displaces points by a distance of 0. We use the symbol 0 not just for the number zero, but also for the trivial vector.

Given a geometric vector v, the vector that translates points by the same distance but along the inverted direction is denoted as $-v$. Because translating a point along v and then back again along $-v$ returns it to its initial position, we have that $v + (-v) = 0$.

You may note that we have verified all the properties that turn the set of translations into an *Abelian group*, with composition given by $+$ and 0 as the neutral element.

There's an additional operation on geometric vectors: They can be "scaled". If v is a geometric vector and λ a non-negative number, then one defines λv to be the map that translates points in the same direction as v, but by a distance that's λ times the distance of v. Sending v to λv is called *scalar multiplication*.

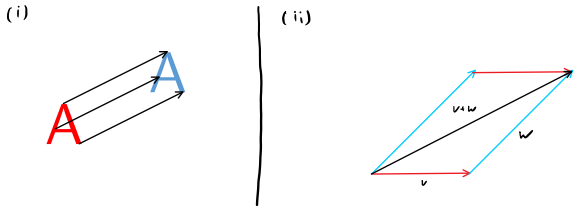


Figure 0.1: (i) A *geometric vector* is a translation of points in space. These can be visualized by an arrow from any point to its translate. (ii) When performing two translates after one another, the order does not matter.

Remark. Don't confuse scalar multiplication with *scalar products* (Sec. 1.3). A "scalar" is quantity that can be described by a single number. "Scalar multiplication" is multiplication by a scalar, (namely λ). A "scalar product" turns two vectors into a scalar.

Addition and scalar multiplication fulfill a number of geometrically obvious compatibility conditions. For example, take a geometric vector v, real numbers λ, μ, and consider $\lambda v + \mu v$. It displaces points along the direction of v, first by μ units and then by λ units. Geometrically, it is clear that this sequence is equivalent to displacing along the direction of v by $\lambda + \mu$ units. That is, we have the *distributivity property*

$$(\lambda v + \mu v) = (\lambda + \mu)v. \tag{2}$$

0.1.1 Coordinate

Calculations with vectors are usually performed in *coordinates*: Choose three geometric vectors e_1, e_2, e_3 that displace points by the same distance, but along mutually orthogonal directions. Then for any choice of real numbers x_1, x_2, x_3, the geometric vector $v = x_1e_1 + x_2e_2 + x_3e_3$ translates points by x_i units along e_i. It is another empirical fact that *any* translation in real space can be realized in this way, for a suitable choice of numbers. One says that *real space is three-dimensional*.

It is then convenient to organize the numbers that define a geometric vector into a vertically written array, called a *column vector*, like so:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The numbers are called the *components*, *elements* or *entries* of the column vector. We tend to use boldface lower-case letters for column vectors.

Note that we are now working with two strongly related but distinct objects: The geometric translation v and its representation $\boldsymbol{x}$ with respect to the *basis* $\{e_1, e_2, e_3\}$.

It is then natural to ask how the operations on geometric vectors that we discussed above are expressed in coordinates. In particular, if v, w are geometric vectors with representations $\boldsymbol{x}, \boldsymbol{y}$ respectively, what is the representation of $v + w$? Let's work it out. Plugging in the definition and using commutativity (1) and distributivity (2):

$$\begin{aligned} v + w &= x_1e_1 + x_2e_2 + x_3e_3 + y_1e_1 + y_2e_2 + y_3e_3 \\ &= (x_1e_1 + y_1e_1) + (x_2e_2 + y_2e_2) + (x_3e_3 + y_3e_3) \\ &= (x_1 + y_1)e_1 + (x_2 + y_2)e_2 + (x_3 + y_3)e_3. \end{aligned}$$

In other words: The sum $v + w$ of geometric vectors is represented in coordinates by the component-wise sum

$$\boldsymbol{x} + \boldsymbol{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}$$

of the column vectors. Likewise, in coordinates, scalar multiplication by λ corresponds to multiplying every component of the column vector by λ.

0.1.2 Euclidean length

We have distinguished between the direction and the distance of a translation. Given a column vector representation $\boldsymbol{x}$ of a translation v, it turns out that its distance (in units of the length of the basis vectors) can be computed by this expression:

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2}. \tag{3}$$

It is called the *Euclidean length*, or *norm*, of $\boldsymbol{x}$.

0.2 Who cares?

The notions introduced above are sufficient to study large swaths of geometry. But let's admit it: "translations in physical space" aren't the most inspiring objects. Who *cares*?

Fundamental reasons to care

It turns out that the concept of a "geometric vector" can be fruitfully generalized, and these generalizations turn up **e v e r y w h e r e**.

Example.

- In mechanics, quantities like velocity, acceleration, or force are naturally associated with a "direction in real space" and a magnitude. They are thus closely related to translations, and are also considered vectors in a more general sense.
- In quantum mechanics, the state of a system is described by a vector with complex-valued components. These vectors have absolutely nothing to do with physical space. In the QM course, we'll discuss how to use abstract linear algebra manipulations to extract physical predictions from these vectors.
- Now turn to analysis. Recall that the derivative $f'(x)$ of a function f can be interpreted in terms of its tangent at x. Tangents sure look like the kind of straight lines that describe translations. And, indeed, it turns out that "differentiating" turns functions that send points to points into linear maps that send vectors to vectors. You don't need to understand this now, but be assured that one *cannot* do any kind of advanced calculus without a thorough understanding of linear algebra. This really opens the floodgates, because differentials are central to large parts of math: Geometry, optimization theory, the study of symmetries...
- Functions themselves can be seen as vectors, e.g. by re-interpreting the argument as an index "$f(x) = f_x$". From this point of view, linear differential equations (Newton's, Maxwell's, Schrödinger's, Dirac's...) can all be treated using tools from linear algebra. The same is true for signal analysis on functions that represent e.g. sound waves or images. We'll see how this works when we look at the Fourier transform later in this course.

- Finally, let's mention *discrete vector spaces* that provide models for bit strings. They are used to solve combinatorial problems in in computer science, with applications e.g. for fault-tolerant communications. These objects have not been traditionally studied in physics, but this is changing rapidly through the rise of quantum information theory.

Pragmatic reasons to care

In all these examples, objects that aren't obviously vectors turn out to be vectors after all, if one generalizes the notion sufficiently. But there is a second, more subtle reason for why linear algebra permeates all of science.

Remember how in high school and undergraduate classes, you always get simple-looking exercise problems that have somewhat simple solutions? Well, here's a piece of dirty truth: That's a bit of a lie by omission perpetrated by your teachers. In reality, almost any math problem is wayyy too hard to be treated analytically ("with pen and paper") or even on a computer. We, as a species, can do basically *nothing*.

Bummer.

But you know which kind of problems we can solve? Linear algebra problems! In particular, there are extremely efficient computer algorithms for many of them.

Thus, when faced with a complicated real-world issue, it is common for scientists to hit it and shake it, until after many reformulations and simplifications, it can be shoehorned into a linear algebra problem. This is then solved. If the process fails, more often than not, we just give up and don't talk about it any more. Thereby, over time, "science" comes to encompass the subset of reality for which linear algebra methods work.

So, linear algebra is to math and science what peeling vegetables is to cooking. Not always glorious, but omnipresent and a skill you can't do without. "Who cares?" – You better do!

The good news

Linear algebra is kinda easy as far as mathematical theories go. That's the good news!

Chapter 1

Vector spaces

1.1 Definition

Wrap you head around this: The general theory does *not* actually tell you what a vector *is*. It only says how to *recognize* one.

Mathematicians call this an *axiomatic characterization*. That is to say, we will define "vector spaces" abstractly, as objects that satisfy a list of properties, or *axioms*.

Every day, mathematicians and scientists come up with novel examples that fit the definition. But the theory is so general that there's no expectation that we could ever describe all realizations.

The disadvantage of this strategy is that we'll be working with abstract objects, which may be hard to develop an intuition for. But it's a price considered worth paying in return for the wide applicability of its results. (Also, even abstract LA isn't actually *that* difficult).

> ***Remark.***
>
> - Def. 2 involves the notion of a *field* $\mathbb{F}$. Don't worry in case you momentarily forgot what that means. For physics applications, you can think of $\mathbb{F}$ as a placeholder for $\mathbb{R}$ or $\mathbb{C}$.
> - While Def. 2 is central to the theory, and you should definitely get a feeling for it, there's no need to memorize it!
> - In fact, here's the tl;dr. *A* vector space *is a collection of objects that can be* added *and* scaled. The rest are just natural compatibility conditions.

With these preparations, here's he central definition of linear algebra:

Definition 2. *Let $\mathbb{F}$ be a field. A* vector space over $\mathbb{F}$ *is a set V on which two operations are defined:*

1. Addition*, which assigns to each pair $v, w \in V$ an element $v + w \in V$ and which turns V into an Abelian group.*
2. Scalar multiplication: *For $\lambda \in \mathbb{F}$ and $v \in V$, there is an element $\lambda v \in V$.*

The operations must fulfill the following compatibility conditions: For all $\alpha, \beta \in \mathbb{F}$ and

all $v, w \in V$, it must hold that

$$
\begin{aligned}
1v &= v && \textit{(multiplicative neutral element leaves vectors invariant)}\\
(\alpha\beta)v &= \alpha(\beta v) && \textit{(associativity of scalar multiplication)}\\
(\alpha+\beta)v &= \alpha v + \beta v && \textit{(distributivity with respect to addition in the field)}\\
\alpha(v+w) &= \alpha v + \alpha w && \textit{(distributivity with respect to addition of vectors).}
\end{aligned}
$$

Example.

- For any field $\mathbb{F}$ and any $n \in \mathbb{N}$, the set of column vectors with n entries from $\mathbb{F}$ forms a vector space, denoted by $\mathbb{F}^n$. Addition and scalar multiplication are defined component-wise. We have already met the special case $\mathbb{R}^3$. Another special case is $\mathbb{F}^1 = \mathbb{F}$. That's right: a field is a vector space over itself.
- The set of all complex-valued functions on $\mathbb{R}$, with addition and scalar multiplication defined point-wise, is a vector space over $\mathbb{C}$. "Point-wise" means that for functions f, g, we define their sum $f + g$ by $(f + g)(x) := f(x) + g(x)$. Likewise, $(\lambda f)(x) := \lambda f(x)$.
- More generally, if M is any set and $\mathbb{F}$ any field, then the set of functions $M \to \mathbb{F}$ is a vector space under point-wise addition and scalar multiplication.
  - Taking $M = \{1, \ldots, n\}$, we recover the set of column vectors with n components.
  - For $M = \mathbb{N}$, the construction gives the vector space of infinite sequences with values in $\mathbb{F}$.
- A rectangular array of numbers with m rows and n columns is an $m \times n$-*matrix*. Here's a 2×3-matrix over $\mathbb{R}$:

  $$\boldsymbol{M} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

  The set $\mathbb{F}^{m\times n}$ of matrices with m rows, n columns, and entries in a field $\mathbb{F}$ is a vector space under element-wise addition and scalar multiplication. We will not distinguish between an $m\times 1$-matrix and a column vector with m components. Likewise, $1\times n$-matrices are called *row vectors*.
- The set of vectors (shown in Fig. 1.1)

  $$\underline{1}^{\perp} = \Big\{ \boldsymbol{x} \in \mathbb{R}^3 \,\Big|\, \sum_{i=1}^{3} x_i = 0 \Big\}$$

  whose components sum to 0 forms a vector space. (We'll come back to it as a simple example of a space without a "canonical basis").
- The space ℓ_2 ("ell two") of *square-summable complex sequences* is

  $$\ell_2 = \Big\{ x : \mathbb{N} \to \mathbb{C} \,\Big|\, \|x\|^2 := \sum_{i=1}^{\infty} |x_i|^2 \leq \infty \Big\}.$$

  The space plays a central role in quantum mechanics. (The condition that the series $\|x\|^2$ converges to a finite number allows one to define a probability distribution by setting $p_i = |x_i|^2/\|x\|^2$. We'll give a physical interpretation in the QM lecture.) Verifying that component-wise operations turn ℓ_2 into a vector space isn't completely trivial. One has to show that for $x, y \in \ell_2$, the sum $\sum_i |x_i + y_i|^2$ converges to a finite number.

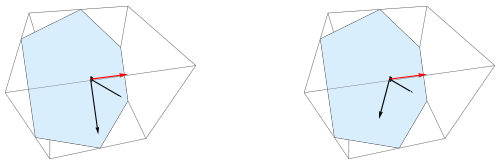


Figure 1.1: The space $\underline{1}^\perp$ of vectors in 3D whose components sum to 0. The red arrow is the vector $(1,1,1)^t$. Left: The $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ basis. Right: The orthonormal basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.

- The set $\mathbb{F}_2 = \{0, 1\}$ is a field with addition and multiplication defined "modulo 2":

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}, \qquad \begin{array}{c|cc} \times & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}.$$

  Column vectors $\mathbb{F}_2^n$ over $\mathbb{F}_2$ are a finite set that form a vector space: $|\mathbb{F}_2^n| = 2^n$. Elements $x \in \mathbb{F}_2^n$ can be interpreted as bit strings of length n. As such, they are important in computer science and information theory. Finite vector spaces used to not matter much in physics, but this is changing with the rise of quantum information theory.

Remark. The high school definition of a vector is "a quantity with a direction and a length". But Def. 2 makes no mention of "length". That's because there are vector spaces for which that notion does not make sense. Example: In Hamiltonian mechanics, the state of a point particle is given in terms of its position and momentum. Collect this information in a column vector $\binom{x}{p}$. If you naively try to compute its Euclidean length as $\sqrt{x^2 + p^2}$, you run into trouble, as x^2 and p^2 have different units and can't be added. Remember: *Not every vector has a length*!

Computer algebra systems have extensive support for LA. We will often show examples in the Julia language. It uses square brackets for arrays and semicolons to separate rows:

```
julia> M = [1 2 3; 4 5 6]  # A matrix
2×3 Matrix{Int64}:
 1  2  3
 4  5  6

julia> v = [1; 2; 3] # A column vector
3-element Vector{Int64}:
 1
 2
 3

julia> 2M+[1 1 1; 0 0 0] # Scalar mult., addition
2×3 Matrix{Int64}:
 3   5   7
 8  10  12
```

We can now answer the question "What is a vector?" in the abstract theory. An *abstract vector* is an element of a vector space. Not more, not less.

Remark. Unfortunately, instead of "column vector", people commonly say "vector". And instead of "abstract vector", they *also* say "vector". Thus if person A claims

- "A function on the real line *is not* a vector",

then they are right! (Because they mean "not a column vector"). And if person B states

- "A function on the real line *is* a vector",

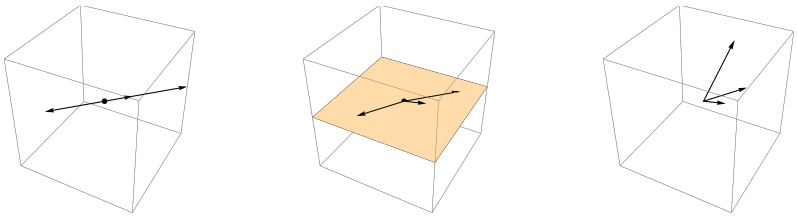


Figure 1.2: The span of three non-zero vectors in $\mathbb{R}^3$ can be a line, a plane, or the entire space. The latter happens if none of the vectors is a linear combination of the two others.

then they are also right! (Because they mean "abstract vector"). Often, person A and person B are the same person.

Mathematical language is pretty precise, but not as precise as e.g. programming languages. So there remains a lot of ambiguity that you will have to learn to deal with. In fact, we will from now on drop the clarifying attributes ourselves. To prepare you for math "as found in the wild".

1.2 Basic constructions

In the following, let V be a vector space over a field $\mathbb{F}$.

1.2.1 Linear combinations

Take a finite set of vectors $v_1, \dots, v_r \in V$, and scalars $x_1, \dots, x_r \in \mathbb{F}$. The sum $\sum_{i=1}^r x_i v_i$ is called a *linear combination of the* v_i.

For a set $S \subset V$ of a vectors, the set $\mathrm{span}(S)$ of all linear combinations of vectors from S is the *span* or *linear hull* of S. The set S is *complete* if its linear hull equals V.

Example.

- For examples in $\mathbb{R}^3$, see Fig. 1.2.
- Choose a field $\mathbb{F}$. A *polynomial* is a function of the form

$$p : \mathbb{F} \to \mathbb{F}, \qquad p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0, \tag{1.1}$$

  for *coefficients* $c_i \in \mathbb{F}$. Its *degree* is the largest power of x with a non-zero coefficient. Write $\mathcal{P}_n(\mathbb{F})$ for the set of all polynomials with degree $\leq n$ and $\mathcal{P}(\mathbb{F}) = \bigcup_{n=0}^{\infty} \mathcal{P}_n(\mathbb{F})$ for all polynomials over $\mathbb{F}$. Then $\mathcal{P}(\mathbb{F})$ is the linear hull of the *monomials* $\{x^n\}_{n \in \mathbb{N}}$.

1.2.2 Subspaces

A subset W of a vector space V is a *subspace* if any linear combination of elements of W lies again in W.

Example.

- For subspaces in $\mathbb{R}^3$, see Fig. 1.3
- For any $n \in \mathbb{N}$, the set $\mathcal{P}_n(\mathbb{F})$ of polynomials of degree $\leq n$ forms a subspace of $\mathcal{P}(\mathbb{F})$. Non-example: The set of polynomials of degree *exactly* n does not form a subspace!
- The solutions to the *wave equation*

$$\left(\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_z^2\right) u(t, x, y, z) = 0$$

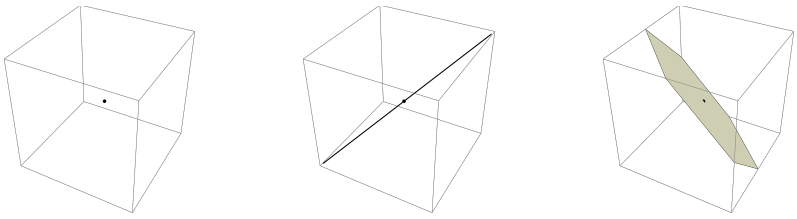


Figure 1.3: A subspace in $\mathbb{R}^3$ can be have dimension 0 (a point), dimension 1 (a line), dimension 2 (a plane), or dimension 3 (the entire space, not shown).

form a subspace of the space of twice-differentiable functions on $\mathbb{R}$. In the context of wave equations and of quantum mechanics, linear combinations are also called *superpositions* and the statement just made is the *superposition principle*.

Exercise.

1. The intersection of two subspaces is a subspace.
2. The union of two subspaces is generally *not* a subspace!
3. To show that $W \subset V$ is a subspace, it suffices to verify that for all $v, w \in W$ and $\lambda \in \mathbb{F}$ it holds that $v + w \in W$ ("W is closed under addition") and $\lambda v \in W$ ("W is closed under scalar multiplication").
4. For any set $S \subset V$ of vectors, $\mathrm{span}(S)$ is a subspace of V.
5. A subspace $W \subset V$ is a vector space by itself.

1.2.3 Linear independence, bases, dimension

Let $S \subset V$ be a set of vectors. The elements of S are *linearly independent* (LI) if no element of S can be written as a linear combination of the other elements of S. Otherwise, they are *linearly dependent*.

Exercise: S is linearly independent if and only if

$$x_1 v_1 + \cdots + x_r v_r = 0 \qquad v_i \in S, x_i \in \mathbb{F} \tag{1.2}$$

implies $x_1 = \cdots = x_r = 0$. Phrased differently: S is linearly independent if only trivial linear combinations are zero.

A set $\mathcal{B} \subset V$ of vectors is a *basis* if $\mathcal{B}$ is complete and linearly independent.

The space V is *n-dimensional* if it has a basis with $n \in \mathbb{N}$ elements. Otherwise, it is *infinite-dimensional*.

Example.

- Let $\boldsymbol{e}_i$ be the column vector whose i-th element is 1 and all other elements are 0. (The total number of elements must be inferred from context). Then $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is a basis for $\mathbb{F}^n$. It is called the *standard basis*. The space $\mathbb{F}^n$ is therefore n-dimensional.
- The set $\{1, i\} \subset \mathbb{C}$ forms a basis of $V = \mathbb{C}$, interpreted as a vector space over $\mathbb{F} = \mathbb{R}$. *Attention*: The set $\{1\} \subset \mathbb{C}$ forms a basis of $V = \mathbb{C}$, interpreted as a vector space over $\mathbb{F} = \mathbb{C}$. The same set can thus be a vector space in different ways! Oof.

- Consider the space $\underline{1}^\perp$ of vectors $\boldsymbol{v} \in \mathbb{R}^3$ with $\sum_{i=1}^3 v_i = 0$. Here are two bases:

$$\boldsymbol{v}_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}; \qquad \boldsymbol{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \boldsymbol{w}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}. \quad (1.3)$$

  The second is more complicated, but has the advantage of being *orthonormal* (Sec. 1.3.5).

- A basis of the vector space of 2×2-matrices over $\mathbb{C}$ is given by the *Pauli matrices*

$$\boldsymbol{\sigma}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \boldsymbol{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.4)$$

  They are very important in the quantum-mechanical theory of spin. And for fault-tolerant quantum computers. And for the representation theory of relativistic space-time transformations. Did I mention the theory of quaternions, with applications e.g. in robotics and attitude control for spacecraft? Look, the Pauli matrices are kind of important.

- The monomials $\{x^n \,|\, n \in \mathbb{N}\}$ form a basis for the space of polynomial function $\mathcal{P}(\mathbb{F})$. Completeness is clear from the definition of $\mathcal{P}(\mathbb{F})$. Independence is less obvious. Therefore, $\mathcal{P}(\mathbb{F})$ is an infinite-dimensional space.

- A **non-example**: f is *analytic* (around 0) if it can be represented by a Taylor series

$$f(x) = \sum_{k=0}^{\infty} c_k x^k, \qquad c_k = \frac{1}{k!} \frac{\partial^k}{(\partial k)^k} f(0).$$

  One could be forgiven to conclude that the monomials are a basis for the space of analytic functions. This would be wrong, though, because linear combinations are defined in terms of *finite* sums, whereas the series is defined as a limit involving infinitely many terms. We'll later generalize the notion of a basis to cover infinite series, too.

Here are two important statements about bases, which we will not prove.

- [Dimension is unique] A priori, it could be that there is a vector space V that has two bases with a different number of elements. In $\mathbb{R}^3$, it is geometrically intuitive that this problem does not arise: A line is a line and a plane a plane, i.e. "dimension" is an intrinsic property of a space. And fortunately, it is not difficult to prove that uniqueness of dimension does hold for all vector spaces. (FK 14.6.4).

- [Basis extension theorem] The *basis extension theorem* says that any set of linearly independent vectors can be extended to a basis. In other words, bases can be built "iteratively": Starting with the empty set, just add linearly independent vectors until you arrive at a collection that spans the space. It cannot happen that this process runs into a "dead-end" where one would be forced to backtrack and make different choices. Again, this is intuitive in $\mathbb{R}^3$. (FK 14.6.5).

Remark. [Every vector space has a basis (trust me, bro!)] Many textbooks say "One can prove that any vector space has a basis." Hm, OK. That's not wrong. But, boy, does it underplay the stakes. Trying to decide this statement leads to a rabbit hole that ends in meta-mathematical questions of what "truth" and "provability" mean in the first place. In the early 20th century, the titans of math fought bitter fights about this and related questions. If you want to waste some time in the lecture, ask me about it. But, yeah, whatever, I guess every vector space has a basis.

1.2.4 Coordinate representation of vectors

Bases allow us to work with arbitrary vector spaces in terms of column vectors.

Let $\mathcal{B}$ be a basis. We assume that there is a discrete *index set* I such that the vectors in the basis can be labeled by elements $i \in I$. To indicate this, we'll write $\mathcal{B} = \{b_i\}_{i\in I}$ or $\mathcal{B} = \{b_i\}_i$ or just $\mathcal{B} = \{b_i\}$. Usually, $I = \{1, \dots, n\}$ or $I = \mathbb{N}$.

By the completeness condition in the definition of basis, every vector $v \in V$ can be written as a linear combination

$$v = \sum_i x_i b_i, \qquad x_i \in \mathbb{F}.$$

This is called a *basis expansion* of v. The numbers x_i are the *expansion coefficients* or *coordinates* of v w.r.t. $\mathcal{B}$.

The linear independence condition guarantees that expansions are unique:

$$\sum_i x_i b_i = \sum_i y_i b_i \quad \Rightarrow \quad x_i = y_i.$$

Proof. Compute:

$$0 = \sum_i x_i b_i - \sum_i y_i b_i = \sum_i (x_i - y_i) b_i.$$

By Eq. (1.2), $x_i - y_i = 0$. □

Assume now for simplicity that $\mathcal{B} = \{b_1, \dots, b_n\}$, i.e. that V is has dimension $n < \infty$. Then any $v \in V$ can be mapped to a unique column vector

$$v = \sum_{i=1}^{n} x_i b_i \mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} =: \boldsymbol{\phi}^{\mathcal{B}}(v) \tag{1.5}$$

composed of the coordinates of v w.r.t. $\mathcal{B}$.

Arguing as in Sec. 0.1.1, we see that the coordinates of a sum are the sum of the coordinates, and the coordinates of a scalar multiple is the scalar multiple of the coordinates:

$$\boldsymbol{\phi}^{\mathcal{B}}(v + w) = \boldsymbol{\phi}^{\mathcal{B}}(v) + \boldsymbol{\phi}^{\mathcal{B}}(w), \qquad \boldsymbol{\phi}^{\mathcal{B}}(\lambda v) = \lambda \boldsymbol{\phi}^{\mathcal{B}}(v).$$

A one-to-one map between two vector spaces that commutes with addition and scalar multiplication is called a *(linear) isomorphism*. Two vector spaces are *isomorphic* if there is a linear isomorphism between the two.

Because linear algebra is concerned only with the effects of addition and scalar multiplication, a linear isomorphism ϕ preserves all "linear algebraic properties". For example:

- A set $S \subset V$ of vectors is complete or linearly independent if and only if $\phi(S)$ is.
- Two isomorphic spaces have the same dimension.

1.3 Scalar products

In some vector spaces, it makes sense to assign a length to every vector, and an angle to every pair. Such geometric notions are derived from a *scalar product*, introduced next.

1.3.1 Euclidean spaces

In the special case of $V = \mathbb{R}^3$, the *standard scalar product* is

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_1 w_1 + v_2 w_2 + v_3 w_3.$$

The *Euclidean length* (or 2-*norm* or the *norm induced by the scalar product*) is

$$\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} \tag{1.6}$$

and the *angle* ϕ enclosed by two vectors $\boldsymbol{v}, \boldsymbol{w}$ is (Fig. 1.4)

$$\phi = \arccos\left(\frac{\langle \boldsymbol{v}, \boldsymbol{w} \rangle}{\|\boldsymbol{v}\| \, \|\boldsymbol{w}\|} \right). \tag{1.7}$$

We now face the same situation as in Sec. 1.1: One could develop a concrete and intuitive theory of scalar products in $\mathbb{R}^3$. But soon, we will need to assign notions of "length" and "angles" to more general vectors, e.g. to elements of function spaces that appear in QM. Thus, as before, we will take the more abstract axiomatic route.

Definition 3 (Scalar product over $\mathbb{R}$)**.** *Let V be a vector space over $\mathbb{R}$. A* scalar product *is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that is*

1. bilinear:

$$\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \langle v, w \rangle$$
$$\langle u, v + \lambda w \rangle = \langle u, v \rangle + \lambda \langle u, w \rangle,$$

2. symmetric:

$$\langle u, v \rangle = \langle v, u \rangle,$$

3. and positive definite:

$$\langle u, u \rangle > 0$$

for all $u \neq 0$.

A real vector space with a scalar product is called a Euclidean space.

Example.

- The standard inner product on $\mathbb{R}^n$:

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_1 w_1 + \cdots + v_n w_n. \tag{1.8}$$

- For an interval $[a, b] \subset \mathbb{R}$, let $C^0([a, b], \mathbb{R})$ be the real vector space of continuous functions $[a, b] \to \mathbb{R}$. Then

$$\langle f, g \rangle = \int_a^b f(x) g(x) \, \mathrm{d}x \tag{1.9}$$

is a scalar product. Continuity implies that the integral is well-defined and that the scalar product is definite (why?).

- Consider a collection of N classical point particles of masses m_i moving in one dimension with velocity v_i. The kinetic energy is given by in terms of their momenta $p_i = m_i v_i$ by

$$K = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2m_i} p_i^2,$$

which is the squared norm $(\boldsymbol{p}, \boldsymbol{p})_m$ with respect to the scalar product

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle_m := \frac{1}{2m_1} p_1 q_1 + \cdots + \frac{1}{2m_N} p_N q_N.$$

- Consider a 3D rigid body with mass distribution $\rho(\boldsymbol{x})$. The *inertia tensor* is

$$I_{kl} = \int \rho(\boldsymbol{x}) \left(\|\boldsymbol{x}\|^2 \delta_{ij} - x_i x_j \right) \, \mathrm{d}^3 \boldsymbol{x}. \tag{1.10}$$

It defines a scalar product via

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle_I := \sum_{kl} v_k I_{kl} w_l. \tag{1.11}$$

Exercise: Show that (1.11) is positive definite. (Hint: Use Cauchy-Schwartz).

Physical interpretation: The rotational motion of a rigid body can be described in terms of a vectorial angular velocity $\boldsymbol{\omega} \in \mathbb{R}^3$. One can then show that the kinetic energy of the rotation is given by $K_{\mathrm{rot}} = \langle \boldsymbol{\omega}, \boldsymbol{\omega} \rangle_I$.

- The standard Gaussian probability density is $\rho(x) = (2\pi)^{-1/2} e^{-\frac{1}{2}x^2}$. Given a function $f(x)$, its *expectation value* with respect to $\rho(x)$ is

$$\mathbb{E}[f] = \int f(x) \rho(x) \, \mathrm{d}x.$$

On the space of continuous functions with finite expectation values,

$$\langle f, g \rangle = \mathbb{E}[fg]$$

is a scalar product. It appears in probability theory, and in the context of Hermite polynomials, with applications to the quantum theory of the harmonic oscillator.

1.3.2 Hermitian spaces

In quantum mechanics, we'll work with complex vector spaces and interpret the squared-norm of a vector as a probability. For this to make sense, the norm needs to be non-negative. But because $i^2 = -1$, a naive extension of Eq. (1.8) does not define a positive definite form. But because for any $z \in \mathbb{C}$, it is true that $\bar{z} z = |z|^2 \geq 0$, it follows that

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \sum_{i=1}^{n} \bar{v}_i w_i \tag{1.12}$$

does fulfill $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$.

Here's the general definition for a scalar product on a complex vector space:

Definition 4 (Scalar product over $\mathbb{C}$)**.** *Let V be a vector space over $\mathbb{C}$. A* scalar product *is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ that is*

1. sesquilinear:

$$\langle u + \lambda v, w\rangle = \langle u, w) + \bar{\lambda}(v, w\rangle$$
$$\langle u, v + \lambda w\rangle = \langle u, v) + \lambda(u, w\rangle,$$

2. conjugate symmetric:

$$\langle u, v\rangle = \overline{\langle v, u\rangle},$$

3. and positive definite:

$$\langle u, u\rangle > 0$$

for all $u \neq 0$.

A complex vector space with a scalar product is called a Hermitian space.

Remark.

- "Sesqui" in "sesquilinear" apparently means one-and-a-half.
- In physics, Hermitian scalar products are defined to be anti-linear in the first, and linear in the second argument. In math, it's the other way round. Oh dear.
- If $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^n$ happen to have real coefficients, the sesquilinear scalar product (1.12) reduces to the bilinear scalar product (1.8). The two cases are thus closely related, and it shouldn't be surprising that we can treat their properties simultaneously.
- You may have heard the term *Hilbert space*. A Hilbert space is a Hermitian space with an extra property ("completeness") – we'll get there later.

Example. A Hermitian scalar product on continuous functions $[a, b] \to \mathbb{C}$ is

$$\langle f, g\rangle = \int_a^b \bar{f}(x)g(x)\,\mathrm{d}x.$$

In the following, V is a Euclidean or a Hermitian space.

1.3.3 Inequalities

Two important inequalities are (Fig. 1.4):

$$|\langle v, w\rangle| \leq \|v\|\,\|w\| \qquad \text{Cauchy-Schwartz inequality,}$$
$$\|v + w\| \leq \|v\| + \|w\| \qquad \text{Triangle inequality.}$$

Cauchy-Schwartz is tight (i.e. "holds with equality") iff $v = \lambda w$ with $\lambda \in \mathbb{R}$. The triangle inequality is tight iff $v = \lambda w$ with $\lambda > 0$.

By the Cauchy-Schwartz inequality, $\left|\frac{\langle v,w\rangle}{\|v\|\,\|w\|}\right| \leq 1$, so that the inverse cosine in (1.7) is well-defined and can be used to define the notion of an *angle* in general (Fig. 1.4).

1.3.4 Projection onto a vector

A vector $v \in V$ is *normalized* if $\|v\| = 1$. For any $v \neq 0$, the vector $\frac{v}{\|v\|}$ is normalized.

For a normalized v, the *(orthogonal) projection onto* v is

$$P_v : w \mapsto v\,\langle v, w\rangle$$

(Fig. 1.4). Its length equals the inner product $\langle v, w\rangle$.

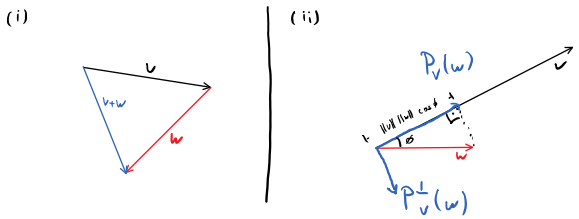


Figure 1.4: (i) The triangle inequality says that a direct path is the shortest connection between two points. (ii) The geometry of two vectors v, w, the decomposition of w into components parallel $P_v(w)$ and orthogonal $P_v^\perp(w)$ to v, as well as the relation between the angle ϕ and the length of the projection.

1.3.5 Orthogonality

Two non-zero vectors $v, w \in V$ are *orthogonal* if $\langle v, w\rangle = 0$. In this case, we write $v \perp w$.

A set $\{b_i\}$ of vectors is an *orthonormal system* if its elements are normalized and mutually orthogonal, i.e. if

$$\langle b_i, b_j\rangle = \delta_{ij}.$$

An orthonormal system that's also a basis is called an *orthonormal basis* or *ONB*.

ONBs are particularly easy to work with. Indeed, let $\{b_i\}$ be an ONB and consider a basis expansion

$$v = \sum_i x_i b_i.$$

In general, given a vector v and a basis $\{b_i\}$, it isn't trivial to find the expansion coefficients x_i. But for an ONB, the coefficients are just given by inner products with the basis vectors:

$$\langle b_i, v\rangle = \sum_j x_j \langle b_i, b_j\rangle = \sum_j x_j \delta_{ij} = x_i \quad \Rightarrow \quad v = \sum_i b_i \langle b_i, v\rangle = \sum_i P_{b_i}(v).$$

You will be using these formulas *a lot*!

Examples & exercises.

- The standard basis (1.12) is an ONB for $\mathbb{R}^n$ and $\mathbb{C}^n$.
- The set $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ defined in Eq. (1.3) is an ONB for $\underline{1}^\perp$.
- Consider the space $C^0([-\pi, \pi], \mathbb{R})$ with the scalar product defined in (1.9). The set

  $$\mathcal{B} = \left\{\frac{\cos(kx)}{\sqrt{\pi}}\right\}_{k=0}^{\infty} \cup \left\{\frac{\sin(kx)}{\sqrt{\pi}}\right\}_{k=1}^{\infty}$$

  forms an orthonormal system.

  We will see later that the system is in fact also complete in some sense. The expansion coefficients of a function $f(x)$ w.r.t. this ONB is a variant of the *Fourier transform*.
- Show that an orthonormal system is automatically linearly independent.
- Show *Parseval's identity*, which generalizes the Pythagorean theorem. It says that for an orthonormal system $\{b_i\}$,

  $$v = \sum_i x_i b_i \qquad \Rightarrow \qquad \|v\|^2 = \sum_i |x_i|^2.$$

For a normalized v, define

$$P_v^\perp : w \mapsto w - P_v(w). \tag{1.13}$$

It maps w to a vector that is orthogonal to v:

$$\langle v, P_v^\perp(w)\rangle = \langle v, w\rangle - \big\langle v, (v\,\langle v, w\rangle)\big\rangle = \langle v, w\rangle - \langle v, w\rangle\langle v, v\rangle = 0.$$

By definition, $w = P_v(w) + P_v^\perp(w)$. We have thus decomposed w into a component $P_v(w)$ parallel to v, and a component $P_v^\perp(w)$ orthogonal to v (Fig. 1.4).

1.3.6 Gram-Schmidt algorithm

ONBs are nice to work with. Thus, given a set of linearly independent vectors $S = \{v_i\}$, it'd be good if we could convert them into an ONB $\mathcal{B} = \{b_i\}$ for $\mathrm{span}(\{v_i\})$.

This is achieved by the *Gram-Schmidt* algorithm. You will very rarely perform a Gram-Schmidt process manually. But it's still important to have seen it once: It is related to the *QR decomposition*, a central tool in numerical linear algebra. It is also used in the construction of orthogonal polynomials that play an important role in the quantum theory of orbital angular momentum and the excited states of the harmonic oscillator.

From $\{v_i\}$, we will iteratively build an orthogonal system $\{w_i\}$ and an orthonormal system $\{b_i\}$. The first step is special:

- Set

  $$w_1 = v_1, \qquad b_1 = \frac{w_1}{\|w_1\|}.$$

  Then b_1 is normalized and, obviously,

  $$\mathrm{span}(\{w_1\}) = \mathrm{span}(\{b_1\}) = \mathrm{span}(\{v_1\}).$$

The meat is in the following step:

- We need w_2 to be orthogonal to w_1, so choose

  $$w_2 = v_2 - b_1\langle b_1, v_2\rangle.$$

  By (1.13), w_2 a linear combination of v_1, v_2 with non-zero coefficients. By linear independence, it is non-zero. Therefore, it can be normalized, and we set

  $$b_2 = \frac{w_2}{\|w_2\|}.$$

  At the end of this step:

  $$\{w_1, w_2\} \text{ is orthogonal,}$$
  $$\{b_1, b_2\} \text{ is orthonormal,}$$

  and, because we can express v_1, v_2 as a linear combination of $\{w_1, w_2\}$ or $\{b_1, b_2\}$,

  $$\mathrm{span}(\{w_1, w_2\}) = \mathrm{span}(\{b_1, b_2\}) = \mathrm{span}(\{v_1, v_2\}).$$

Gram-Schmidt just repeats the previous construction for the remainder of the v_i:

- At the i-th step, set

$$w_i = v_i - \sum_{j=1}^{i-1} b_j \langle b_j, v_i \rangle \qquad \text{“orthogonalize”},$$
$$b_i = \frac{w_i}{\|w_i\|} \qquad \text{“normalize”}.$$

Arguing as before, one then shows by induction that

$$\begin{gathered} \{w_1, \ldots, w_i\} \text{ is orthogonal}, \\ \{b_1, \ldots, b_i\} \text{ is orthonormal}, \\ \operatorname{span}(\{w_1, \ldots, w_i\}) = \operatorname{span}(\{b_1, \ldots, b_i\}) = \operatorname{span}(\{v_1, \ldots, v_i\}). \end{gathered}$$

Chapter 2

Linear maps

A map $L : V \to W$ between two vector spaces V, W over the same field $\mathbb{F}$ is *linear* if it commutes with vector addition and scalar multiplication

$$L(u + \lambda v) = L(u) + \lambda L(v) \qquad u, v \in V, \lambda \in \mathbb{F}.$$

For linear maps, it is common to omit the parentheses around the argument: $Lv = L(v)$. Linear maps are also called *(linear) operators*. In math, this term is associated with maps on infinite-dimensional vector spaces, but physicists aren't as discerning.

Example.

1. For examples on $V = W = \mathbb{R}^2$, see Fig. **??**.
2. The *identity* map $\mathbb{1} : v \mapsto v$ is linear, as is the zero map $0 : v \mapsto 0$.
3. If V is a Euclidean space and $b \in V$, then $v \mapsto \langle b, v\rangle$ is linear.
4. If V is a Hermitian space and $b \in V$, then $v \mapsto \langle b, v\rangle$ is linear, but the function $F : v \mapsto \langle v, b\rangle$ is *not* linear, because it does not commute with scalar multiplication:
$$F(\lambda v) = \langle \lambda v, b\rangle = \bar{\lambda}\langle v, b\rangle = \bar{\lambda} F(v).$$
5. Related, $f \mapsto \int_a^b f(x)\,\mathrm{d}x$ is linear on the space of integrable functions.
6. In a vector space of functions, every element x of the domain of definition defines a linear function $f \mapsto f(x)$, called the *evaluation map*.
7. Let $p = \sum_{k=0}^{n} c_k x^k$ be a polynomial. Then the *multiplication operator* M_p that sends a function f to $M_p f$ where $(M_p f)(x) = p(x) f(x)$ is linear. With p, we an also associate the differential operator
$$D_p : f \mapsto \sum_{k=0}^{n} (-i)^k \frac{\partial^k}{(\partial x)^k} f$$
which is a linear function on the vector space of n-times differentiable functions f. E.g., the squared norm $p = x^2 + y^2 + z^2$ gives rise to $D_p = -\Delta$, the negative *Laplacian*, which describes kinetic energy in QM. What's up with the funky factor of $(-i)^k$? We'll see that it turns M_p and D_p into each others' Fourier transforms.
8. Let $\boldsymbol{R}_\phi$ be the rotation matrix of (2.5). Define U_ϕ on functions $\mathbb{R}^2 \to \mathbb{C}$ by
$$(U_\phi f)(\boldsymbol{x}) = f(\boldsymbol{R}_\phi^{-1}\boldsymbol{x}).$$
The linear map U_ϕ is central to the quantum theory of orbital angular momentum.
9. Some maps that are *not* linear: Translations $v \mapsto v + w$, the norm $v \mapsto \|v\|^2$, or complex conjugation of the components of column vectors $\boldsymbol{v} \mapsto \bar{\boldsymbol{v}}$.

2.1 Matrix representations

Let $\mathcal{A} = \{v_j\}$ be a basis of V. Then

$$L(v) = L\Big(\sum_j x_j v_j\Big) = \sum_j x_j L(v_j). \tag{2.1}$$

It follows that a linear map is determined once we know its values $L(v_j)$ on a basis. Now choose a basis $\mathcal{B} = \{w_i\}$ of W. Then there are expansion coefficients $a_{ij} \in \mathbb{F}$ such that

$$L(v_j) = \sum_i a_{ij}\, w_i. \tag{2.2}$$

If $n = \dim V, m = \dim W$ are finite, the a_{ij} are usually presented as an $m \times n$ *matrix*

$$\phi^{\mathcal{B}}_{\mathcal{A}}(L) := \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \tag{2.3}$$

The numbers a_{ij} are called the *matrix elements* of L (with respect to the bases $\mathcal{A}, \mathcal{B}$). The notation (a_{ij}) refers to the entire matrix in (2.7). We will also use bold upper-case symbols for matrices, as in $(a_{ij}) = \boldsymbol{A}$.

Remark.

- Confusion alert! In 2D geometry, we specify the horizontal coordinate first, "(x, y)". Matrix elements are indexed by their *vertical* coordinate first!
- If $W = V$, we usually choose $\mathcal{B} = \mathcal{A}$ and sometimes write $\phi_{\mathcal{A}}$ instead of $\phi^{\mathcal{A}}_{\mathcal{A}}$.
- One says that (2.1) *linearly extends* L from $\{v_j\}$ to all of V.

Example.

- The zero map $0 : v \mapsto 0$ has matrix representation

  $$\phi^{\mathcal{B}}_{\mathcal{A}}(0) = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = 0 \in \mathbb{F}^{m\times n}$$

  with respect to any choice of bases $\mathcal{A}, \mathcal{B}$.

- For $V = W$, the identity map has matrix representation

  $$\phi^{\mathcal{A}}_{\mathcal{A}}(\mathbb{1}) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} = \mathbb{1}_n \in \mathbb{F}^{n\times n} \tag{2.4}$$

  with respect to any choice of basis $\mathcal{A}$. In other words, the matrix elements are $a_{ij} = \delta_{ij}$.

- Let $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ be the standard basis of $\mathbb{R}^2$. The matrix representation $\boldsymbol{R}_\phi$ of a rotation R_ϕ by an angle ϕ is

  $$\boldsymbol{R}_\phi = \phi^{\mathcal{A}}_{\mathcal{A}}(R_\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \tag{2.5}$$

The representation of a reflection S_x about the x-axis is

$$\boldsymbol{\phi}^{\mathcal{A}}_{\mathcal{A}}(S_x) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Another basis is $\mathcal{B} = \{\boldsymbol{f}_1, \boldsymbol{f}_2\}$ where $\boldsymbol{f}_1 = \boldsymbol{e}_1 + \boldsymbol{e}_2$, $\boldsymbol{f}_2 = \boldsymbol{e}_1 - \boldsymbol{e}_2$. Then

$$\boldsymbol{\phi}^{\mathcal{B}}_{\mathcal{B}}(S_x) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

- Let $\mathcal{A} = \{x^0, x^1, x^2, \dots\}$ be the basis of monomials for the space of polynomial functions. On the subspace $\mathcal{P}_3$ of polynomials of degree at most 3, the differential operator ∂_x has matrix representation

$$\boldsymbol{\phi}^{\mathcal{A}}_{\mathcal{A}}(\partial_x) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.6}$$

In Julia, the element a_{ij} of an array `A` is indexed as `A[i, j]`. Entire rows and columns can be accessed using the *slice notation*, where a colon represents all values an index can take:

```
julia> A = [1 2 3; 4 5 6]
2×3 Matrix{Int64}:
 1  2  3
 4  5  6

julia> A[1,1] # Julia indices start at 1 (unlike Python / C)
1

julia> A[:,1] # First column
2-element Vector{Int64}:
 1
 4

julia> A[1,:] # First row
3-element Vector{Int64}:
 1
 2
 3
```

2.1.1 Matrix-vector multiplication

Choose bases $\mathcal{A} = \{v_i\}$ for V and $\mathcal{B} = \{w_i\}$ for W. Given $v \in V$, what is the relation between the coordinate representations

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{\phi}^{\mathcal{A}}(v) && \text{of the vector } v, \\ \boldsymbol{A} &= \boldsymbol{\phi}^{\mathcal{B}}_{\mathcal{A}}(L) && \text{of the linear map } L, \text{ and} \\ \boldsymbol{y} &= \boldsymbol{\phi}^{\mathcal{B}}(Lv) && \text{of the image } Lv? \end{aligned}$$

Using (2.2),

$$L(v) = L\Big(\sum_j x_j v_j\Big) = \sum_i \Big(\sum_j a_{ij} x_j\Big) w_i \quad \Rightarrow \quad y_i = \sum_j a_{ij} x_j. \tag{2.7}$$

This is the simple, but enormously important *matrix-vector multiplication* formula. The sum is called a *contraction* of the two indices labeled by j's. We abbreviate Eq. (2.7) as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

I'm afraid you'll have to memorize those:

The i-th element of $\boldsymbol{Ax}$ is the contraction of the i-th row of $\boldsymbol{A}$ with $\boldsymbol{x}$. (2.8)

The j-th column of a matrix is the image of the j-th standard basis vector. (2.9)

Multiplication by an $m \times n$ matrix defines a linear map $\mathbb{F}^n \to \mathbb{F}^m$. Conversely, every linear map between column vector spaces is given by multiplication with a matrix. It is common not to distinguish these two concepts. For example, a matrix is said to be "invertible" if the corresponding linear map is.

Remark. Equation (2.7) is one of the most practically important relation in all of science and engineering. A significant percentage of all operations in scientific computing and machine learning / inference is spent evaluating matrix-vector products.

Example. (Differentiation, the linalg way). Let's apply $(\partial_x - 1) : p \mapsto p' + p$ to $p = 5x^2 + 3x$. Instead of working with functions directly, we'll use matrix/column vector representations w.r.t. the monomial basis of $\mathcal{P}_3$. Combine (2.4) with (2.6) to get $\boldsymbol{\phi}^{\mathcal{B}}_{\mathcal{B}}(\partial_x - 1)$ and conclude

$$\begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ -5 \\ 0 \end{pmatrix} = \boldsymbol{\phi}^{\mathcal{B}}(-5x^2 + 7x + 3).$$

2.2 Composition of maps and matrix-matrix multiplication

The composition $K \circ L$ of two linear maps K, L is linear (why?). Recall: $K \circ L$ is the map that sends v to $K(L(v))$. In the linear case, one typically writes KL for $K \circ L$.

Example.

- On $\mathbb{R}^2$, let K the rotation by ϕ, and L be the reflection about the x-axis. Then

$$\begin{aligned} KL : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) \\ &= \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \end{aligned}$$

  so KL is described by a rotation matrix with the signs of the second column inverted.

- The *gradient* is a linear map from the space of scalar functions on $\mathbb{R}^3$ to the space of vector-valued functions. The *divergence* is a linear map from the vector-valued functions to the scalar functions. Their composition

$$\operatorname{div}(\operatorname{grad} f) = \operatorname{div} \begin{pmatrix} \partial_x f \\ \partial_y f \\ \partial_z f \end{pmatrix} = \partial_x^2 f + \partial_y^2 f + \partial_z^2 f = \Delta f$$

  is the *Laplacian* Δ. (All functions are assumed to be twice differentiable).

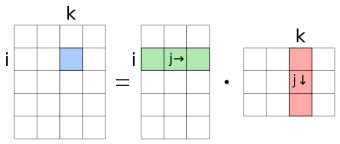
Let's work out the composition of matrix representations. For matrices $\boldsymbol{A}, \boldsymbol{B}$, we want to find the matrix $\boldsymbol{C}$ such that $\boldsymbol{Cx} = \boldsymbol{A}(\boldsymbol{Bx})$ for all $\boldsymbol{x}$. Using (2.9):

- $\boldsymbol{Ce}_j$ is the j-th column of $\boldsymbol{C}$. So it suffices to ensure $\boldsymbol{Ce}_j = \boldsymbol{A}(\boldsymbol{Be}_j)$ for all j.

- But $\boldsymbol{B}\boldsymbol{e}_j$ is the j-th column of $\boldsymbol{B}$,
- so $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{e}_j)$ is the matrix-vector product of $\boldsymbol{A}$ and the j-th column of $\boldsymbol{B}$.

Therefore:

The j-th column of $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ is the product of $\boldsymbol{A}$ and the j-th column of $\boldsymbol{B}$.



Expressed in terms of the matrix elements:

$$c_{ik} = \sum_j a_{ij} b_{jk}. \tag{2.10}$$

"Contract right index of $\boldsymbol{A}$ with left index of $\boldsymbol{B}$".

Exercise. Rotations in 2D by an angle ϕ are realized by the matrix $\boldsymbol{R}_\phi$ given in Eq. (2.5). It is intuitive that $\boldsymbol{R}_\phi \boldsymbol{R}_\psi = \boldsymbol{R}_{\phi+\psi}$. Prove that, using matrix multiplication and this useful page.

2.2.1 The algebra of linear maps

If V, W are vector spaces over some field $\mathbb{F}$, then the set of linear maps $V \to W$ is itself a vector space. It is denoted as $L(V, W)$, or, if $W = V$, just as $L(V)$.

Remark. This can be iterated. The set $L\big(L(V)\big)$ of linear maps on linear maps is again a vector space. In QM, it is sometimes called the space of *superoperators*. Sounds bombastic, doesn't it?

The extra structure $L(V)$ has compared to a generic vector space is that elements of $L(V)$ can be multiplied with each other. Such spaces are called *algebras*.

The most important property of the algebra of linear maps is that it is not commutative: The *commutator* $[K, L] = KL - LK$ between two linear maps does not, in general, vanish.

Example.

- Remember the Pauli matrices from (1.4)? They *anti-commute* in the sense that

$$\boldsymbol{\sigma}_i \boldsymbol{\sigma}_j = -\boldsymbol{\sigma}_j \boldsymbol{\sigma}_i \qquad \forall\, i \neq j \in \{1, 2, 3\}.$$

- On differentiable functions, define the multiplication and the differentiation operator by

$$(Xf)(x) = x f(x), \qquad (Pf)(x) = -i \frac{\partial}{\partial x} f(x).$$

  Their commutator is *extremely* important in QM. Letting it act on a function f, one finds

$$[X, P] = i\mathbb{1}.$$

- Let $\mathrm{diag}((a_i))$ be the matrix that has a_i as its i-th diagonal element and is zero else. Then $\mathrm{diag}((a_i))\,\mathrm{diag}((b_i)) = \mathrm{diag}((a_i b_i))$. It follows that diagonal matrices commute with each other.

Composition of maps (linear or not) is *associative*: $K(LM) = (KL)M$. We can therefore write KLM without ambiguity.

2.3 Linear forms

A linear map $L : V \to \mathbb{F}$ is called *linear form*, *(linear) functional*, or *covector*.

Example. Examples 3 – 6 in Sec. 2 are linear functionals.

Remark. Confusion alert: In other contexts, a *functional* is a function acting on functions, linear or not. This use is common in the *calculus of variation*, in particular in Lagrangian mechanics.

Choose a basis $\mathcal{A} = \{v_i\}$ for V. The one-dimensional space $\mathbb{F}$ has a canonical basis given by $\mathcal{B} = \{1\}$. Therefore, the coordinate representation

$$\boldsymbol{\phi}_{\mathcal{A}}(L) := \boldsymbol{\phi}_{\mathcal{A}}^{\{1\}}(L) = \begin{pmatrix} L(v_1) & \dots & L(v_n) \end{pmatrix} \in \mathbb{F}^{1\times n} \tag{2.11}$$

of a linear form is a row vector. Its elements are the values of L on the basis.

Some definitions:

- The space $L(V, \mathbb{F})$ of linear forms is denoted by V^* and called the *dual space* of V.
- Let $\boldsymbol{\epsilon}_i \in \mathbb{F}^{1\times n}$ be the row vector with 1 in the i-th column and 0 else. The set $\{\epsilon_i\}$ is the *standard dual basis*. It fulfills $\boldsymbol{\epsilon}_i\, \boldsymbol{e}_j = \delta_{ij}$.
- Let $\mathcal{B} = \{v_j\}$ be a basis in V. Define the linear form ν_i by linear extension of $\nu_i(v_j) = \delta_{ij}$. Then $\{\nu_i\}$ is the *dual basis* associated with $\{v_j\}$ and $\boldsymbol{\phi}_{\mathcal{B}}(\nu_i) = \boldsymbol{\epsilon}_i$.

Example. For a linear form L and a vector v with coordinate representations

$$\boldsymbol{\phi}_{\mathcal{A}}(L) = \begin{pmatrix} \alpha_1 & \dots & \alpha_n \end{pmatrix}, \qquad \boldsymbol{\phi}^{\mathcal{A}}(v) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

an application of the formula (2.7) for matrix-vector multiplication gives

$$L(v) = \begin{pmatrix} \alpha_1 & \dots & \alpha_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_i \alpha_i x_i.$$

In coordinates, the evaluation of a linear form on a vector corresponds to a product between a row and a column vector, which in turn is the contraction $\sum_i \alpha_i x_i$ of their coefficients.

2.3.1 Transpose and adjoint of column vectors

The *transpose* of an $m \times n$ matrix $\boldsymbol{A} = (a_{ij})$ is the $n \times m$ matrix $\boldsymbol{A}^t = (b_{ij})$ with rows and columns interchanged: $b_{ij} = a_{ji}$. In particular, the transpose of a column vector is a row vector with the same components.

On $\mathbb{R}^n$, the inner product with a vector $\boldsymbol{w}$ is

$$\boldsymbol{v} \mapsto \langle \boldsymbol{w}, \boldsymbol{v} \rangle = \sum_i w_i v_i = \boldsymbol{w}^t \boldsymbol{v}.$$

Thus: The inner product with $\boldsymbol{w}$ is the linear form represented by the row vector $\boldsymbol{w}^t$.

Likewise, the *adjoint* of a matrix $\boldsymbol{A}$ is $\boldsymbol{A}^\dagger = \bar{\boldsymbol{A}}^t$, the conjugate-transpose (pronounced "A dagger"). On $\mathbb{C}^n$ with the standard Hermitian scalar product, we have

$$\boldsymbol{v} \mapsto \langle \boldsymbol{w}, \boldsymbol{v} \rangle = \sum_i \bar{w}_i v_i = \boldsymbol{w}^\dagger \boldsymbol{v}.$$

In computer algebra systems, the standard Euclidean / Hermitian scalar products are often implemented this way. In Julia, where an apostrophe (`'`) denotes the adjoint:

```
julia> v=[1;im] # Column vector. 'im' is the imaginary unit i.
2-element Vector{Complex{Int64}}:
 1 + 0im
 0 + 1im

julia> v' # Its adjoint is a row vector. Note the conjugation.
1×2 adjoint(::Vector{Complex{Int64}}) with eltype Complex{Int64}:
 1+0im  0-1im

julia> v'*v # Standard Hermitian norm squared
2 + 0im
```

Remark. The adjoint of a vector v is denoted by $v^\dagger$ in physics. In analysis, it's v^*. In Julia, v'. In QM, vectors are written as "kets" $|v\rangle$ and their adjoints as "bras" $\langle v|$. In covariant notation, taking the transpose corresponds to "lowering the index": $v^i \mapsto v_i$. In geometry, one writes $v^\flat$ (because the musical symbol $\flat$ denotes the... wait for it... *lowering* of the pitch). Madness.

2.4 Differentials

Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable. Its first-order Taylor expansion around $\boldsymbol{p} \in \mathbb{R}^n$ is

$$f_i(\boldsymbol{p} + \boldsymbol{x}) \simeq f_i(\boldsymbol{p}) + \sum_j \frac{\partial f_i(\boldsymbol{p})}{\partial p_j} x_j, \qquad i = 1, \dots, n.$$

We see that changing the argument from $\boldsymbol{p}$ to $\boldsymbol{p} + \boldsymbol{x}$ produces, to first order, a change in the function value that is given by a matrix-vector product. We can write it as

$$\mathrm{d}_{\boldsymbol{p}}\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m, \quad \boldsymbol{y} \mapsto \boldsymbol{J}^{\boldsymbol{f}}_{\boldsymbol{p}} \boldsymbol{x}, \qquad \text{where} \qquad \boldsymbol{J}^{\boldsymbol{f}}_{\boldsymbol{p}} = \begin{pmatrix} \frac{\partial f_1(\boldsymbol{p})}{\partial p_1} & \cdots & \frac{\partial f_1(\boldsymbol{p})}{\partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\boldsymbol{p})}{\partial p_1} & \cdots & \frac{\partial f_m(\boldsymbol{p})}{\partial p_n} \end{pmatrix}.$$

$\boldsymbol{J}_{\boldsymbol{p}}$ is the *Jacobian matrix*, and the linear map $\mathrm{d}_{\boldsymbol{p}}\boldsymbol{f}$ it defines is the *differential*, of $\boldsymbol{f}$ at $\boldsymbol{p}$. The differential is thought of as a "linearization" of the function $\boldsymbol{f}$ around a given point.

Remark.

- For $m = 1$, i.e. for scalar functions f on $\mathbb{R}^n$, the differential is a linear form

$$\mathrm{d}_{\boldsymbol{p}} f : \boldsymbol{x} \mapsto \begin{pmatrix} \frac{\partial f(\boldsymbol{p})}{\partial p_1} & \cdots & \frac{\partial f(\boldsymbol{p})}{\partial p_n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}. \tag{2.12}$$

  In physics, this is more commonly written as the inner product between the *gradient*

$$\operatorname{grad} f = \begin{pmatrix} \frac{\partial f(\boldsymbol{p})}{\partial p_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{p})}{\partial p_n} \end{pmatrix}$$

  and the deviation $\boldsymbol{x}$. Mathematically, the differential is the more fundamental notion: It can also be defined on spaces that do not have a scalar product to change the column vector $\operatorname{grad} f$ to the row vector representing the linear form that's actually meant.

- Let's be a bit flexible with notation. Use x_i to denote, depending on context, (1) the function $\mathbb{R}^n \to \mathbb{R}$ that maps a column vector to its ith component; (2) the result $x_i = x_i(\boldsymbol{v}) \in \mathbb{R}$ of that function acting on a vector; (3) a "variable name" which indicates that $\frac{\partial}{\partial x_i}$ refers to the partial derivative w.r.t. the ith argument. As a function, x_i has a differential $\mathrm{d}_{\boldsymbol{p}} x_i$. Because $\frac{\partial x_i}{\partial x_j} = \delta_{ij}$, it is represented by the standard dual basis element $\boldsymbol{\epsilon}_i$. Expanding (2.12) in that basis (and suppressing $\boldsymbol{p}$) gives

$$\mathrm{d}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, \mathrm{d}x_i.$$

  This formula can be read in two ways: (1) Intuitively, as relating "infinitesimal" changes of the arguments to those of the function value. (2) Mathematically rigorous, as an equation between well-defined linear maps.

- The chain rule of differentiation states

$$\frac{\partial (f \circ g)_i}{\partial x_k} = \sum_j \frac{\partial f_i}{\partial g_j} \frac{\partial g_j}{\partial x_k}. \tag{2.13}$$

  That sure looks like a matrix product! And indeed, if one decodes the terse notation (the "flexibility" sketched above now also applies to f_i, g_j), Eq. (2.13) says that the Jacobian $\boldsymbol{J}_{\boldsymbol{p}}^{\boldsymbol{f}\circ\boldsymbol{g}}$ of a composition is the matrix product of the individual Jacobians:

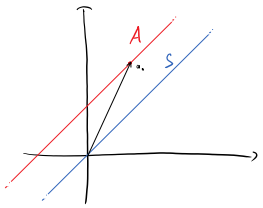
$$\boldsymbol{J}_{\boldsymbol{p}}^{\boldsymbol{f}\circ\boldsymbol{g}} = \boldsymbol{J}_{\boldsymbol{g}(\boldsymbol{p})}^{\boldsymbol{f}} \boldsymbol{J}_{\boldsymbol{p}}^{\boldsymbol{g}} \qquad \text{and thus} \qquad \mathrm{d}(\boldsymbol{f} \circ \boldsymbol{g}) = (\mathrm{d}\boldsymbol{f})(\mathrm{d}\boldsymbol{g}).$$

2.5 Kernels and images

2.5.1 Affine subspaces

Take a subspace $S \subset V$ and a vector $a_0 \in V$. Now consider the "shifted subspace"

$$A = a_0 + S := \{a_0 + s \,|\, s \in S\}.$$



A set of this form is called an *affine subspace of* V.

Elementary facts:

1. One can recover S from A as

$$S = A - A := \{a_1 - a_2 \,|\, a_1, a_2 \in A\}.$$

   It is called the *directional vector space* of A.

2. A set $A \subset V$ is affine if and only if $A - A$ is a linear subspace.

Examples are considered in the next section.

2.5.2 Kernels and images

With a linear map $L : V \to W$, associate the sets

$$\begin{aligned} \operatorname{img} L &:= \{Lv \,|\, v \in V\} \subset W && \text{the } \textit{image} \text{ or } \textit{range} \text{ of } L, \\ \ker L &:= \{v \in V \,|\, Lv = 0\} \subset V && \text{the } \textit{kernel} \text{ or } \textit{null space} \text{ of } L. \end{aligned}$$

Both the image and the kernel are linear spaces.

A not insignificant part of your time as a scientist will be spent solving *(inhomogeneous) linear equations*: Given L, f, find u such that

$$Lu = f. \tag{2.14}$$

The set of solutions to (2.14) is an affine space with directional vector space equal to $\ker L$.

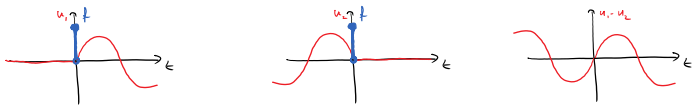
Proof. If u_1 and u_2 are solutions, then $L(u_1 - u_2) = f - f = 0$, so that $u_1 - u_2 \in \ker L$. Conversely, if u is a solution and $v \in \ker L$, then $L(u + v) = L(u) = f$ is a solution. □

Example. **Driven harmonic oscillator.** Newton's equation for the position $u(t)$ of a harmonic oscillator with frequency ω subject to a driving force $mf(t)$ is

$$\left(\partial_t^2 + \omega^2\right) u = f.$$

The kernel of $L = \partial_t^2 + \omega^2$ on the space of bounded functions has basis $\{\cos(\omega t), \sin(\omega t)\}$. Thus the set of solutions is of the form $u(t) = u_0(t) + A\cos(\omega t) + B\sin(\omega t)$, where u_0 is any particular solution, and A, B are constants. Finding a u_0 given f is considerably harder.

Consider, e.g., a driving force acting on an oscillator at $t = 0$.



If the system was previously at rest, it will then start oscillating (left). If it was moving before, then, with a bit of luck, the driving force will exactly cancel the upwards momentum (center). The difference between the two solutions is a free oscillation (right).

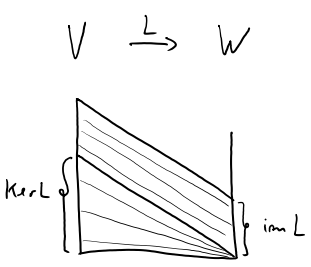
We'll talk about practical ways for solving linear equations in Sec. 3.6.2.
The discussion implies that L is injective if and only if $\ker L = \{0\}$.

2.5.3 Rank-nullity theorem, down to earth approach

If $L : V \to W$ and V is finite-dimensional, then

$$\dim \operatorname{img} L + \dim \ker L = \dim V. \tag{2.15}$$

The dimension of the image of L is the *rank* of L. The dimension of the kernel of L is sometimes called the *nullity* of L. Thus (2.15) is also known as the *rank-nullity theorem*.



The textbook by Fischer and Kaul contains this figure, illustrating the "no-expansion-of-dimension" property of linear maps, which is made precise by the rank-nullity theorem. I mean, (2.15) is an equation relating three natural numbers, so I'm not sure it really *needs* a visualization per se. But then again, F&K are much more experienced in teaching this material than I am. And, gotta admit, the figure has a certain memetic *je ne sais quoi*, so I didn't want to withhold it from you.

Proof. Choose a basis $\{b_1, \ldots, b_s\}$ of $\ker L$. By the basis extension theorem, one can complete it to a basis $\{b_1, \ldots, b_s, c_1, \ldots, c_r\}$ of V. Then $s + r = \dim V$ and $s = \dim \ker L$.

The set $\mathcal{B} = \{Lc_i\}_{i=1}^s$ spans $\operatorname{img} L$. It is also linearly independent (hence a basis):

$$\sum_i x_i L(c_i) = L\Big(\sum_i x_i c_i\Big),$$

so if the left hand side is 0, then $v = \sum_i x_i c_i \in \ker L$. But the b_i form a basis of $\ker L$, so that also $v = \sum_i y_i b_i$. By the uniqueness of basis expansions, all the x_i must be 0.

Thus we have $\dim \operatorname{img} L = s$, proving the claim. □

Remark. Confusion alert (English, *German*): rank = *Rang* ≠ range = *Bild*.

Example. Let V be Euclidean or Hermitian, $0 \neq v \in V$. Consider the projection P_v. Then $\operatorname{img} P_v$ consists of the multiples of v, thus $\operatorname{rank} P_v = 1$. The kernel $\ker P_v$ is the space $v^\perp = \{w \in V \,|\, \langle v, w\rangle = 0\}$ of vectors orthogonal to v (its *orthocomplement*). One can extend v to a basis and use Gram-Schmidt to obtain an orthogonal basis $\{w_1 = v, w_2, \dots\}$. Then $\mathcal{B} = \{w_2, \dots\}$ is a basis of $v^\perp$. If $\dim V = n < \infty$, then $|\mathcal{B}| = n - 1$, compatible with (2.15).

2.5.4 Optional: Rank-nullity, what's actually going on

Quotient spaces

Fix a subspace $S \subset V$. Remarkably, the set of affine subspaces with directional vector space S is itself a vector space, the *quotient space V up to S*, denoted by V/S.

Addition and scalar multiplication between sets are defined as

$$A + B = \{a + b \,|\, a \in A, b \in B\}, \qquad \lambda A = \{\lambda a \,|\, a \in A\}.$$

We need to verify: (1) If A, B are affine subspaces with directional vector space S, then the same is true for their sum and scalar multiples; and (2) These operations satisfy Def. 2.

Associate with every element $v \in V$ the affine subspace it lies in:

$$v \mapsto [v]_S := v + S.$$

Claim: This map commutes with addition and scalar multiplication

$$[v + w]_S = [v]_S + [w]_S, \qquad \lambda[v]_S = [\lambda v]_S. \tag{2.16}$$

Proof. Using that S, as a subspace, is closed under addition and scalar multiplication,

$$\begin{aligned} \lambda[v]_S &= \lambda(v + S) = \lambda v + \lambda S = \lambda v + S = [\lambda v]_S \\ [v + w]_S &= (v + w) + S = (v + w) + S + S = (v + S) + (w + S) = [v]_S + [w]_S. \end{aligned}$$

□

Then condition (1) follows immediately, and condition (2) is an easy exercise.

The map $v \mapsto [v]_S$ is called the *canonical projection* from V to V/S.

Rank-nullity revisited

Isomorphic vector spaces have the same dimension. It'd be cool if we could associate a vector space with each side of rank-nullity theorem, and prove that the two are isomorphic. This would "explain the equality in a deeper way".

Not only is this possible, but this formulation also works in infinite dimensions (and analogous constructions exist in group theory and in non-commutative algebra).

The central claim is (Fig. **??**): $\operatorname{img} L$ is isomorphic to the quotient space $V/\ker L$.

Proof. Write $K = \ker L$. Define a linear map $\tilde{L}$ from V/K to $\operatorname{img} L$ by

$$\tilde{L}[v]_K = Lv.$$

$\tilde{L}$ is well-defined, because any $v' \in [v]_K$ differs from v by an element of K, so that $Lv' = Lv$. It is surjective, because if $u = Lv \in \operatorname{img} L$, then $\tilde{L}[v]_K = u$. It is injective, because if $\tilde{L}[v_1]_K = \tilde{L}[v_2]_K$, then $0 = \tilde{L}[v_1 - v_2]_K = L(v_1 - v_2)$ so that $(v_1 - v_2) \in K$ and thus $[v_1 - v_2]_K = [0]_K$. □

In the special case where V is finite-dimensional, the proof in the previous section can be re-read to show that

$$\dim V/K = \dim V - \dim K,$$

so we recover the rank-nullity theorem in the re-arranged way

$$\dim \operatorname{img} L = \dim V - \dim \ker L.$$

2.6 Inverse maps

If $L : V \to W$ is invertible, then the inverse function $L^{-1} : W \to V$ is also linear.

Proof. Given $w \in W$, let $v = L^{-1}(w)$. By linearity, $L(\lambda v) = \lambda w$. Applying L^{-1} to both sides gives $L^{-1}(\lambda w) = \lambda v = \lambda L^{-1}(w)$.

Showing that L^{-1} commutes with addition works analogously. □

If V, W are finite-dimensional, then the rank-nullity theorem implies:

1. L can be invertible only if $\dim V = \dim W$.
2. If $\dim V = \dim W$, then:

$$\begin{aligned} & L \text{ is invertible} \\ \Leftrightarrow\ & \dim \ker L = 0, && \text{i.e. the nullity of } L \text{ is 0, or } L \text{ is injective} \\ \Leftrightarrow\ & \dim \operatorname{img} L = \dim W, && \text{i.e. the rank of } L \text{ is maximal, or } L \text{ is surjective.} \end{aligned}$$

Remark. These equivalences do not hold in infinite dimensions. Take $V = W = \mathcal{P}(\mathbb{C})$, the space of polynomials. The multiplication operator $p \mapsto xp$ injective. But it is not surjective, because the constant function $p(x) = 1$ is not in its range. The differentiation operator $p \mapsto \partial_x p$ is surjective. But it is not injective, because its kernel contains all constant functions.

Example. Let V be the vector space of all functions on $\mathbb{R}$. Let m be a function without a zero: $m(x) \neq 0$. The inverse of the multiplication operator $(M_m f)(x) = m(x)f(x)$ is $M_{1/m}$.

Taking the inverse reverses compositions: $(KL)^{-1} = L^{-1}K^{-1}$. Proof:

$$\left(L^{-1}K^{-1}\right)(KL) = L^{-1}(K^{-1}K)L = L^{-1}L = \mathbb{1}.$$

2.6.1 Inverse matrices

Now assume that V, W are both n-dimensional, and that bases have been chosen.

If L is represented by a matrix $\boldsymbol{A}$, then $\boldsymbol{B}$ is a matrix representation of L^{-1} if and only if $\boldsymbol{AB} = \mathbb{1}$. In this case, we write $\boldsymbol{B} = \boldsymbol{A}^{-1}$, the *inverse matrix* of $\boldsymbol{A}$.

Example.

- If $\boldsymbol{A} = \mathrm{diag}((a_i))$ is diagonal with non-zero entries, then $\boldsymbol{A}^{-1} = \mathrm{diag}((a_i^{-1}))$.
- The Pauli matrices (1.4) square to identity $\boldsymbol{\sigma}_i^2 = \mathbb{1}$. Hence $\boldsymbol{\sigma}_i^{-1} = \boldsymbol{\sigma}_i$.
- Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ be invertible. By Sec. 2.4, the Jacobi matrices fulfill $\left(\boldsymbol{J}^{\boldsymbol{f}}_{\boldsymbol{p}}\right)^{-1} = \boldsymbol{J}^{\boldsymbol{f}^{-1}}_{\boldsymbol{f}(\boldsymbol{p})}$.
- A real square matrix $\boldsymbol{O}$ is *orthogonal* if $\boldsymbol{O}^t\boldsymbol{O} = \mathbb{1}$, i.e. if $\boldsymbol{O}^{-1} = \boldsymbol{O}^t$. Orthogonal matrices preserve the Euclidean scalar product—we'll look at them in detail later.

For 2×2-matrices, there is an explicit formula for the matrix inverse (check it!):

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - cb} \begin{pmatrix} d & -b \\ c & a \end{pmatrix}.$$

There are a few more classes of matrices for which the inverse can be easily found. But in general, computing matrix inverses is painful for humans and even computers are only moderately good at it, with speed and numerical stability being issues. Experience has it that problems that seem to require a matrix inverse can often be re-formulated to avoid it.

2.7 Coordinate changes

2.7.1 General case

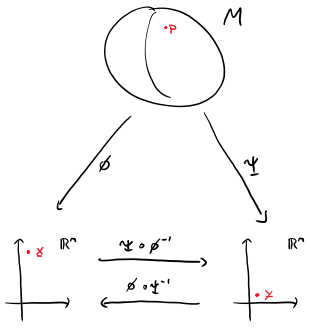
We start with a few remarks on coordinates in general, not just linear ones on vector spaces.

Manifolds

Let M be a set of points on which we want to do physics. For example, M could be Euclidean space, or a model of the surface of the earth or of the entire universe. (In the mathematical theory, these sets go by the pleasant-sounding name of *manifolds*).

Coordinate functions

In order to talk about the elements of M concretely, it is helpful to choose coordinates. A *coordinate function* is an invertible function from (a subset of) M to $\mathbb{R}^n$. With respect to a coordinate function $\boldsymbol{\phi}$, a point $p \in M$ is represented by the column vector $\boldsymbol{x} = \boldsymbol{\phi}(p)$.

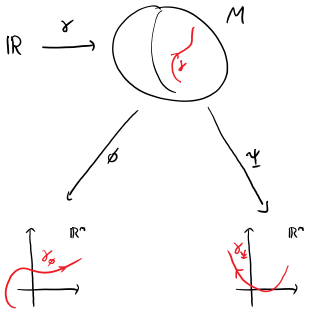


For example, if M models earth, it contains a point $p_{\wedge\wedge}$ corresponding to the center of Cologne. If $\boldsymbol{\phi} : M \to \mathbb{R}^2$ assigns to each point their latitude and longitude in degrees, then $\boldsymbol{\phi}(p_{\wedge\wedge}) = \left(\begin{smallmatrix} 50.933594 \\ 6.961899 \end{smallmatrix}\right)$.

If one has chosen two different coordinate systems $\boldsymbol{\phi}, \boldsymbol{\psi}$, one can ask which function converts the first type to the second type. Read off the figure: Starting with $\boldsymbol{x} = \boldsymbol{\phi}(p)$, move up $\boldsymbol{\phi}^{-1}$ to get to p, and then down again along $\boldsymbol{\psi}$, reaching $\boldsymbol{y} = \boldsymbol{\psi}(p)$. Thus, the coordinate change is given by $\boldsymbol{\psi} \circ \boldsymbol{\phi}^{-1}$. In the other direction, it's $\boldsymbol{\phi} \circ \boldsymbol{\psi}^{-1}$.

Curves

In the mechanics of point particles, you will study *curves* $\gamma : \mathbb{R} \to M$, where $\gamma(t)$ is the position of a particle at time t. Particles have traced out trajectories long before humans started choosing reference systems, so γ exists independently of any coordinates. But if we want to, we can consider representations

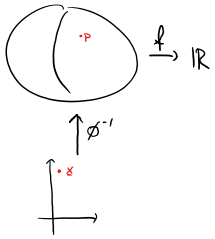


$$\gamma_\phi := \phi \circ \gamma : \mathbb{R} \to \mathbb{R}^n, \quad \text{or}$$
$$\gamma_\psi := \psi \circ \gamma : \mathbb{R} \to \mathbb{R}^n.$$

These can be converted into each other using the coordinate changes described above:

$$(\psi \circ \phi^{-1}) \circ \gamma_\phi = \psi \circ \phi^{-1} \circ \phi \circ \gamma = \psi \circ \gamma = \gamma_\psi. \tag{2.17}$$

Scalar functions



Another important set of objects are scalar functions $f : M \to \mathbb{R}$. Think of, say, temperature or a potential, that take a real value at every point. A coordinate representation f_ϕ of f maps coordinates $\boldsymbol{x} = \phi(p)$ of a point $p \in M$ to the value $f(p)$. Thus:

$$f_\phi = f \circ \phi^{-1} \tag{2.18}$$

("start from $\boldsymbol{x}$, move up through ϕ^{-1} to p, then apply f").

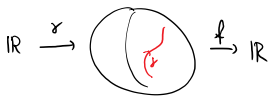
A very important lesson:

> Coordinate representations of functions *to* M depend on ϕ *directly*.
> Coordinate representations of functions *on* M depend on the *inverse* ϕ^{-1}.

And indeed, converting f_ϕ to f_ψ involves the inverse of the map encountered in (2.17):

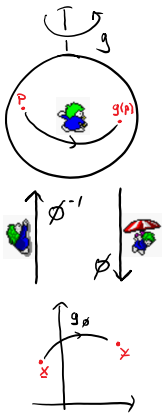
$$f_\phi \circ (\phi \circ \psi^{-1}) = f \circ \phi^{-1} \circ \phi \circ \psi^{-1} = f \circ \psi^{-1} = f_\psi.$$

Curves and scalar functions can be concatenated to $f \circ \gamma : \mathbb{R} \to \mathbb{R}$, describing the evolution of f as t is varied, without referencing any coordinates.



(2.19)

Transformations



Consider a map $g : M \to M$, describing, e.g., a rotation. In coordinates

$$g_\phi = \phi \circ g \circ \phi^{-1}. \tag{2.20}$$

("Start with $\boldsymbol{x}$, crawl up ϕ^{-1} to get to p, slide along g, then down again through ϕ, thereby reaching $\boldsymbol{y} = \phi(g(p))$.") Because g is both a function *on* M and *to* M, its coordinate representation depends on ϕ and ϕ^{-1}.

Changing coordinates:

$$\begin{aligned}&(\psi \circ \phi^{-1}) \circ g_\phi \circ (\phi \circ \psi^{-1})\\ =&\psi \circ \phi^{-1} \circ \phi \circ g \circ \phi^{-1} \circ \phi \circ \psi^{-1} = g_\psi.\end{aligned}$$

Passive transformations

Ever been confused about whether your train just started rolling out of the station, or whether the one on the neighboring track began moving in the opposite direction?

Let $g : M \to M$. Its action $p \mapsto g(p)$ on points is called *active*. If g is invertible, one can also use it to act on coordinate systems (analogous to the reference system defined by the neighboring train). This *passive* action takes a coordinate function ϕ and creates a new coordinate function $\psi := \phi \circ g^{-1}$.

The coordinate change from the original ϕ to the new system ψ is

$$\phi \circ g^{-1} \circ \phi^{-1}. \tag{2.21}$$

Formally, this is the inverse g_ϕ^{-1} of the coordinate representation of the active action (recall the opposite direction of motion of the reference train).

Remarks.

- "Passive transformation" sometimes refers to the map $\phi \mapsto \phi \circ g^{-1}$ (as in our definition) and sometimes, if ϕ is clear from context, to the coordinate change $\phi \circ g^{-1} \circ \phi^{-1}$.
- All other objects encountered (M, curves, scalar functions, (active) transformations) exist independently of coordinates. That's obviously not the case for passive transformations. It thus comes as no surprise that passive transformations are often seen in science and engineering, but much less talked about in math, with its emphasis on canonic objects.
- Compare "If a physical process is possible, then so is any rotated version of it" to "The equations of physics are invariant under rotations of the coordinate system". These statements express the same fact about reality, respectively from an active and a passive mindset.

Notation

We have encountered many objects in quick succession. In physics books, most of them are not mentioned explicitly. Let me try to give a fair summary of the standard terminology:

concept	math	physics	comment
point	$p \in M$	$\boldsymbol{r}$	
coordinate function	$\boldsymbol{\phi} : M \to \mathbb{R}^n$	$\boldsymbol{r}$	
different coordinate function	$\boldsymbol{\psi} : M \to \mathbb{R}^n$	$\boldsymbol{r}'$	
coordinates of a point	$\boldsymbol{x} = \boldsymbol{\phi}(p)$	$\boldsymbol{r}$	I'm sensing a pattern!
curve	$\gamma : \mathbb{R} \to M$	$\boldsymbol{r}(t)$, $\boldsymbol{r}$	
position of curve at time t	$\gamma(t)$	$\boldsymbol{r}(t)$	
coordinate rep of curve	$\gamma_{\boldsymbol{\phi}} = \phi \circ \gamma$	$\boldsymbol{r}(t), \boldsymbol{r}$	
scalar function	$f : M \to \mathbb{R}$	$f(\boldsymbol{r}), f$	
coordinate rep of function	$f_{\boldsymbol{\phi}} = f \circ \phi^{-1}$	$f(\boldsymbol{r}), f$	
function along curve	$f \circ \gamma$	$f(t)$	Wait, where's the curve?
coordinate change	$\boldsymbol{\psi} \circ \boldsymbol{\phi}^{-1}$	$\boldsymbol{r}'(\boldsymbol{r})$	I'm not making this up.
⋮	⋮	⋮	¯_(ツ)_/¯

As long as you can keep all the different meanings separated in your head, the terse physics notation is extremely clean and efficient. But God help you should you ever get confused!

2.7.2 In linear algebra

In linear algebra, coordinates are derived from bases. For a basis $\mathcal{B}$, recall that $\boldsymbol{\phi}^{\mathcal{B}}$ is the linear maps $V \to \mathbb{F}^n$ that gives the expansion coefficients w.r.t. $\mathcal{B}$ (Eq. (1.5)).

Vectors

Assume two bases $\mathcal{B} = \{v_i\}_{i=1}^n$ and $\mathcal{B}' = \{v'_i\}_{i=1}^n$ have been chosen on a space V. There are thus two sets of coordinates associated with every vector:

$$\boldsymbol{x} = \boldsymbol{\phi}^{\mathcal{B}}(v), \qquad \boldsymbol{x}' = \boldsymbol{\phi}^{\mathcal{B}'}(v).$$

The coordinate change

$$\boldsymbol{\phi}^{\mathcal{B}'} \circ \left(\boldsymbol{\phi}^{\mathcal{B}}\right)^{-1} : \mathbb{F}^n \to \mathbb{F}^n \tag{2.22}$$

is a linear function and thus realized by multiplication with an $n \times n$-matrix. Denote it as $\boldsymbol{T}_{\mathcal{B}}^{\mathcal{B}'}$, or, if the bases are clear from context, as $\boldsymbol{T}$:

$$\boldsymbol{x}' = \boldsymbol{\phi}^{\mathcal{B}'} \circ \left(\boldsymbol{\phi}^{\mathcal{B}}\right)^{-1}(\boldsymbol{x}) = \boldsymbol{T}\boldsymbol{x}. \tag{2.23}$$

Concretely,

$$\boldsymbol{T}\boldsymbol{e}_i = \boldsymbol{\phi}^{\mathcal{B}'} \circ \left(\boldsymbol{\phi}^{\mathcal{B}}\right)^{-1}(\boldsymbol{e}_i) = \boldsymbol{\phi}^{\mathcal{B}'}(v_i),$$

so the i-th column of $\boldsymbol{T}$ is the representation of the i-th old basis vector v_i in terms of the new basis.

Example. Take $V = \mathbb{R}^2$ and $\boldsymbol{v}_i = \boldsymbol{e}_i$ the standard basis. Then $\boldsymbol{\phi}^{\mathcal{B}}$ is just the identity map. Now choose, say,

$$\boldsymbol{v}'_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad \boldsymbol{v}'_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Expanding the old basis vectors in the new basis,

$$\boldsymbol{T} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 \end{pmatrix}.$$

Relation (2.23) is called the *contravariant transformation law*.

Linear forms

Now let $f : V \to \mathbb{F}$ be a linear form. By (1.9), it is represented by the row vectors

$$\boldsymbol{y} = \phi_{\mathcal{B}}(f), \qquad \boldsymbol{y}' = \phi_{\mathcal{B}'}(f).$$

As in (2.18): With $\boldsymbol{T}$ the same matrix as above, they are related by

$$\boldsymbol{y}' = \boldsymbol{y}\,\boldsymbol{T}^{-1}. \tag{2.24}$$

Relation (2.24) is the *covariant transformation law*. As a consistency check, let's verify that the contraction between a co- and a contravariant quantity is invariant:

$$\boldsymbol{y}'\,\boldsymbol{x}' = \boldsymbol{y}\,\boldsymbol{T}^{-1}\,\boldsymbol{T}\,\boldsymbol{x} = \boldsymbol{y}\,\boldsymbol{x}.$$

This had better be the case, because both expressions describe the coordinate-independent number $f(v)$ (c.f. (2.19)).

Example. If $\boldsymbol{T}\,\boldsymbol{x} = \lambda\boldsymbol{x}$ stretches vectors, then $\boldsymbol{y}\,\boldsymbol{T}^{-1} = \frac{1}{\lambda}\boldsymbol{y}$ shrinks to compensate.

Linear maps

Let $L : V \to V$ be a linear map with matrix representations

$$\boldsymbol{A} = \phi^{\mathcal{B}}_{\mathcal{B}}(L), \qquad \boldsymbol{A}' = \phi^{\mathcal{B}'}_{\mathcal{B}'}(L).$$

As in (2.20),

$$\boldsymbol{A}' = \boldsymbol{T}\,\boldsymbol{A}\,\boldsymbol{T}^{-1}. \tag{2.25}$$

One says that $\boldsymbol{T}$ *acts on* $\boldsymbol{A}$ *by conjugation*. Two matrices $\boldsymbol{A}, \boldsymbol{A}'$ that are related by conjugation with an invertible $\boldsymbol{T}$ are *similar*.

Consistency check, as above:

$$f(Lv) = \boldsymbol{y}'\boldsymbol{A}'\boldsymbol{x}' = (\boldsymbol{y}\boldsymbol{T}^{-1})(\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1})(\boldsymbol{T}\boldsymbol{x}) = \boldsymbol{y}\boldsymbol{A}\boldsymbol{x} = f(Lv).$$

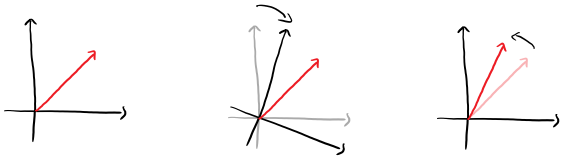
Passive action

If $L : V \to V$ is invertible, it maps bases to bases:

$$\mathcal{B} = \{v_i\} \mapsto \mathcal{B}' = L(\mathcal{B}) = \{Lv_i\}.$$

Then $\phi^{L(\mathcal{B})} = \phi^{\mathcal{B}} \circ L^{-1}$, i.e. the action of invertible maps on bases gives rise to the "passive action".

Let $\boldsymbol{T} = \phi^{L(\mathcal{B})} \circ (\phi^{\mathcal{B}})^{-1}$ be the matrix implementing the change from the old to the new coordinate system. Formally, it equals the basis representation $\phi^{\mathcal{B}} \circ L^{-1} \circ (\phi^{\mathcal{B}})^{-1}$ of the inverse of L. "Rotating the basis to the right has the same effect on coordinates as rotating all vectors to the left":



Chapter 3

Eigenvalues and eigenvectors

Two matrix representations $\boldsymbol{A}$, $\boldsymbol{A}'$ of a linear map are related by a similarity transformation $\boldsymbol{A}' = \boldsymbol{T}\,\boldsymbol{A}\,\boldsymbol{T}^{-1}$ (c.f. (2.25)). Matrices that are "similar" in this technical sense can look very different in practice. Given L, it is of great interest to find a basis w.r.t. which the representation $\boldsymbol{A}$ is particularly simple. (In fact, in QM, "solving a system" has come to mean "finding a basis in which the operator describing energy is diagonal".) It is this problem we'll address in the present chapter.

We'll first look at *invariants* of similarity transforms, i.e. functions of matrices that do not change under $\boldsymbol{A} \mapsto \boldsymbol{T}\,\boldsymbol{A}\,\boldsymbol{T}^{-1}$, then turn to the discussion of eigenvectors.

3.1 The trace

The *trace* of a square matrix is the sum over its diagonal elements

$$\operatorname{tr} \boldsymbol{A} = \sum_i a_{ii}.$$

The trace has the following *cyclicity property*

$$\operatorname{tr} \boldsymbol{A}\boldsymbol{B} = \operatorname{tr} \boldsymbol{B}\boldsymbol{A} \tag{3.1}$$

which implies

$$\operatorname{tr} \boldsymbol{A}_1\boldsymbol{A}_2 \ldots \boldsymbol{A}_{k-1}\boldsymbol{A}_k = \operatorname{tr} \boldsymbol{A}_2\boldsymbol{A}_3 \ldots \boldsymbol{A}_k\boldsymbol{A}_1 \quad \text{and} \quad \operatorname{tr} \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \operatorname{tr} \boldsymbol{A}. \tag{3.2}$$

Proof. Homework. □

If L is an abstract linear map, define $\operatorname{tr} L := \operatorname{tr} \boldsymbol{A}$, where $\boldsymbol{A} = \phi^{\mathcal{B}}(L)$ is the matrix representation w.r.t. some basis. By (3.2), the result does not depend on the basis chosen.

Example [Bloch representation]. The *trace scalar product*

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\operatorname{tr}} := \operatorname{tr} \boldsymbol{A}^\dagger \boldsymbol{B}.$$

is a Hermitian scalar product on the space $\mathbb{C}^{n\times n}$ of complex $n \times n$-matrices.

The Pauli matrices $\{\frac{1}{\sqrt{2}}\boldsymbol{\sigma}_i\}_{i=0}^3$ form an ONB for $\mathbb{C}^{2\times 2}$:

$$\frac{1}{2}\langle \boldsymbol{\sigma}_i^\dagger, \boldsymbol{\sigma}_j \rangle = \delta_{ij}.$$

It follows that any 2×2-matrix $\boldsymbol{\rho}$ can be written as

$$\boldsymbol{\rho} = \sum_i c_i \boldsymbol{\sigma}_i, \qquad c_i = \frac{1}{2} \operatorname{tr} \boldsymbol{\sigma}_i^\dagger \boldsymbol{\rho}.$$

The coefficient vector $\boldsymbol{c}$ is called the *Bloch representation* of $\boldsymbol{\rho}$ in quantum physics.

3.2 The determinant

The determinant is important in linear algebra and for multidimensional integration. In $\mathbb{R}^n$, it can be geometrically interpreted as measuring "oriented volume". We'll first explain that interpretation in $\mathbb{R}^2$ and then treat the more abstract general case.

3.2.1 Oriented volume in two dimensions

Consider a pair of vectors $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ in $\mathbb{R}^2$. They span a parallelogram. It turns out (surprisingly?) to be important to classify these according to their *area* and their *orientation*.

Some shapes, like shoes, gloves, or the letter R are not congruent to their mirror images. Roughly, "orientation" is the property that distinguishes the original from its reflection. For an ordered pair $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ of vectors in $\mathbb{R}^2$, define their *orientation* by

$$\begin{aligned}
\operatorname{ori}(\boldsymbol{a}_1, \boldsymbol{a}_2) &= +1 && \text{if } \boldsymbol{a}_1 \text{ lies to the } \textit{right} \text{ of the line through } \boldsymbol{a}_2,\\
\operatorname{ori}(\boldsymbol{a}_1, \boldsymbol{a}_2) &= -1 && \text{if } \boldsymbol{a}_1 \text{ lies to the } \textit{left} \text{ of the line through } \boldsymbol{a}_2,\\
\operatorname{ori}(\boldsymbol{a}_1, \boldsymbol{a}_2) &= \ \ 0 && \text{if } \boldsymbol{a}_1 \text{ lies } \textit{on} \text{ the line through } \boldsymbol{a}_2.
\end{aligned}$$

Let the *volume* $\operatorname{vol}(\boldsymbol{a}_1, \boldsymbol{a}_2)$ be the area of the parallelogram spanned by $\boldsymbol{a}_1, \boldsymbol{a}_2$.

The *determinant* is the "oriented volume"

$$\det(\boldsymbol{a}_1, \boldsymbol{a}_2) = \operatorname{ori}(\boldsymbol{a}_1, \boldsymbol{a}_2)\ \operatorname{vol}(\boldsymbol{a}_1, \boldsymbol{a}_2).$$

If $\boldsymbol{A}$ is the matrix with columns $\boldsymbol{a}_1, \boldsymbol{a}_2$, one also writes $\det \boldsymbol{A}$ for $\det(\boldsymbol{a}_1, \boldsymbol{a}_2)$.

Below, we will turn the geometric definition into a formula, in two steps:

(S1): The determinant is bilinear and *alternating*

$$\det(\boldsymbol{a}_1, \boldsymbol{a}_2) = -\det(\boldsymbol{a}_2, \boldsymbol{a}_1).$$

Bilinear functions are also called *bilinear forms*. Denote the set of alternating bilinear forms by $\bigwedge^2(\mathbb{R}^2)$ ("*wedge*-2"). For the next statement, write

$$\boldsymbol{a}_1 = \begin{pmatrix} a \\ c \end{pmatrix}, \quad \boldsymbol{a}_2 = \begin{pmatrix} b \\ d \end{pmatrix}, \quad \text{so that in matrix form} \quad (\boldsymbol{a}_1, \boldsymbol{a}_2) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

(S2): $\bigwedge^2(\mathbb{R}^2)$ is a 1-dimensional vector space. Any $K \in \bigwedge^2(\mathbb{R}^2)$ fulfills

$$K(\boldsymbol{a}_1, \boldsymbol{a}_2) = (ab - cd) K(\boldsymbol{e}_1, \boldsymbol{e}_2). \tag{3.3}$$

Interpret (3.3) this way: All alternating bilinear forms agree that the oriented volume of a general parallelogram is $(ab - cd)$ times the one of the unit square. But one is free to "choose units" by assigning an arbitrary value $K(\boldsymbol{e}_1, \boldsymbol{e}_2)$ to this reference shape. Because the standard choice is 1, we get

$$\det(\boldsymbol{a}_1, \boldsymbol{a}_2) = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Proof of the formula for the two-dimensional determinant

The following properties can be seen from (Fig. **??**):

(P1): Exchanging $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$...

- leaves the volume invariant, and
- flips the sign of the orientation.

(P2): Multiplying $\boldsymbol{a}_1$ or $\boldsymbol{a}_2$ by $\lambda \in \mathbb{R}$...

- multiplies the volume by $|\lambda|$, and
- preserves orientation if $\lambda > 0$ and flips the sign of the orientation if $\lambda < 0$.

(P3): Adding a multiple of one vector to the other preserves both volume and orientation.

Proof of (S1). Property (P1) implies that $\det$ is alternating, (P2) gives *homogeneity*

$$\det(\lambda \boldsymbol{a}_1, \boldsymbol{a}_2) = \lambda \det(\boldsymbol{a}_1, \boldsymbol{a}_2) = \det(\boldsymbol{a}_1, \lambda \boldsymbol{a}_2). \tag{3.4}$$

Now assume that $(\boldsymbol{a}_1, \boldsymbol{a}_2)$ are LI. Then any vector is of the form $\lambda \boldsymbol{a}_1 + \mu \boldsymbol{a}_2$ and we have

$$\begin{aligned} \det(\boldsymbol{a}_1 + (\lambda \boldsymbol{a}_1 + \mu \boldsymbol{a}_2), \boldsymbol{a}_2) &= \det((1+\lambda)\boldsymbol{a}_1, \boldsymbol{a}_2) && \text{by (P3),} \\ &= \det(\boldsymbol{a}_1, \boldsymbol{a}_2) + \lambda \det(\boldsymbol{a}_1, \boldsymbol{a}_2) && \text{by (3.4),} \\ &= \det(\boldsymbol{a}_1, \boldsymbol{a}_2) + \det(\lambda \boldsymbol{a}_1 + \mu \boldsymbol{a}_2, \boldsymbol{a}_2) && \text{by (3.4), (P3).} \end{aligned}$$

The remaining cases are treated similarly. □

Proof of (S2). Let $K \in \bigwedge^2(\mathbb{R}^2)$. Using $K(\boldsymbol{v}, \boldsymbol{v}) = -K(\boldsymbol{v}, \boldsymbol{v})$ and hence $K(\boldsymbol{v}, \boldsymbol{v}) = 0$,

$$\begin{aligned} & K(a\boldsymbol{e}_1 + c\boldsymbol{e}_2, b\boldsymbol{e}_1 + d\boldsymbol{e}_2) \\ =& ab \underbrace{K(\boldsymbol{e}_1, \boldsymbol{e}_1)}_{0} + cd \underbrace{K(\boldsymbol{e}_2, \boldsymbol{e}_2)}_{0} + adK(\boldsymbol{e}_1, \boldsymbol{e}_2) + cb \underbrace{K(\boldsymbol{e}_2, \boldsymbol{e}_1)}_{-K(\boldsymbol{e}_1, \boldsymbol{e}_2)} \\ =& (ad - cb) K(\boldsymbol{e}_1, \boldsymbol{e}_2). \end{aligned} \tag{3.5}$$

□

3.2.2 The determinant in general

In $\mathbb{R}^2$ the determinant can be described equivalently in geometric terms as oriented volume, or in algebraic terms, as the unique alternating bilinear form that assigns $+1$ to the standard basis. For general vector spaces $\mathbb{F}^n$, the notions of "orientation" or "volume" might not exist, but the algebraic approach does generalize.

Alternating forms

A function

$$K : \underbrace{V \times \cdots \times V}_{k \text{ times}} \to \mathbb{F}$$

is a *multilinear form* if it is linear in each argument. A multilinear form is *alternating* if it changes the sign whenever two arguments are swapped. Let $\bigwedge^k(\mathbb{F}^n)$ be the space of alternating k-linear forms on $\mathbb{F}^n$.

Permutations

In Eq. (3.3), we used the alternating property to map $K(\boldsymbol{e}_2, \boldsymbol{e}_1)$ to $-K(\boldsymbol{e}_1, \boldsymbol{e}_2)$. Below, we'll be dealing with expressions of the form $K(\boldsymbol{e}_{j_1}, \ldots, \boldsymbol{e}_{j_n})$. To track the sign changes incurred when ordering the arguments, we need the concept of "permutations" and their "sign".

A *permutation* π of n letters is a re-arrangement $1 \mapsto \pi_1, \ldots, n \mapsto \pi_n$ of $(1, \ldots, n)$. The set of all such permutations is S_n, the *symmetric group on n letters*. The *sign* of a permutation is a number $\mathrm{sign}(\pi) \in \pm 1$ computed as follows. Take π and sort its entries iteratively, for i ranging from 1 to n: If $\pi_i \neq i$, find k such that $\pi_k = i$ and exchange $i \leftrightarrow k$. Then $\mathrm{sign}\,\pi = +1$ if an even number of swaps occurred, and -1 else.

Example.

$$\begin{aligned} (2,3,1) \to (1,3,2) \to (1,2,3) \quad &\Rightarrow \quad \mathrm{sign}(2,3,1) = +1. \\ (2,1,3) \to (1,2,3) \quad &\Rightarrow \quad \mathrm{sign}(2,1,3) = -1. \end{aligned}$$

The general definition

1. There is a unique alternating n-form on $\mathbb{F}^n$ that assigns $+1$ to the standard basis. Def.: This form is the *determinant*. The determinant of an $n \times n$ matrix is the determinant of its columns.

2. The determinant satisfies the *Leibniz formula*

$$\det(\boldsymbol{A}) = \sum_{\pi \in S_n} (\mathrm{sign}\,\pi)\, a_{1\pi_1} \ldots a_{n\pi_n}. \tag{3.6}$$

3. Any alternating n-form $K \in \bigwedge^n(\mathbb{F}^n)$ is proportional to it

$$K(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) = \det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)\, K(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n). \tag{3.7}$$

Proof. From Eq. (3.6), $\det$ is multilinear, alternating, and evaluates to 1 on the standard basis. Now use multilineraity and the alternating property to find

$$\begin{aligned} K\Big(\sum_{j_1} a_{1j_1}\boldsymbol{e}_{j_1}, \ldots, \sum_{j_n} a_{nj_n}\boldsymbol{e}_{j_n}\Big) &= \sum_{j_1,\ldots,j_n} a_{1j_1} \ldots a_{nj_n} K\big(\boldsymbol{e}_{j_1}, \ldots \boldsymbol{e}_{j_n}\big) \\ &= \sum_{\pi \in S_n} (\mathrm{sign}\,\pi)\, a_{1\pi_1} \ldots a_{n\pi_n} K\big(\boldsymbol{e}_1, \ldots \boldsymbol{e}_n\big). \end{aligned} \tag{3.8}$$

□

Remark. The size of the symmetric group is $n! = n \cdot (n-1) \ldots 2 \cdot 1$, which is *ginormous*. For $n \simeq 60$, the number of terms in (3.6) exceeds the number of hadrons in the universe. Thus, (3.6) is useful for mathematical reasoning, but does not provide a practical algorithm for actually computing $\det$. See Sec. 3.6.2 for an efficient algorithm.

3.2.3 Properties of the determinant

Most importantly, $\det$ turns products of matrices into the product of numbers:

$$\det(\boldsymbol{A}\,\boldsymbol{B}) = \det \boldsymbol{A} \det \boldsymbol{B}. \tag{3.9}$$

Proof. Write

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{Ab}_1, \ldots, \boldsymbol{Ab}_n) =: K(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n).$$

Then K is is multilinear and alternating, so that (3.7) implies (3.9) . □

In particular, matrix inverses are mapped to inverse numbers:

$$1 = \det(\mathbb{1}) = \det\left(\boldsymbol{T}\,\boldsymbol{T}^{-1}\right) = \det\left(\boldsymbol{T}\right)\det\left(\boldsymbol{T}^{-1}\right) \tag{3.10}$$

$$\Rightarrow \qquad \det\left(\boldsymbol{T}^{-1}\right) = [\det\left(\boldsymbol{T}\right)]^{-1}. \tag{3.11}$$

It follows that $\det$ is indeed invariant under similarity transformations:

$$\det\left(\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}\right) = \det\left(\boldsymbol{T}\right)\det\left(\boldsymbol{A}\right)\det\left(\boldsymbol{T}^{-1}\right) = \det(\boldsymbol{A}).$$

Like we did for the trace, one can therefore define the determinant of an abstract $L : V \to V$ as the determinant of any matrix representation $\det L := \det\left(\boldsymbol{\phi}_{\mathcal{B}}(L)\right)$.

Another important consequence:

$$\det(\boldsymbol{A}) \neq 0 \qquad \Leftrightarrow \qquad \boldsymbol{A} \text{ is invertible.} \tag{3.12}$$

Proof. If $\boldsymbol{A}$ is invertible, then $\det(\boldsymbol{A})\det(\boldsymbol{A}^{-1}) = 1$, hence $\det(\boldsymbol{A}) \neq 0$. If $\boldsymbol{A}$ is not invertible, then there is a non-zero $\boldsymbol{v}$ such that $\boldsymbol{Av} = 0$. Let i be such that $v_i \neq 0$. Then

$$\begin{aligned}\det(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i, \ldots, \boldsymbol{a}_n) &= v_i^{-1}\det\Big(\boldsymbol{a}_1, \ldots, \sum_j v_j\boldsymbol{a}_j, \ldots, \boldsymbol{a}_n\Big)\\ &= v_i^{-1}\det(\boldsymbol{a}_1, \ldots, 0, \ldots, \boldsymbol{a}_n) = 0.\end{aligned}$$

□

Remarks & Exercises.

1. $\det(\lambda\boldsymbol{A}) = \lambda^n\det(\boldsymbol{A})$. (*Not* $\lambda\det(\boldsymbol{A})$, a common mistake!)
2. From the Leibniz formula, $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^t)$.
3. The determinant of a diagonal matrix is the product of the diagonal elements: $\det\left(\operatorname{diag}(a_i)\right) = a_1 \ldots a_n$.
4. Slightly less obvious: the determinant of an *upper triangular matrix* is also the product of the diagonal elements.
5. For $n = 3$, the *triple product* (*Spatprodukt*) formula links $\det$ to the *cross product*:

$$\det(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \langle \boldsymbol{a} \times \boldsymbol{b}, \boldsymbol{c}\rangle, \quad \text{where} \quad \begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix} \times \begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix} = \begin{pmatrix}a_2b_3 - a_3a_2\\a_3b_1 - a_1b_3\\a_1b_2 - a_2b_2\end{pmatrix}.$$

6. The array of numbers $\epsilon_{j_1\ldots j_n} := \det\left(\boldsymbol{e}_{j_1}, \ldots \boldsymbol{e}_{j_n}\right)$ is called the *Levi-Civita symbol*. Expanding as in Eq. (3.8),

$$\det(\boldsymbol{A}) = \sum_{j_1,\ldots,j_n} a_{1j_1}\ldots a_{nj_n}\det\left(\boldsymbol{e}_{j_1}, \ldots \boldsymbol{e}_{j_n}\right) = \sum_{j_1,\ldots,j_n} a_{1j_1}\ldots a_{nj_n}\epsilon_{j_1\ldots j_n}.$$

3.2.4 Volume and orientation in higher dimensions

In $\mathbb{R}^n$, one *defines* volume and orientation of a prallelepiped in terms of the determinant:

$$\begin{aligned}\operatorname{ori}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n) &:= \operatorname{sign}(\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)),\\ \operatorname{vol}(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n) &:= |\det(\boldsymbol{a}_1,\dots,\boldsymbol{a}_n)|.\end{aligned}$$

The volume of an arbitrary body in $\mathbb{R}^n$ is defined by approximating it by parallelepipeds (Fig. **??**).

From this point of view, the product formula

$$\det(\boldsymbol{A}\boldsymbol{b}_1,\dots,\boldsymbol{A}\boldsymbol{b}_n) = \det(\boldsymbol{A})\det(\boldsymbol{b}_1,\dots,\boldsymbol{b}_n)$$

says that

> *multiplying with the matrix $\boldsymbol{A}$ changes* volume *by a factor of* $|\det(\boldsymbol{A})|$,
> *multiplying with the matrix $\boldsymbol{A}$ changes* orientation *by* $\operatorname{sign}(\det(\boldsymbol{A}))$.

Examples.

1. Rotations preserve volume and orientation. Reflections about a coordinate plane preserve volume, but invert orientation. And indeed:

$$\det\begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix} = \cos^2\phi+\sin^2\phi = 1, \qquad \det\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = -1.$$

We can now outline what determinants have to do with integration (though the detailed theory is not part of this course). Assume we want to integrate $f:\mathbb{R}^n\to\mathbb{R}$ over a region $R\subset\mathbb{R}^n$. It might be that there is a coordinate change $\boldsymbol{T}$ such that $\boldsymbol{T}(R) = \{\boldsymbol{T}\boldsymbol{x} \mid \boldsymbol{x}\in R\}$ is easier to describe. Rewriting the integral in terms of the new coordinates, we have to insert a factor of $|\det\boldsymbol{T}|^{-1}$ to compensate for the distortion of volume under $\boldsymbol{T}$:

$$\int_R f(\boldsymbol{x})\,\mathrm{d}^n\boldsymbol{x} = \int_{\boldsymbol{T}(R)} f(\boldsymbol{T}^{-1}\boldsymbol{y})\,|\det\boldsymbol{T}|^{-1}\,\mathrm{d}^n\boldsymbol{y}.$$

3.3 Eigenvectors and eigenvalues

Let $L:V\to V$ be a linear map. An *eigenvector* of L is a non-zero $v\in V$ such that

$$Lv = \lambda v \tag{3.13}$$

for some $\lambda\in\mathbb{F}$. The number λ is the *eigenvalue* of L associated with v.

Examples.

1. The eigenvectors of a diagonal matrix are the standard basis vectors $\{\boldsymbol{e}_i\}$. The eigenvalue associated with $\boldsymbol{e}_i$ is the ith diagonal element. Every non-zero vector is an eigenvector of $\mathbb{1}$ with eigenvalue 1, and an eigenvector of the null matrix with eigenvalue 0.
2. A non-trivial rotation on $\mathbb{R}^2$ does not have any eigenvectors.
3. The eigenvectors of $\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$ are the vectors of the form $\boldsymbol{v} = \left(\begin{smallmatrix}\lambda\\0\end{smallmatrix}\right)$ for $\lambda\neq 0$.

4. On the space of differentiable functions $\mathbb{R} \to \mathbb{C}$, the eigenvectors of ∂_x are of the form $f(x) = \mu e^{\lambda x}$ for $\mu \neq 0$. Functions representing physical quantities are usually *bounded* in the sense that for every f, there is a $C \in \mathbb{R}$ s.t. $\forall x, |f(x)| \leq C$. The eigenfunctions above are bounded iff $\lambda = i|\lambda|$ is purely imaginary, in which case f is a plane wave.
5. The map $L + \mu\mathbb{1}$ has the same eigenvectors as L, but all eigenvalues are shifted by μ.

While not every linear map has an eigenvector, some have entire bases of them! An *eigenbasis* for $L : V \to V$ is a basis $\mathcal{B} = \{v_i\}$ for V consisting of eigenvectors

$$Lv_i = \lambda_i v_i \qquad \forall\, i.$$

This concept is of great importance, because the representation of L w.r.t. an eigenbasis is extremely simple. In particular, if V is finite-dimensional, we immediately have

$$\mathcal{B} \text{ is an eigenbasis} \qquad \Leftrightarrow \qquad \phi^{\mathcal{B}}_{\mathcal{B}}(L) \text{ is diagonal.}$$

In this case, the diagonal elements of $\phi_{\mathcal{B}}(L)$ are nothing but the eigenvalues of L. If L has an eigenbasis, it is said to be *diagonalizable*.

Let's make this concrete for the case where L is given by a square matrix $\boldsymbol{A}$. Because

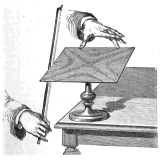
$$\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i \quad \Leftrightarrow \quad \boldsymbol{A}(\boldsymbol{T}^{-1}\boldsymbol{e}_i) = \lambda_i(\boldsymbol{T}^{-1}\boldsymbol{e}_i),$$

$\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}$ is diagonal iff the columns of $\boldsymbol{T}^{-1}$ are an eigenbasis of $\boldsymbol{A}$.

Why this* really *matters. Given a linear L, it can be quite difficult to find an eigenbasis (assuming it exists). But once we have it, most problems involving L become trivial. For example:

- To decide whether L is invertible: Just check whether 0 is among the eigenvectors.
- To compute the inverse of L: Just invert the eigenvalues.
- To solve the linear equation $Lu = f$: For the representations $\boldsymbol{f} = \boldsymbol{\phi}^{\mathcal{B}}(f)$, $\boldsymbol{u} = \boldsymbol{\phi}^{\mathcal{B}}(u)$ w.r.t. the eigenbasis, the solution is just $u_i = \lambda_i^{-1} f_i$.

Also, eigenvectors/eigenvalues often have direct physical interpretations. Examples:



- The frequencies of the harmonics of a musical instrument are eigenvalues of a mechanical differential operator. The zeros of the eigenvectors can be visualized as Chladni figures.
- The stable axes of rotation of a rigid body are the eigenvectors of the inertia tensor associated with the largest & smallest eigenvalues.
- The spectral lines of atoms are the differences between eigenvalues of the quantum mechanical energy operator.

Examples.

- The *Fibonacci numbers* are defined by the recursion relation $F_{k+2} = F_{k+1} + F_k$ with initial conditions $F_0 = 0, F_1 = 1$. Apparently, they describe mating rabbits... whatever. To understand the asymptotic behavior, write the relation in matrix form

$$\underbrace{\begin{pmatrix} F_{k+1} \\ F_k \end{pmatrix}}_{\boldsymbol{f}_k} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}}_{\boldsymbol{F}} \underbrace{\begin{pmatrix} F_k \\ F_{k-1} \end{pmatrix}}_{\boldsymbol{f}_{k-1}} \quad \Rightarrow \quad \boldsymbol{f}_k = \boldsymbol{F}^k \boldsymbol{f}_0, \quad \boldsymbol{f}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

  The powers of the *transfer matrix* $\boldsymbol{F}$ are best computed in its eigenbasis $\{\boldsymbol{v}_+, \boldsymbol{v}_-\}$:

$$\boldsymbol{f}_k = \boldsymbol{F}^k(c_+\boldsymbol{v}_+ + c_-\boldsymbol{v}_-) = c_+\lambda_+^k \boldsymbol{v}_+ + c_-\lambda_-^k \boldsymbol{v}_-, \quad \boldsymbol{f}_0 = c_+\boldsymbol{v}_+ + c_-\boldsymbol{v}_-, \quad (3.14)$$

  where $c_\pm$ are the expansion coefficients of $\boldsymbol{f}_0$ w.r.t. the eigenbasis. Thus, one only has to find the eigenbasis of $\boldsymbol{F}$ to get a closed formula for F_k! Details are an exercise.

- Eigenbases allow us to decouple systems of linear differential equations. This is one of the major applications in physics.

  First, consider a single pendulum. The restoring force is

  $$F = -mg\sin(\phi).$$

  For small deflections, the motion is approximately linear, with horizontal distance from equilibrium given by $r = l\sin(\phi)$. Newton's equation $m\ddot{r} = F$ are thus solved by

  $$\ddot{r} = -\frac{g}{l}r \qquad \Rightarrow \qquad r(t) = A\sin(\omega t + \phi_0), \qquad \omega = \sqrt{\frac{g}{l}}.$$

  Now consider two coupled pendulums. Writing the Hook constant as $k = m\kappa$, the forces are

  $$F_1 = -m\omega^2 r_1 + m\kappa(r_2 - r_1),$$
  $$F_1 = -m\omega^2 r_2 + m\kappa(r_1 - r_2).$$

  so that Newton's equations can be written in matrix form

  $$\begin{pmatrix}\ddot{r}_1\\ \ddot{r}_2\end{pmatrix} = \begin{pmatrix}-\omega^2-\kappa & \kappa\\ \kappa & -\omega^2-\kappa\end{pmatrix}\begin{pmatrix}r_1\\ r_2\end{pmatrix}.$$

  Now we are in trouble. To solve for $r_1(t)$, need to know $r_2(t)$, which in turn depends on $r_1(t)$. It's not obvious how to get out of this cyclic dependency.

  However, *any* system of the form $\ddot{\boldsymbol{r}} = \boldsymbol{M}\boldsymbol{r}$ can easily be solved if $\boldsymbol{M}$ has an eigenbasis $\{\boldsymbol{b}_i\}$. Just expand $\boldsymbol{r} = \sum_i c_i\boldsymbol{b}_i$, plug it into $\ddot{\boldsymbol{r}} = \boldsymbol{M}\boldsymbol{r}$ to get

  $$\sum_i \ddot{c}_i\boldsymbol{b}_i = \sum_i \lambda_i c_i \boldsymbol{b}_i \qquad \Leftrightarrow \qquad \ddot{c}_i = \lambda_i c_i,$$

  where the final equivalence follows from the uniqueness of basis expansions. If all eigenvalues are negative, this is just a system of uncoupled harmonic oscillators with frequencies $\omega_i = \sqrt{-\lambda_i}$. Hence, the solution to the system is

  $$\boldsymbol{r}(t) = \sum_i A_i\sin(\omega_i + \phi_i)\,\boldsymbol{b}_i.$$

  The ω_i are called the *eigenfrequencies* of the system, and the $\boldsymbol{b}_i$ the *eigenmodes*.

  Back to the example. It's easy to guess eigenvectors:

  $$\begin{pmatrix}-\omega^2-\kappa & \kappa\\ \kappa & -\omega^2-\kappa\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}-\omega^2\\-\omega^2\end{pmatrix} = -\omega^2\begin{pmatrix}1\\1\end{pmatrix} =: \lambda_1\boldsymbol{b}_1,$$
  $$\begin{pmatrix}-\omega^2-\kappa & \kappa\\ \kappa & -\omega^2-\kappa\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}-(\omega^2+2\kappa)\\(\omega^2+2\kappa)\end{pmatrix} = -(\omega^2+2\kappa)\begin{pmatrix}1\\-1\end{pmatrix} =: \lambda_2\boldsymbol{b}_2.$$

  Physical interpretation: The vector $\boldsymbol{b}_1$ is the *center of mass mode*. Here, the two particles swing in unison. Because the frequency of small oscillations is independent of mass, the eigenfrequency of the c.o.m. mode is just ω. The vector $\boldsymbol{b}_2$ describes an opposing motion of the two masses. Because they have to work both against gravity and against the Hook constant, the eigenfrequency is higher.

3.3.1 Eigenspaces

If v, w are eigenvectors of L for the same eigenvalue λ, then

$$L(v + \mu w) = Lv + \mu Lw = \lambda(v + \mu w).$$

Therefore, the set of all eigenvectors for λ, together with the null vector, form a subspace V_λ of V. It is called the *eigenspace* for λ.

Rearranging the *eigenvalue equation* (3.13) gives

$$Lv = \lambda v \quad \Leftrightarrow \quad (\lambda \mathbb{1} - L)v = 0 \quad \Leftrightarrow \quad v \in \ker(\lambda \mathbb{1} - L) \tag{3.15}$$

and thus $V_\lambda = \ker(\lambda \mathbb{1} - L)$. The dimension $\dim V_\lambda$ is called the *(geometric) multiplicity* of λ. Eigenvalues with multiplicity larger than 1 are called *degenerate*.

Remark. If $Lv = \lambda v$, then $L(\mu v) = \lambda(\mu v)$. Thus, even if the eigenspace is non-degenerate, eigenvectors are determined at most up to a non-zero factor $\mu \in \mathbb{F}$. In Euclidean/Hermitian spaces, one often demands that eigenvectors be normalized. But this still leaves an ambiguity of ± 1 or $e^{i\phi}$, known as a "global phase" in QM.

Example

1. On $\mathbb{C}^n$, consider the projection $P_{\boldsymbol{w}} : \boldsymbol{v} \mapsto \boldsymbol{w}\langle \boldsymbol{w}, \boldsymbol{v}\rangle$ for a normalized $\boldsymbol{w}$. Extend $\mathcal{B} = \{\boldsymbol{b}_1 = \boldsymbol{w}, \boldsymbol{b}_2, \dots\}$ to an ONB. Then $P_{\boldsymbol{w}}(\boldsymbol{b}_1) = \boldsymbol{b}_1$ and $P_{\boldsymbol{w}}(\boldsymbol{b}_i) = 0$ for $i > 1$. Thus $\mathcal{B}$ is an eigenbasis. The eigenvalue 1 is non-degenerate, but 0 has multiplicity $n - 1$.
2. On the space of smooth bounded functions $\mathbb{R} \to \mathbb{C}$, the eigenvalues of $P = -i\partial_x$ are the real numbers: The eigenspace associated with $k \in \mathbb{R}$ consists of the multiples of $f_k(x) = e^{ikx}$. All eigenvalues are hence non-degenerate.

   The square $P^2 = -\partial_x^2$ has the same eigenvectors. The eigenvalue of f_k is now $E := k^2$, and thus f_k and f_{-k} belong to the same eigenspace of P^2. Therefore: $E = 0$ is non-degenerate, but the positive eigenvalues have multiplicity 2.

   In the quantum treatment of a particle in one dimension, P is associated with the particle's momentum, and P^2 with its kinetic energy. Physically, the degeneracy reflects the fact that each non-zero value of kinetic energy is compatible with two directions of motion.

Let $S = \{v_i\}$ be a set of eigenvectors with distinct eigenvalues λ_i. Then S is LI.

Proof. By induction on $|S| = k$. Assume the statement holds for $k - 1$. If

$$0 = \sum_{i=1}^{k} c_i v_i \quad \text{then} \quad 0 = (\lambda_k \mathbb{1} - L) \sum_{i=1}^{k} c_i v_i = \sum_{i=1}^{k-1} c_i \underbrace{(\lambda_k - \lambda_i)}_{\neq 0} v_i.$$

Thus $c_1 = \cdots = c_{k-1} = 0$ by the induction hypothesis, and hence $c_k = 0$ as well. □

Thus: If $\dim V = n < \infty$ and L has n distinct eigenvalues, then it has an eigenbasis.

Remark. If L has n distinct eigenvalues, there is therefore a basis, unique up multiplicative factors, labeled by the eigenvalues. This is the basis of a concise notation used in QM. E.g. "$|E\rangle$" is the eigenvector with eigenvalue E for the "energy operator" H, i.e. the essentially unique solution to $H|E\rangle = E|E\rangle$.

Going even further, one sometimes uses the same letter for both the operator and its eigenvalues. To tell them apart, a "hat" is placed on top of the operator, like so: $\hat{P}|p\rangle = p|p\rangle$. Four p's, three different meanings: "The operator $\hat{P}$ acting on the eigenvector $|p\rangle$ results in a factor of p". Got it?

3.3.2 The characteristic polynomial

In this section, assume that $\dim V = n < \infty$. By (3.15), λ is an eigenvalue iff $\det(L - \lambda\mathbb{1}) = 0$. The degree-$n$ polynomial

$$p_L : x \mapsto \det(x\mathbb{1} - L)$$

is the *characteristic polynomial* of L. Thus, the eigenvalues of L are among the roots of its characteristic polynomial.

Example. For the rotation matrices $\boldsymbol{R}_\phi$, the characteristic polynomials and its roots are

$$p_{\boldsymbol{R}_\phi}(x) = \det \begin{pmatrix} \cos\phi - x & \sin\phi \\ -\sin\phi & \cos\phi - x \end{pmatrix} = (\cos\phi - x)^2 + \sin^2\phi$$

solved by $\lambda_\pm = \cos\phi \pm i\sin\phi = e^{\pm i\phi}$. The roots are complex and thus are *not* eigenvalues of $\boldsymbol{R}_\phi$ as a linear map $\mathbb{R}^2 \to \mathbb{R}^2$.

If $\mathbb{F} = \mathbb{C}$, the characteristic polynomial factorizes

$$p_L(x) = \prod_{i=1}^{n} (x - \lambda_i), \qquad \lambda_i \in \mathbb{C}. \tag{3.16}$$

Consequences:

- By the *fundamental theorem of algebra*, every polynomial has at least one root over $\mathbb{C}$. Therefore, over $\mathbb{C}$, every L has at least one eigenvalue.
- If p_L has n distinct roots, then L has an eigenbasis.

Example. If a real matrix $\boldsymbol{A}$ has no eigenbasis in $\mathbb{R}^n$, it may have one in $\mathbb{C}^n$. For example, the choice $\boldsymbol{T} = \frac{1}{\sqrt{2}} \left(\begin{smallmatrix} 1 & i \\ 1 & -i \end{smallmatrix}\right)$ diagonalizes the 2D rotation matrices: $\boldsymbol{T}\boldsymbol{R}_\phi\boldsymbol{T}^{-1} = \mathrm{diag}(e^{i\phi}, e^{-i\phi})$.

In Hamiltonian mechanics, the dynamics of a harmonic oscillator is a rotation in *phase space*, i.e. the x-p-plane. In dimensionless coordinates: $\left(\begin{smallmatrix} x(t) \\ p(t) \end{smallmatrix}\right) = A \left(\begin{smallmatrix} \cos(-\omega t + \phi_0) \\ \sin(-\omega t + \phi_0) \end{smallmatrix}\right)$. By the above (or directly by Euler's formula), the dynamics is diagonalized by the complex coordinates

$$\begin{pmatrix} a(t) \\ a^\dagger(t) \end{pmatrix} := \boldsymbol{T} \begin{pmatrix} x \\ p \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} x + ip \\ x - ip \end{pmatrix} = \begin{pmatrix} a(0)\, e^{-i\omega t} \\ a^\dagger(0)\, e^{i\omega t} \end{pmatrix}.$$

This formulation is central to the "algebraic solution" of the quantum harmonic oscillator.

The number of times a given value λ appears in the decomposition (3.16) is the *algebraic multiplicity* of the eigenvalue. Counting and using LI of different eigenspaces:

- If for all eigenvalues, the geometric multiplicity equals the algebraic multiplicty, then L has an eigenbasis.

3.4 The Jordan Normal Form (and why it doesn't matter)

The geometric multiplicity can be strictly smaller than the algebraic one. For example,

$$\boldsymbol{N} := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{F}^{m\times m}, \quad \det(x\mathbb{1} - \boldsymbol{N}) = x^m. \tag{3.17}$$

It "shifts components up, inserting 0 at the bottom". The only eigenvalue is 0, and the only vectors that get annihilated are the multiplies of $\boldsymbol{e}_1$. Thus the geometric multiplicity is 1, compared to m for the algebraic one.

It then follows that

$$\boldsymbol{J} = \lambda \mathbb{1} + \boldsymbol{N} := \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}, \quad \det(x\mathbb{1} - \boldsymbol{J}) = (x-\lambda)^m. \tag{3.18}$$

also has geometric multipicty 1, with λ being the unique eigenvalue. A matrix of the form (3.18) is called a *Jordan block*. It is specified by an *eigenvalue* λ and *multiplicity* m.

Jordan blocks are "the most general" examples of a non-diagonalizable matrix, in the following sense:

Jordan Normal Form (JNF): Let V be over $\mathbb{C}$, $\dim V < \infty$, and $L : V \to V$. If the roots of p_L are elements of $\mathbb{F}$, then there exists a basis $\mathcal{B}$ such that $\phi^{\mathcal{B}}_{\mathcal{B}}(L)$ is a *block matrix*

$$\phi^{\mathcal{B}}_{\mathcal{B}}(L) = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix},$$

with each J_i a Jordan block. The map L has an eigenbasis iff all Jordan blocks are 1×1. The JNF is unique up to the order of the blocks.

Remark. The JNF is a spectacular mathematical result. It says that, up to choosing a basis, a linear map is specified exactly by a set of eigenvalues and multiplicites $\{(\lambda_i, m_i)\}$.

It also doesn't matter much in practice.

One reason is that a matrix can fail to be diagonalizable over $\mathbb{C}$ only if two roots of the characteristic polynomial fall together. But "$\lambda_i = \lambda_j$" is an unstable condition: The tiniest bit of noise will cause all roots to become distinct. It thus doesn't make sense to compute the JNF of matrices whose entries represent noisy real-world quantities. Numerical computer programs don't usually even provide functions for the JNF, because floating point rounding errors would lift the multiplicities (though computer *algebra* systems can find the JNF for matrices with exact coefficients, e.g. specified as rational numbers).

Another reason is that many classes of matrices treated in physics are automatically diagonalizable, degeneracies or not. For example: Call a matrix *normal* $\boldsymbol{A}$ if it commutes with its adjoint $[\boldsymbol{A}, \boldsymbol{A}^\dagger] = 0$. Examples include real symmetric matrices or complex self-adjoint ones. We will show in Sec. 3.7 that all normal matrices can be diagonalized.

For these reasons, the JNF isn't very prominent in physics.

Example. From (3.17), $\boldsymbol{N}^m = 0$. A matrix with the property that a finite power is the null matrix is called *nilpotent*. An example of a nilpotent matrix in physics is the Fermionic creation operator $a_i^\dagger$, which "creates a particle in the ith state". It fulfills $(a_i^\dagger)^2 = 0$. Why? That's *Pauli's exclusion principle* at work: There can be only one Fermion in a given state.

3.5 Functions of linear maps

Linear maps form an algebra: It makes sense to take powers and linear combinations of them. Thus, if $p(X) = \sum_{k=0}^{d} c_k X^k$ a polynomial, it maps linear maps to linear maps via

$$p(L) = \sum_{k=0}^{d} c_k L^k.$$

The same applies to functions $f(x) = \sum_{k=0}^{\infty} c_k x^k$ that have a series expansion, assuming that no convergence issues arise.

> ***Remark***. Under no circumstances should you confuse this with an application of p to matrix elements. For example, for $p(X) = X^2$ and L given by the $\boldsymbol{\sigma}_x$ Pauli matrix,
>
> $$\begin{pmatrix} p(0) & p(1) \\ p(1) & p(0) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \neq p(\boldsymbol{\sigma}_x) = \boldsymbol{\sigma}_x^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On the other hand, if $Lv = \lambda v$, then

$$p(L)v = \sum_i c_i L^i v = \sum_i c_i \lambda^i v = p(\lambda_i)v.$$

That is: When applying L to eigenvectors, we *can* replace the possibly quite complicated "function of linear map" with the generally much simpler "function of scalar". And if L has an eigenbasis $\{v_i\}$, then $p(L)$ is just the linear extension of $v_i \mapsto p(\lambda_i)v_i$.

Now let $f : \mathbb{F} \to \mathbb{F}$ be any function, not necessarily polynomial. For a linear map L with eigenbasis $\{v_i\}$, *define* $f(L)$ as the linear extension of $v_i \mapsto f(\lambda_i)v_i$. This convention is known as the *functional calculus*.

In particular, we get this recipe for evaluating $f(\boldsymbol{A})$ for a diagonalizable matrix $\boldsymbol{A}$:

Find $\boldsymbol{T}$ s.t. $\boldsymbol{A} = \boldsymbol{T}^{-1} \operatorname{diag}(\lambda_i)\boldsymbol{T}$. Then $f(\boldsymbol{A}) = \boldsymbol{T}^{-1} \operatorname{diag}(f(\lambda_i))\boldsymbol{T}$.

3.5.1 Matrix exponentials and linear differential equations

By far the most important application is to the exponential function $e^t = \sum_{k=0}^{\infty} \frac{1}{k!} t^k$. From the series, one directly verifies that $e^{\lambda t}$ satisfies the differential equation $f'(t) = \lambda f(t)$ with boundary condition $f(0) = 1$.

Essentially the same argument applies to matrices:

$$\boldsymbol{M}'(t) = \boldsymbol{\Lambda}\boldsymbol{M}(t), \quad \boldsymbol{M}(0) = \mathbb{1} \qquad \Leftrightarrow \qquad \boldsymbol{M}(t) = e^{t\boldsymbol{\Lambda}} = \sum_{k=0}^{\infty} \frac{1}{k!}(t\boldsymbol{\Lambda})^k. \tag{3.19}$$

While this looks pretty abstract, this type of differential equation is extremely important. If the following list doesn't convince you, I don't know what to say:

- It describes the time evolution of any quantum system.
- It describes the behavior of continuous symmetries.
- It describes the time evolution of "linear" systems in classical mechanics.
- It approximately describes the behavior of any mechanical system for short times.

Recall how we solved coupled differential equations by writing them in matrix form an passing to an eigenbasis decomposition? Taking exponentials of matrices allows for a more compact representation of this technique. We'll have a brief look at this next.

Dynamical systems

Consider a *dynamical system* whose state is described by a vector $\boldsymbol{x}(t) \in \mathbb{R}^n$ subject to the equations of motion

$$\dot{\boldsymbol{x}} = \boldsymbol{G}(\boldsymbol{x}) \qquad \boldsymbol{G} : \mathbb{R}^n \to \mathbb{R}^n.$$

It is called *linear* there is an $n \times n$-matrix $\boldsymbol{\Lambda}$ such that $\boldsymbol{G}(\boldsymbol{x}) = \boldsymbol{\Lambda}\boldsymbol{x}$. The solution is (exercise!)

$$\boldsymbol{x}(t) = e^{t\boldsymbol{\Lambda}}\boldsymbol{x}(0). \tag{3.20}$$

Sometimes, the powers $\boldsymbol{\Lambda}^k$ and hence the series expansion $e^{t\boldsymbol{\Lambda}}$ can be computed explicitly. Alternatively, expanding in eigenbasis of $\boldsymbol{\Lambda}$, we get the solution

$$\boldsymbol{x}(0) = \sum_i c_i(0)\boldsymbol{b}_i \quad \Rightarrow \quad \boldsymbol{x}(t) = \sum_i c_i(0)e^{t\lambda_i}\boldsymbol{b}_i.$$

Linear system in mechanics

Let's apply this to a mechanical system, whose configuration is given by a point $\boldsymbol{r} \in \mathbb{R}^n$. Now Newton's equations involve *second* derivatives. But they can easily be re-expressed in terms of first-order time derivatives by introducing auxiliary variables. Indeed, with every coordinate r_i, associate its velocity $v_i = \dot{r}_i$. Let $\boldsymbol{x} = (\boldsymbol{r}, \boldsymbol{v}) \in \mathbb{R}^{2n}$ be the *phase space vector* that describes the position and velocities. Assume that the force

$$\boldsymbol{F}(\boldsymbol{r}, \boldsymbol{v}) = \boldsymbol{F}\boldsymbol{x} = \begin{pmatrix} \boldsymbol{F}_r & \boldsymbol{F}_v \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{v} \end{pmatrix}$$

is a linear function of the positions and the coordinates. Then

$$\dot{\boldsymbol{x}} = \begin{pmatrix} \dot{\boldsymbol{r}} \\ \dot{\boldsymbol{v}} \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{0} & \mathbb{1} \\ \boldsymbol{F}_r & \boldsymbol{F}_v \end{pmatrix}}_{\boldsymbol{\Lambda}} \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{v} \end{pmatrix},$$

and so we can solve it using (3.20). An example is on the exercise sheet.

Stability analysis

Now consider a mechanical system with forces $\boldsymbol{F}(\boldsymbol{r}, \boldsymbol{v})$ that are not necessarily linear. An *equilibrium point* $\boldsymbol{x}_0 := (\boldsymbol{r}_0, \boldsymbol{v}_0)$ is one where the forces vanish

$$\boldsymbol{F}(\boldsymbol{r}_0, \boldsymbol{v}_0) = 0.$$

To analyze the behavior of small perturbations around equilibrium, we may replace $\boldsymbol{F}$ by its first-order Taylor expansion, i.e. its Jacobi matrix $\boldsymbol{J}$. Then the dynamics of the perturbations $\boldsymbol{\delta} = \boldsymbol{x} - \boldsymbol{x}_0$ is governed by the approximate equations of motion $\dot{\boldsymbol{\delta}} \simeq \boldsymbol{J}\boldsymbol{\delta}$. Using (3.20), we find that the equilibrium is stable iff $\operatorname{Re}\lambda_i < 1$ for all i, and unstable iff $\operatorname{Re}\lambda_i > 1$ for at least one i. In this context, the λ_i, describing the reaction of a system to small perturbations, are called *Lyapunov exponents*.

Example. Here, we revisit the *damped harmonic oscillator* in the language of linear algebra. It has velocity-dependent forces

$$F(r,v) = -\omega_0^2 r - \gamma v = \begin{pmatrix} -\omega_0^2 & -\gamma \end{pmatrix} \begin{pmatrix} r \\ v \end{pmatrix}.$$

Let's compute the Lyaponuv exponents around the equilibrium point $\boldsymbol{x}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$:

$$\dot{\boldsymbol{x}} = \begin{pmatrix} \dot{\boldsymbol{r}} \\ \dot{\boldsymbol{v}} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{pmatrix}}_{\boldsymbol{\Lambda}} \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{v} \end{pmatrix} \quad \Rightarrow \quad \det(x\mathbb{1} - \boldsymbol{\Lambda}) = x^2 + \gamma x - \omega_0^2$$

and so the eigenvalues are

$$\lambda_\pm = -\frac{1}{2}\gamma \pm \frac{1}{2}\sqrt{\gamma^2 - 4\omega_0^2}$$

which can be split into real and imaginary parts:

$$\operatorname{Re}\lambda_\pm = \begin{cases} 0 & \gamma = 0 \\ -\frac{1}{2}\gamma & \gamma \le 2\omega_0 \\ \lambda_\pm & \gamma > 2\omega_0 \end{cases}, \qquad \operatorname{Im}\lambda_\pm = \begin{cases} \pm\omega_0 & \gamma = 0 \\ \pm\sqrt{\omega_0^2 - \gamma^2/4} & \gamma \le 2\omega_0 \\ 0 & \gamma > 2\omega_0 \end{cases}.$$

Interpretation: For $\gamma > 0$, the real part of the eigenvalues is strictly negative. We recover the unsurprising result that the equilibrium is stable in the presence of a damping factor γ: Deviations are exponentially suppressed with Lyapunov exponent $-\operatorname{Re}\lambda_\pm$. If $\gamma \le 2\omega_0$, the exponential decay modulates oscillations with a frequency $\operatorname{Im}\lambda_\pm$ that is reduced as compared to the $\gamma = 0$ case. For $\gamma > 2\omega$, no oscillations take place anymore.

3.6 Algorithms

A key ingredient to the success of linear algebra is the existence of comparatively fast and stable numerical algorithms for many important problems.

3.6.1 Iterative algorithms for finding eigenvalues and eigenvectors

See Exercise 8.

3.6.2 Gaussian elimination

Gaussian elimination is an efficient algorithm for solving inhomogeneous linear equations

$$\text{solve} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \quad \text{for } \boldsymbol{x} \in \mathbb{F}^n, \quad \text{given } \boldsymbol{A} \in \mathbb{F}^{m\times n}, \boldsymbol{y} \in \mathbb{F}^n.$$

Recall from Sec. (2.5.2) that the set of solutions is an affine space

$$\{\boldsymbol{x} \,|\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\} = \{\boldsymbol{x}_0 + \boldsymbol{s} \,|\, \boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{y}, \boldsymbol{s} \in \ker \boldsymbol{A}\}.$$

By "solve", we mean: find a *particular solution* $\boldsymbol{x}_0$ together with a basis $\{\boldsymbol{s}_i\}_i$ of the directional vector space $\ker \boldsymbol{A}$.

Some applications of linear equations:

1. **Eigenvectors from eigenvalues**. If λ is an eigenvector of a matrix $\boldsymbol{A}$, then the corresponding eigenspace is the set of solutions of $(\lambda\mathbb{1} - \boldsymbol{L})\boldsymbol{x} = 0$.

2. **Test for linear independence**. The vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ are linearly independent if and only if the only solution to the equation $\boldsymbol{A}\boldsymbol{x} = 0$ is $\boldsymbol{x} = 0$. Here, $\boldsymbol{A}$ is the matrix with the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ as its columns.

3. **Inverting matrices**. Given a matrix $\boldsymbol{A}$, a matrix $\boldsymbol{B}$ with columns $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ is its inverse if and only if $\boldsymbol{A}\boldsymbol{b}_i = \boldsymbol{e}_i$ for $i = 1, \ldots, n$.

4. **Basis expansions**. Given a basis $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ and a vector $\boldsymbol{y} \in \mathbb{F}^n$, the basis representation $\boldsymbol{c} = \boldsymbol{\phi}^{\mathcal{B}}(\boldsymbol{f})$ is the solution to $\boldsymbol{B}\boldsymbol{c} = \boldsymbol{y}$, where $\boldsymbol{B}$ has columns $\boldsymbol{b}_i$.

We'll see that Gaussian elimination can also be used to compute the determinant of a matrix without having to sum over $n!$ terms.

Gaussian elimination has a status not unlike long division. It's a fundamental algorithm, that's typically taught to students, but, these days, performed almost exclusively by computers.[1] For this reason, we'll treat it only briefly here.

Start with a linear equation

$$\begin{aligned} A_{1,1}u_1 + \cdots + A_{1,n}u_n &= y_1 \\ &\vdots \\ A_{m,1}u_1 + \cdots + A_{m,n}u_n &= y_m. \end{aligned}$$

First, we'll explain how to solve it in the spatial case where $m = n$ and assuming that certain coefficients are non-zero. We'll remark on the general case later. Our running example will be

$$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 2, \\ 3x_1 + 11x_2 - 9x_3 &= 4, \\ -x_1 + x_2 + 6x_3 &= 5. \end{aligned}$$

To save ink, there's a common shorthand notation

$$\left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 3 & 11 & -9 & 4 \\ -1 & 1 & 6 & 5 \end{array}\right),$$

where variable names and equality signs are suppressed.

Step 1: Row reduction. Our goal is to transform the system into an equivalent one that is in *row echolon form*. This form is simple to visualize, but awkward to verbalize. Examples:

$$\left(\begin{array}{ccc|c} \tilde{A}_{1,1} & \tilde{A}_{1,2} & \tilde{A}_{1,3} & \tilde{y}_1 \\ 0 & \tilde{A}_{2,2} & \tilde{A}_{2,3} & \tilde{y}_2 \\ 0 & 0 & \tilde{A}_{3,3} & \tilde{y}_3 \end{array}\right) \quad \text{or} \quad \left(\begin{array}{ccc|c} \tilde{A}_{1,1} & \tilde{A}_{1,2} & \tilde{A}_{1,3} & \tilde{y}_1 \\ 0 & 0 & \tilde{A}_{2,3} & \tilde{y}_2 \\ 0 & 0 & 0 & \tilde{y}_3 \end{array}\right).$$

In words: If in any row, the first k coefficients are zero, then in the next row, at least the first $k + 1$ coefficients must be zero.

Every linear system can be brought row echelon form, like so:

[1] Though the current lecturer once constructed a custom variant of Gaussian elimination as part of an attack on a cryptographic witness of *quantum computational supremacy*. So knowing about the algorithm might come in handy in unexpected ways.

Assume $A_{11} \neq 0$. Add $-A_{1,2}/A_{1,1}$ times the first row to the second one. Example:

$$\left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 3 & 11 & -9 & 4 \\ -1 & 1 & 6 & 5 \end{array}\right) \quad \mapsto \quad \left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 0 & 2 & 6 & -2 \\ -1 & 1 & 6 & 5 \end{array}\right)$$

Claim: This operation does not change the solution space. Indeed, if $\boldsymbol{x}$ is a solution for the system on the left, it is also a solution of the system on the right. That's because if an equation (like $x_1 + 3x_2 - 5x_3 = 2$) is true, then any multiple of it is also true; and if two equations are true, then so is their sum. But the converse also holds, because we can go back from the right to the left in the same way: Just add $+A_{2,1}/A_{1,1}$ times the first row to the second one. The two systems are thus equivalent, as claimed.

We can apply the same logic to all rows: For $i = 2 \ldots m$: Add $-A_{i,1}/A_{1,1}$ times the first row from the ith one. In the example, there's only one more row left:

$$\left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 0 & 2 & 6 & -2 \\ -1 & 1 & 6 & 5 \end{array}\right) \quad \mapsto \quad \left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 0 & 2 & 6 & -2 \\ 0 & 4 & 1 & 7 \end{array}\right).$$

One says that the first column has been *cleared*.

Denote the coefficients of the transformed equivalent system by $\tilde{A}_{i,j}, \tilde{y}_j$.

Next, we clear the second column. Assume that $\tilde{A}_{2,2} \neq 0$. For $i = 3, \ldots, m$, add $-\tilde{A}_{i,2}/\tilde{A}_{2,2}$ times the second row from the ith one. Example:

$$\left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 0 & 2 & 6 & -2 \\ 0 & 4 & 1 & 7 \end{array}\right) \quad \mapsto \quad \left(\begin{array}{ccc|c} 1 & 3 & -5 & 2 \\ 0 & 2 & 6 & -2 \\ 0 & 0 & -11 & 11 \end{array}\right).$$

In the example, we have now reached the row echelon form. In general, one consecutively clears all columns $j = 1, \ldots, n-1$ this way.

Step 2: Back substitution A system in row echelon form is trivial to solve. Look at the last line of the example. It depends only on a single variable, so we can immediately conclude

$$-11x_3 = 11 \qquad \Leftrightarrow \qquad x_3 = -1.$$

The next-to-last line depends on two variables x_2, x_3. But we have already found x_3, and can thus solve easily for the remaining one:

$$2x_2 + 6x_3 = -2 \qquad \Leftrightarrow \qquad 2x_2 - 6 = -2 \qquad \Leftrightarrow \qquad x_2 = 2.$$

Likewise,

$$x_1 + 3x_2 - 5x_3 = 2 \qquad \Leftrightarrow \qquad x_1 + 6 + 5 = 2 \qquad \Leftrightarrow \qquad x_1 = -9$$

and we are done.

In general, a few things can turn out differently from the example:

- We could encounter a row of the form

$$\left(\begin{array}{ccc|c} 0 & \ldots & 0 & y_j \end{array}\right)$$

  for some non-zero y_j. In this case, the system clearly has no solution.

- More than one variable can disappear in a single step. For example, the row echelon form could read

$$\left(\begin{array}{ccc|c} A_{1,1} & A_{1,2} & A_{1,3} & y_1 \\ 0 & 0 & A_{2,3} & y_2 \end{array}\right).$$

  The first row gives just one equation for the two variables x_1, x_2. There are thus infinitely many solutions: To label them, introduce a parameter $t \in \mathbb{F}$ and set $x_2 = t$. Then

$$x_1 = \frac{y_1}{A_{1,1}} - t\frac{A_{1,2}}{A_{1,1}}$$

  and the full solution space is

$$\begin{pmatrix} \frac{y_1}{A_{1,1}} \\ \frac{y_2}{A_{2,3}} \end{pmatrix} + t \begin{pmatrix} \frac{A_{1,2}}{A_{1,1}} \\ 0 \end{pmatrix} \qquad t \in \mathbb{F}.$$

  As expected, it is an affine space and we have found a base point and a basis for the (one-dimensional) directional vector space.

- It can happen that a coefficient is zero, preventing us from clearing the column below it. Example: $A_{1,1}$ in

$$\left(\begin{array}{cc|c} 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right).$$

  In this case, we would swap the first and second row – also an operation that clearly does not change the solution space.

What's the time complexity of Gaussian elimination? If $\boldsymbol{L}$ is an $n \times n$ matrix, then about n columns will be cleared. Each time, about n^2 coefficients are modified, for a total of $O(n^3)$ operations. Not great, not terrible.

LU decomposition and computing the determinant

Let $\boldsymbol{A}$ be an $n \times n$ matrix. Recall that we often care about $\det \boldsymbol{A}$, but the only algorithm available to us at the moment, a literal reading of the Leibniz formula, involves $n!$ steps to compute, which quickly exceeds the number of particles in the known universe.

In addition to solving linear systems, Gaussian elimination can also be used to find determinants efficiently. The key insight is that the process of bringing the matrix $\boldsymbol{A}$ into row echelon form $\tilde{\boldsymbol{A}}$ can be described as multiplying $\boldsymbol{A}$ by a *normalized lower triangular matrix* $\boldsymbol{\Lambda}$. (*Lower triangular* means that $\Lambda_{ij} \neq 0$ only if $i > j$, and *normalized* means that $\Lambda_{ii} = 1$, see below).

Indeed, assume for now that no swaps of rows are required. In the first step, the Gaussian algorithm adds $-\frac{A_{2,1}}{A_{1,1}} =: \Lambda_{2,1}$ times the first row to the second row. This can be expressed as the matrix multiplication

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} 1 & & & \\ \Lambda_{2,1} & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \boldsymbol{A}.$$

But *all* the Gaussian algorithm ever does is adding certain multiples of the jth row to the ith one, for $j < i$. Therefore, we can express the final row echelon matrix $\boldsymbol{U}$ (for *upper triangular*) as:

$$\boldsymbol{U} = \begin{pmatrix} U_{1,1} & U_{1,2} & \dots & U_{1,n} \\ & U_{2,2} & \ddots & \vdots \\ & & \ddots & U_{n-1,n} \\ & & & U_{n,n} \end{pmatrix} = \begin{pmatrix} 1 & & & \\ \Lambda_{2,1} & 1 & & \\ \vdots & \ddots & \ddots & \\ \Lambda_{n,1} & \dots & \Lambda_{n,n-1} & 1 \end{pmatrix} \boldsymbol{A} =: \boldsymbol{\Lambda A}. \quad (3.21)$$

Now the determinant of an upper or lower triangular matrix *can* be computed using the Leibniz formula, because in this case, one easily sees that only one single term survives: the product of the main diagonal. Applying this to both sides of (3.21) gives

$$\det \boldsymbol{A} = \prod_{i=1}^{n} U_{i,i}.$$

Remark. Applying $\boldsymbol{\Lambda}^{-1} =: \boldsymbol{L}$ to both sides of (3.21) gives $\boldsymbol{A} = \boldsymbol{LU}$. The inverse matrix $\boldsymbol{L}$ is again normalized lower triangular and can be computed using one more round of Gaussian elimination. That's because $\boldsymbol{L}$ is the solution to $\boldsymbol{L\Lambda} = \mathbb{1}$, i.e. now $\boldsymbol{\Lambda}$ takes the role of the coefficient matrix that is being brought into echelon form. Thus:

$$\boldsymbol{A} = \begin{pmatrix} 1 & & & \\ L_{2,1} & 1 & & \\ \vdots & \ddots & \ddots & \\ L_{n,1} & \dots & L_{n,n-1} & 1 \end{pmatrix} \begin{pmatrix} U_{1,1} & U_{1,2} & \dots & U_{1,n} \\ & U_{2,2} & \ddots & \vdots \\ & & \ddots & U_{n-1,n} \\ & & & U_{n,n} \end{pmatrix} = \boldsymbol{LU}. \quad (3.22)$$

This form, which plays an important role in numerical analysis, is the *LU factorization* of $\boldsymbol{A}$

Remark. We arrived at Eqs. (3.21, 3.22) assuming that now row swaps are required to achieve row reduced echelon form. It is not hard to see that the general case leads to $\boldsymbol{PA} = \boldsymbol{LU}$ with $\boldsymbol{P}$ a *permutation matrix*: $\boldsymbol{P}\boldsymbol{e}_i = \boldsymbol{e}_{\pi_i}$ for some permutation $\pi \in S_n$. This form is called a *LUP decomposition*. Taking determinants gives

$$\det \boldsymbol{A} = \operatorname{sign} \pi \prod_i U_{i,i},$$

where $\operatorname{sign} \pi = (-1)^k$ for k the number of row swaps.

3.7 Normal matrices

Recall how we didn't bother too much with the Jordan normal form? One reason is that linear maps in physics often have a simple-to-check property called *normalcy*, which ensures not only that they have an eigenbasis, but even an orthonormal one!

The condition is stated in terms of the *adjoint* of a matrix. The adjoint has appeared a number of times in the lecture and the exercises. We quickly recap some of its properties.

Recap of the adjoint

Let $\boldsymbol{A}$ be a complex $n \times m$-matrix. Its *adjoint* $\boldsymbol{A}^\dagger$ is the complex conjugate of the transpose, i.e. an $m \times n$-matrix with components $A^\dagger_{ij} = \bar{A}_{ji}$. Then:

- In terms of the standard Hermitian inner product, "the adjoint acts on the left argument as the matrix does on the right":

$$\langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{w}\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} \bar{v}_i A_{ij} w_j = \sum_{i=1}^{m}\sum_{j=1}^{n} \overline{A_{ij} v_i} w_j = \langle \boldsymbol{A}^\dagger \boldsymbol{v}, \boldsymbol{w}\rangle.$$

- Taking the adjoint of a product reverses the order:

$$(\boldsymbol{A}\boldsymbol{B})^\dagger = \boldsymbol{B}^\dagger \boldsymbol{A}^\dagger.$$

- The adjoint of the adjoint is the original matrix: $(\boldsymbol{A}^\dagger)^\dagger = \boldsymbol{A}$.

- For a column vector $\boldsymbol{v} \in \mathbb{C}^n$, the adjoint $\boldsymbol{v}^\dagger$ is a row vector. As a linear functional, it equals the inner product with $\boldsymbol{v}$:

$$\boldsymbol{v}^\dagger \boldsymbol{w} = \sum_i \bar{v}_i w_i = \langle \boldsymbol{v}, \boldsymbol{w}\rangle.$$

3.7.1 Normal matrices have orthonormal eigenbases

A square matrix $\boldsymbol{A}$ is *normal* if it commutes with its adjoint $[\boldsymbol{A}, \boldsymbol{A}^\dagger] = 0$.

For concreteness, we mention two important classes of normal matrices:

- A matrix $\boldsymbol{A}$ is *Hermitian* or *self-adjoint* if $\boldsymbol{A}^\dagger = \boldsymbol{A}$. Because every matrix commutes with itself, Hermitian matrices are normal.

- A matrix $\boldsymbol{U}$ is *unitary* if $\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1}$. Because a matrix commutes with its inverse, $\boldsymbol{U}\boldsymbol{U}^{-1} = \mathbb{1} = \boldsymbol{U}^{-1}\boldsymbol{U}$, unitary matrices are normal.

Examples

- The Pauli matrices are all Hermitian. Because the square to $\mathbb{1}$, they are *also* unitary

$$\boldsymbol{\sigma}\boldsymbol{\sigma}^\dagger = \boldsymbol{\sigma}\boldsymbol{\sigma} = \mathbb{1} \qquad \Rightarrow \qquad \boldsymbol{\sigma}^\dagger = \boldsymbol{\sigma}^{-1}.$$

- If $\boldsymbol{A}$ is any $n \times m$-matrix, then

$$(\boldsymbol{A}\boldsymbol{A}^\dagger)^\dagger = (\boldsymbol{A}^\dagger)^\dagger \boldsymbol{A}^\dagger = \boldsymbol{A}\boldsymbol{A}^\dagger$$

  so $\boldsymbol{A}\boldsymbol{A}^\dagger$ is Hermitian. In particular, for $\boldsymbol{v} \in \mathbb{C}^n$, the rank-one map $\boldsymbol{v}\boldsymbol{v}^\dagger$ is Hermitian.

- The rotation matrices $\boldsymbol{R}_\phi$ fulfill

$$\boldsymbol{R}_\phi^\dagger = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}^\dagger = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \boldsymbol{R}_{-\phi} = \boldsymbol{R}_\phi^{-1}$$

  and are therefore unitary.

Here's what makes normal matrices important:

A matrix is normal if and only if it has an ortho-normal eigenbasis. (3.23)

Remark. The condition "$\boldsymbol{A}$ is normal" is almost trivial to verify: Take $\boldsymbol{A}$, transpose-conjugate it to get $\boldsymbol{A}^\dagger$, compute two matrix products $\boldsymbol{A}\boldsymbol{A}^\dagger$, $\boldsymbol{A}^\dagger\boldsymbol{A}$, subtract, check whether you got 0. Easy.

On the other hand, "$\boldsymbol{A}$ has an orthonormal eigenbasis" is a priori *much* more difficult to check for any given $\boldsymbol{A}$! You *could* find all the roots of the characteristic polynomial, solve the linear equation for the eigenspaces, check whether you'll find a basis at all, and if so, whether Gram-Schmidt on degenerate spaces gives you an ONB. Near-impossible for large matrices!

Mapping something a priori difficult to something easy is what makes the statement useful. This suggests that the implication "$\Rightarrow$" is the more difficult one to prove, which is indeed correct.

Proof. Claim: If $\boldsymbol{A}$ is normal, then

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \quad \Rightarrow \quad \boldsymbol{A}^\dagger\boldsymbol{v} = \bar{\lambda}\boldsymbol{v}. \tag{3.24}$$

Proof: We'll show that $\|\boldsymbol{A}\boldsymbol{v} - \lambda\boldsymbol{v}\|^2 = 0$ implies $\|\boldsymbol{A}^\dagger\boldsymbol{v} - \bar{\lambda}\boldsymbol{v}\|^2 = 0$. Indeed:

$$\begin{aligned}
\|\boldsymbol{A}^\dagger\boldsymbol{v} - \bar{\lambda}\boldsymbol{v}\|^2 &= \langle \boldsymbol{A}^\dagger\boldsymbol{v} - \bar{\lambda}\boldsymbol{v}, \boldsymbol{A}^\dagger\boldsymbol{v} - \bar{\lambda}\boldsymbol{v}\rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{v}\rangle - \bar{\lambda}\langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}\rangle - \lambda\langle \boldsymbol{v}, \boldsymbol{A}^\dagger\boldsymbol{v}\rangle + \lambda\bar{\lambda}\langle \boldsymbol{v}, \boldsymbol{v}\rangle \\
&= \langle \boldsymbol{v}, \boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{v}\rangle - \bar{\lambda}\langle \boldsymbol{v}, \boldsymbol{A}\boldsymbol{v}\rangle - \lambda\langle \boldsymbol{v}, \boldsymbol{A}^\dagger\boldsymbol{v}\rangle + \lambda\bar{\lambda}\langle \boldsymbol{v}, \boldsymbol{v}\rangle \quad \text{(by normality)} \\
&= \|\boldsymbol{A}\boldsymbol{v} - \lambda\boldsymbol{v}\|^2 = 0 \qquad\qquad\qquad\qquad\qquad\quad \text{(by assumption).}
\end{aligned}$$

We now show by induction that a normal $n \times n$-matrix has k orthonormal eigenvectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ for all $k \leq n$.

Anchor: The statement is true for $k = 1$, because by the fundamental theorem of algebra, every matrix has at least one eigenvector, which we can scale to be normalized.

Induction step: Now assume it is true for some $k < n$. Let U be the subspace of $\mathbb{C}^n$ consisting of the vectors that are orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$. If $\boldsymbol{w} \in U$, then using (3.24),

$$\langle \boldsymbol{v}_i, \boldsymbol{A}\boldsymbol{w}\rangle = \langle \boldsymbol{A}^\dagger\boldsymbol{v}_i, \boldsymbol{w}\rangle = \lambda_i\langle \boldsymbol{v}_i, \boldsymbol{w}\rangle = 0, \qquad i = 1, \ldots k.$$

Hence $\boldsymbol{A}\boldsymbol{w} \in U$, too. Therefore, the restriction of $\boldsymbol{A}$ to U defines a linear map $U \to U$. Invoking again the fundamental theorem of algebra, that map must have an eigenvector $\boldsymbol{v}_{k+1}$, which we can scale to be normalized. By construction, $\boldsymbol{v}_{k+1}$ is also an eigenvector of $\boldsymbol{A}$, and orthogonal to the $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$.

We have shown $\boldsymbol{A}$ normal $\Rightarrow$ it has an ortho-normal eigenbasis. Now for the converse. Assume $\boldsymbol{A}$ has an orthonormal eigenbasis $\{\boldsymbol{v}_i\}$ with associated eigenvalues λ_i. We claim

$$\boldsymbol{A} = \sum_i \lambda_i\, \boldsymbol{v}_i\boldsymbol{v}_i^\dagger.$$

Indeed,

$$\big(\sum_i \lambda_i\, \boldsymbol{v}_i\boldsymbol{v}_i^\dagger\big)\boldsymbol{v}_j = \sum_i \lambda_i\, \boldsymbol{v}_i \underbrace{\boldsymbol{v}_i^\dagger\boldsymbol{v}_j}_{\delta_{ij}} = \lambda_j\boldsymbol{v}_j$$

so $\boldsymbol{A}$ and $\sum_i \lambda_i\boldsymbol{v}_i\boldsymbol{v}_i^\dagger$ act in the same way on the basis $\{\boldsymbol{v}_j\}$. They therefore define the same linear map. But then

$$\boldsymbol{A}^\dagger = \big(\sum_i \lambda_i\, \boldsymbol{v}_i\boldsymbol{v}_i^\dagger\big)^\dagger = \sum_i \bar{\lambda}_i\, \boldsymbol{v}_i\boldsymbol{v}_i^\dagger, \tag{3.25}$$

which has the same eigenbasis as $\boldsymbol{A}$, and thus commutes with it. □

In the proof, we have also shown

$$\begin{aligned} & \boldsymbol{A} \text{ is normal} \\ \Leftrightarrow \quad & \boldsymbol{A} \text{ has an ortho-normal eigenbasis } \{\boldsymbol{v}_i\}_i \text{ with eigenvalues } \lambda_i \\ \Leftrightarrow \quad & \boldsymbol{A} = \sum_i \lambda_i \, \boldsymbol{v}_i \boldsymbol{v}_i^\dagger \quad \text{with } \{\boldsymbol{v}_i\}_i \text{ an ONB.} \end{aligned} \tag{3.26}$$

Equation (3.26) is called the *spectral representation* or *eigendecomposition* of $\boldsymbol{A}$.

Remark.

- "Spectrum" is a generalization of "the set of eigenvalues". For matrices, the two terms have the same meaning. They are different in infinite dimensions, though. Don't worry about this for now.
- Eq. (3.26) is *very* useful! It explicitly expresses the operator in terms of its eigenbasis and eigenvalues. In QM, the very first step in calculations is typically to write down the eigendecomposition (3.26) for any operator and take it from there.
- Syntax check! (After long discussion in the lecture on this point). The l.h.s. of (3.26) is a matrix. The r.h.s. involves *vectors*. How does that even make formal sense? Well, $\boldsymbol{v}_i \boldsymbol{v}_i^\dagger$ is the product of an $n \times 1$ matrix (i.e. column vector) and a $1 \times n$ matrix (i.e. row vector). Altogether, this gives an $n \times n$-matrix, so (3.26) is at least syntactically correct.

3.7.2 Unitary and orthogonal matrices

We have defined that a matrix $\boldsymbol{U}$ is unitary if

$$\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1}. \tag{3.27}$$

If the matrix has in addition real entries, then the adjoint can be replaced by a transpose. A real matrix $\boldsymbol{O}$ is *orthogonal* if

$$\boldsymbol{O}^t = \boldsymbol{O}^{-1}. \tag{3.28}$$

Here's the interpretation: Unitary matrices are to $\mathbb{C}^n$, and orthogonal matrices to $\mathbb{R}^n$, what rotations are to $\mathbb{R}^3$.

There are many ways of characterizing "rotations" and likewise, many ways of characterizing "unitarity" and "orthogonality". Let's look at some.

- *Rotations preserve lengths and angles*. Generalizing this property:

  A complex matrix $\boldsymbol{U}$ is unitary if and only if it preserves Hermitian inner products

  $$\langle \boldsymbol{U}\boldsymbol{v}, \boldsymbol{U}\boldsymbol{w} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle \qquad \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^n. \tag{3.29}$$

  A real matrix $\boldsymbol{O}$ is orthogonal if and only if it preserves Euclidean inner products

  $$\langle \boldsymbol{O}\boldsymbol{v}, \boldsymbol{O}\boldsymbol{w} \rangle = \langle \boldsymbol{v}, \boldsymbol{w} \rangle \qquad \forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n.$$

Proof. □

- *Rotations map the standard basis to an ONB*. Generalizing this property:

  A complex matrix $\boldsymbol{U}$ is unitary if and only if its columns form a Hermitian ONB.
  A real matrix $\boldsymbol{O}$ is orthogonal if and only if its columns form a Euclidean ONB.

Proof. □

There's one more, very important!, characterization of unitaries. Think of it this way: A diagonal matrix $\boldsymbol{U} = \mathrm{diag}((e^{i\phi_j})_j)$ obviously preserves the Hermitian inner product and is therefore unitary. By switching to the spectral decomposition of $\boldsymbol{U}$, it is easy to see that *every* unitary is of this form, if we allow for the phases being applied w.r.t. to a general ONB, rather than just the standard basis.

- A complex matrix $\boldsymbol{U}$ is unitary if and only if it is normal and has eigenvalues $\lambda_i = e^{i\phi_i}$ of absolute value equal to 1

$$\boldsymbol{U} = \sum_i e^{i\phi_i} \boldsymbol{v}_i \boldsymbol{v}_i^\dagger, \qquad \phi_i \in \mathbb{R}, \{\boldsymbol{v}_i\} \text{ an ONB of } \mathbb{C}^n. \tag{3.30}$$

Proof. □

Below is work very much in progress.

We can write (3.30) as

$$\boldsymbol{U} = e^{i\boldsymbol{H}}, \qquad \boldsymbol{H} := \sum_i \phi_i \boldsymbol{v}_i \boldsymbol{v}_i^\dagger. \tag{3.31}$$

One says that $\boldsymbol{U}$ is *generated* by $\boldsymbol{H}$.

Examples. Homework.

Remarks. If $\boldsymbol{U}$ and $\boldsymbol{V}$ both preserve inner products, then so do the composition $\boldsymbol{UV}$ and their inverses. The set of unitary matrices on $\mathbb{C}^n$ is therefore a group, the *unitary group* $U(n)$.

The determinant of a unitary has absolute value 1. This follows from the fact that the same is true for its eigenvalues, or more directly from

$$|\det \boldsymbol{U}| = \det \boldsymbol{U} \det \boldsymbol{U}^\dagger = \det \boldsymbol{U}\boldsymbol{U}^{-1} = \det \mathbb{1} = 1.$$

From the multipicativity property $\det(\boldsymbol{UV}) = \det \boldsymbol{U} \ \det \boldsymbol{V}$ of the determinant, it follows that the set of unitaries with determinant 1 is a subgroup. It is called the *special unitary group* $SU(n)$.

Analogously, the set of orthogonal matrices from the *orthogonal group* $O(n)$. Repeating the above calculation, $|\det \boldsymbol{O}| = 1$. But because $\boldsymbol{O}$ is real, we actually have $\det = \pm 1$. Because $\det$ is a continuous function, we thus see that it is not possible to continuously deform an orthogonal matrix with positive determinant into one with negative determinant. Therefore, $O(n)$ has (at least, and in fact, exactly) two disconnected components. The connected component that contains the identity is the subgroup $SO(n)$ of *proper rotations*. The connected component $\{\boldsymbol{O} \in O(n) \mid \det \boldsymbol{O} = -1\}$ is not subgroup (why?). It contains e.g. the reflections about hyperplanes.

Let $\boldsymbol{U} \in U(n)$. From (3.31), there is a Hermitian matrix $\boldsymbol{H}$ such that $\boldsymbol{U} = e^{i\boldsymbol{H}}$. Now introduce a parameter $t \in \mathbb{R}$ and consider the family

$$\boldsymbol{U}(t) = e^{-it\boldsymbol{H}}$$

(the minus sign is convention). We have found a continuous family of unitaries such that $\boldsymbol{U}(0) = \mathbb{1}$ and $\boldsymbol{U}(1) = \boldsymbol{U}$. Thus, unlike $O(n)$, the unitary group is connected. What is more,

$$\boldsymbol{U}(t)\boldsymbol{U}(s) = \boldsymbol{U}(t+s), \qquad i\partial_t \boldsymbol{U}(t) = \boldsymbol{H}\boldsymbol{U}(t).$$

The first equation says that the $\boldsymbol{U}(t)$'s form a *(one parameter) group*.

Example. In QM, the state of a system is represented by a complex vector $\boldsymbol{\psi}$ of norm $\|\boldsymbol{\psi}\|^2 = 1$. Time evolution $\boldsymbol{\psi}(t)$ has to preserve normalization. It is therefore not surprising that it is implemented by a unitary $\boldsymbol{\psi}(t) = \boldsymbol{U}(t)\boldsymbol{\psi}(0)$. From the above, it is natural to look at the Hermitian $\boldsymbol{H}(t)$ that generated the unitary. For closed systems, it turns out to be time-independent $\boldsymbol{H}(t) = \boldsymbol{H}$. The quantum-mechanical time evolution is therefore determined by

$$i\partial_t \boldsymbol{\psi}(t) = i\partial_t \boldsymbol{U}(t)\boldsymbol{\psi}(0) = \boldsymbol{H}\boldsymbol{U}(t)\boldsymbol{\psi}(0) = \boldsymbol{H}\boldsymbol{\psi}(t),$$

which is the general form of the *Schrödinger equation*.

3.7.3 Hermitian and symmetric matrices

We have defined that a matrix $\boldsymbol{A}$ is Hermitian if

$$\boldsymbol{A}^\dagger = \boldsymbol{A}.$$

If the matrix has in addition real entries, then the adjoint can be replaced by a transpose. A real matrix $\boldsymbol{S}$ is *symmetric* if

$$\boldsymbol{S}^t = \boldsymbol{S}.$$

From (3.25), it follows that Hermitian (and hence symmetric) matrices have real eigenvalues.

Projections

There is a one-one relation between:

- Subspaces $V \subset \mathbb{C}^d$, and
- *projections* P, i.e. matrices maps fulfilling $P = P^\dagger$, $P^2 = P$.

(see Fig. 3.1)

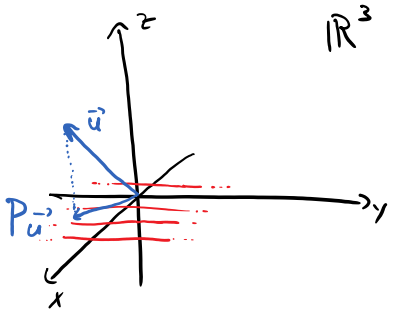


Figure 3.1: Orthogonal projection of $\boldsymbol{u}$ onto the x-y-plane.

→ **Homework**

Chapter 4

Tensors and multilinear algebra

- Good news: Tensors are conceptually really simple.
- Bad news: *Many* simple things happen at the same time, so one can easily get very confused.
- Good news having the last word: It only takes a bit of practice to grok it.

4.1 Tensors as arrays of numbers

Let's recap:

- A *scalar* is a single number λ from some field $\mathbb{F}$.
- A concrete *vector* $\boldsymbol{v}$ is a linear array (v_i) of numbers. Specifically, if $\boldsymbol{v} \in \mathbb{F}^n$, the index i runs from 1 to n and the v_i are elements of the field $\mathbb{F}$.
- A concrete *matrix* $\boldsymbol{M}$ is a rectangular array of numbers M_{ij}. Specifically, if $\boldsymbol{M} \in \mathbb{F}^{n\times m}$, the index i runs from 1 to n, and the index j from 1 to m, and the M_{ij} are elements of the field $\mathbb{F}$.

An obvious generalization suggests itself:

- A *concrete tensor* $\boldsymbol{T}$ is an array of numbers $T_{i_1 i_2 \ldots i_s}$. Specifically,
  - $\mathbb{F}^{n_1\times\cdots\times n_s}$ is the set of arrays $\boldsymbol{T}$ with s indices, where
  - i_1 runs from 1 to n_1,
    ..., and, ...
    i_s runs from 1 to n_s,
  - each number $T_{i_1\ldots i_s}$ lies in the field $\mathbb{F}$.

We call s the *order* of the tensor and the numbers $(n_1, \ldots, n_s)$ its *shape*.

Remark. Unfortunately, the terminology around tensors is not standardized, and often very confusing.

Instead of "order", some authors use *rank* or *dimension*. That's not ideal, because these terms already have different meanings in linear algebra. (In German, we'll use *Stufe* for order, which seems more common than the literal translation "*Ordnung*"). The j-th index i_j is sometime said to correspond to the jth *axis* of the tensor. The Julia language, to add to the confusion, calls it the jth *dimension*.

Tensors (or "multidimensional arrays") in Julia:

```
julia> T = [
       1 2
       3 4;;;
       5 6
       7 8
       ]
2×2×2 Array{Int64, 3}:
[:, :, 1] =
 1  2
 3  4

[:, :, 2] =
 5  6
 7  8

julia> T[1,1,1]
1

julia> T[1,1,2]
5

julia> size(T) # the shape of T
(2, 2, 2)
```

A tensor of shape $(n_1, \dots, n_s)$ is specified by $\prod_{i=1}^{s} n_i$ coefficients. If $n_i = n$ is constant, this number grows exponentially with the order $d = n^s$.

Remark. The exponential growth means that tensors of high order are difficult to handle on a computer. That's felt particularly acutely in quantum many-body theory, where states are specified by tensors, with the order corresponding to the number of degrees of freedom. Take for example the simplest quantum system: a spin-$1/2$ particle. The state of a single one is described by a vector in $\mathbb{C}^2$, i.e. by just two complex numbers. But the state of s spins is an element of $\mathbb{C}^{2\times\cdots\times 2}$, requiring 2^s elements of $\mathbb{C}$. If we arrange spin-$1/2$ particles as a cube with side length 10 – a *tiny* system! – then $s = 1000$ and $2^s \simeq 10^{300}$. This is such an absurdly large number that even galactic references, like the number of hadrons in the universe, pale compared to it.

That's bad news if you want to do computer physics: There's patently no chance to even consider putting all these numbers into computer memory. In some physically relevant situations, it turns out that it suffices to work with low-dimensional subsets of all tensors (e.g. "tensor network states") to simulate the system – but, yes, in general, it means we just cannot simulate even such simple situations. ...or, one can turn lemons into lemonade: Maybe, one could hope, it is possible to channel the computational power that Nature demonstrates when time-evolving a few-spin system in order to perform useful computations that are beyond the reach of classical machines. For all we know, this is indeed possible, and leads to the field of *quantum computation*.

4.2 Operations on tensors

4.2.1 Addition, scalar multiplication

Scalar multiplication and addition between tensors of the same shape are defined componentwise:

$$(\lambda \boldsymbol{T})_{i_1,\dots,i_s} = \lambda T_{i_1,\dots,i_s} \qquad (\boldsymbol{T} + \boldsymbol{S})_{i_1,\dots,i_s} = T_{i_1,\dots,i_s} + S_{i_1,\dots,i_s}$$

In particular, the set $\mathbb{F}^{n_1\times\cdots\times n_s}$ of tensors of given shape $(n_1, \dots, n_s)$ and with elements in some field $\mathbb{F}$ form a vector space.

It's a good idea to think of a tensor $\boldsymbol{T}$ as a function which maps the s arguments $i_1, \dots, i_s$ to $\mathbb{F}$. Then the operations above are just the usual addition and scalar multiplication of functions.

Tensor T as in previous example.

```
julia> S
2×2×2 Array{Int64, 3}:
[:, :, 1] =
 1  1
 1  1

[:, :, 2] =
 0  0
 0  0

julia> T + 5*S
2×2×2 Array{Int64, 3}:
[:, :, 1] =
 6  7
 8  9

[:, :, 2] =
 5  6
 7  8
```

4.2.2 Tensor product

More exciting is the *tensor product* which maps a tensor $\boldsymbol{T}$ of order t and a tensor $\boldsymbol{S}$ of order s to their product $\boldsymbol{T} \otimes \boldsymbol{S}$ of order $t + s$

$$(\boldsymbol{T} \otimes \boldsymbol{S})_{i_1,\dots,i_t,i_{t+1},\dots,i_{t+s}} = T_{i_1,\dots,i_t}\, S_{i_{t+1},\dots,i_{t+s}}.$$

In the tensor-as-a-function-language:

The product of tensors is their product as functions, with independent arguments.

Example.

- If $t = s = 1$, the tensor product is just the outer product between vectors:

$$\boldsymbol{M}_{ij} = \boldsymbol{T}_i \boldsymbol{S}_j.$$

- Consider N dice, where the i-th dice shows j eyes with probability $p_j^{(i)}$. If the N dice are cast one after the other, the *joint probability* $P_{i_1,\dots,i_N}$ of the first dice showing j_1 eyes, the second dice j_2 eyes, and so on, is

$$P_{i_1,\dots,i_N} = p_{i_1}^{(1)} p_{i_2}^{(2)} \dots p_{i_N}^{(N)}.$$

  In other words: The joint distribution of independent events is the tensor product of the individual distributions

$$P = p^{(1)} \otimes \dots \otimes p^{(N)}.$$

A tensor $\boldsymbol{T}$ of order s is called a *product tensor* (or *simple tensor* or *elementary tensor*) if there are vectors $\boldsymbol{v}_1 \dots \boldsymbol{v}_s$ such that $\boldsymbol{T} = \boldsymbol{v}_1 \otimes \dots \otimes \boldsymbol{v}_s$. (I.e. if the function $\boldsymbol{T}$ is product function that only depend on one parameter each).

The number of parameters required to specify a product tensor is roughly $\sum_{i=1}^{s} n_i$. This is much smaller than the number of parameter $\prod_{i=1}^{s} n_i$ that specify a general tensor. In this sense, "being a product tensor" is an exceptional property.

Remark. Actshually... ...the number of parameters that specify a product tensor is $\sum_{i=1}^{s} n_i - \mathbf{1}$. Why?

Remark. An order-2 tensor $\boldsymbol{T}_{ij}$ is a product if and only if $T_{ij} = v_i w_j$ for some vectors $\boldsymbol{v}, \boldsymbol{w}$, i.e. if and only if all columns are proportional to each other, i.e. if and only if the matrix $\boldsymbol{T}$ has rank one. This suggests to generalize the notion of "rank" to tensors. For example, one can define the *tensor rank* of $\boldsymbol{T}$ to be the smallest number r such that $\boldsymbol{T}$ can be written as the of r product tensors. Sounds good, no? Trouble is that there is no analogue of the Gaussian algorithm for most tensorial problems. And indeed, there are natural order-3 tensors for which serious researchers have been trying hard for decades to figure out their rank. Ask me for examples, if you want to waste time.

Product bases

Recall the standard basis vectors $\boldsymbol{e}_i$ with elements $(\boldsymbol{e}_i)_j = \delta_{ij}$ of $\mathbb{F}^n$.

The tensor product $\boldsymbol{e}_{i_1} \otimes \boldsymbol{e}_{i_2}$ has components

$$(\boldsymbol{e}_{i_1} \otimes \boldsymbol{e}_{i_2})_{j_1 j_2} = (\boldsymbol{e}_{i_1})_{j_1} (\boldsymbol{e}_{i_2})_{j_2} = \delta_{i_1,j_1} \delta_{i_2,j_2}.$$

In other words: It is a matrix with all entries zero, except a one at the intersection of row i_1 and column i_2. A matrix $\boldsymbol{M} = (M_{i_1,i_2})$ can therefore be expanded as

$$\boldsymbol{M} = \sum_{i_1,i_2} M_{i_1,i_2} (\boldsymbol{e}_{i_1} \otimes \boldsymbol{e}_{i_2})$$

and so the set $\{\boldsymbol{e}_{i_1} \otimes \boldsymbol{e}_{i_2}\}_{i_1,i_2}$ forms a basis for the space of matrices.

In exactly the same way one sees that for order-s tensors

$$\boldsymbol{T} = \sum_{i_1,\ldots,i_s} T_{i_1,\ldots,i_s} (\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_s}).$$

There are many notations in use for the elements of such *product bases*:

$$\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_s} = \boldsymbol{e}_{i_1} \ldots \boldsymbol{e}_{i_s} = \boldsymbol{e}_{i_1,\ldots,i_s} = |i_1, \ldots i_n\rangle.$$

4.3 Geometric meaning

Linear algebra started with column vectors... ...which we later interpreted as the basis representation of abstract vectors. We then looked at matrices... ...which we later interpreted as basis representations of linear maps.

So the question arises: What, if anything, do concrete tensors represent?

4.3.1 Multilinear forms

Recall: If V is a vector space over $\mathbb{F}$, then a function $K : V \times \cdots \times V \to \mathbb{F}$ is a *multilinear form* if it is linear in each argument. Let $\{e_i\}$ be a basis for V. Given vectors $v_1, \ldots, v_s \in V$, these can be expanded in terms of the basis as

$$\boldsymbol{v}_i = \sum_j c_j^{(i)} e_i$$

and thus, using multilinearity,

$$\begin{aligned} K(v_1,\dots,v_s) &= K\Big(\sum_{j_1} c^{(1)}_{j_1} e_{j_1},\dots,\sum_{j_s} c^{(s)}_{j_s} e_{j_s}\Big) \\ &= \sum_{j_1,\dots,j_s} c^{(1)}_{j_1}\dots c^{(s)}_{j_s} \underbrace{K\big(e_{j_1},\dots e_{j_s}\big)}_{=:R_{j_1,\dots,j_n}}. \end{aligned} \tag{4.1}$$

Therefore:

An s-linear form is described (with respect to a basis) by an order-s tensor.

We can introduce a new isomorphism from the space of s-linear forms over an n-dimensional vector space V to the space of concrete tensors $\mathbb{F}^{n\times\dots\times n}$:

$$\boldsymbol{\phi}_{\mathcal{B},\dots,\mathcal{B}}(K) = \boldsymbol{R}, \qquad R_{j_1,\dots,j_s} = K\big(b_{j_1},\dots b_{j_s}\big).$$

Examples.

- Recap: The concrete order-n tensor $\epsilon \in \mathbb{F}^{n\times\dots\times n}$ with components $\epsilon_{j_1\dots j_n} := \det\big(\boldsymbol{e}_{j_1},\dots \boldsymbol{e}_{j_n}\big)$ is called the *Levi-Civita symbol*. It is therefore the concrete tensor representing the determinant in the standard basis.
- The standard Euclidean inner product is represented by the tensor $g_{ij} = \delta_{ij}$ with respect to the standard basis

  $$\langle \boldsymbol{e}_i, \boldsymbol{e}_j\rangle = \delta_{ij}.$$

- The *Minkowski inner product* which describes the geometry of space-time in relativity is

  $$\eta_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{e}_j\rangle = \begin{cases} 1 & i=j=1 \\ -1 & i=j\in\{2,3,4\} \end{cases}.$$

  In general, an order-2 tensor associated with a bilinear form is called a *metric tensor*.

Remark. From a mathematical perspective, a linear map $L : V \to V$ and a bilinear form $K : V \times V \to \mathbb{F}$ are completely different objects. Yet, in coordinates, they are both represented by order-2 tensors. So if a two-indexed object pops up in a physics calculation, it may not be immediately obvious what its geometric meaning is.

Even if one is disinterested in such questions for their own sake, it is important to distinguish representations of maps from representations of forms. E.g. because the transformation rules under coordinate changes differ. (c.f. Sec. **??**).

Example. The 2×2-determinant is represented w.r.t. the standard basis $\mathcal{B} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ by

$$\big(\boldsymbol{\phi}_{\mathcal{B},\mathcal{B}}(\det)\big)_{ij} = \det(\boldsymbol{e}_i, \boldsymbol{e}_j) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}_{ij}.$$

A rotation $R : \boldsymbol{e}_1 \to \boldsymbol{e}_2, \boldsymbol{e}_2 \to -\boldsymbol{e}_1$ by 90 degrees is described by the same matrix:

$$\big(\phi^{\mathcal{B}}_{\mathcal{B}}(R)\big)_{ij} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}_{ij}.$$

But now change the basis to $\mathcal{B}'$ with elements

$$\boldsymbol{f}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{f}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then

$$\big(\phi_{\mathcal{B}',\mathcal{B}'}(\det)\big)_{ij} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}_{ij}, \quad \text{but} \quad \big(\phi^{\mathcal{B}'}_{\mathcal{B}'}(R)\big)_{ij} = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

Tensor product of forms

Recall: For a vector space V, the set of linear forms $V \to \mathbb{F}$ is again a vector space, the dual space V^*.

Take two linear forms $\alpha, \beta \in V^*$. We define:

$$(\alpha \otimes \beta)(v, w) = \alpha(v)\beta(w).$$

Thus:

A tensor product of functionals is their product, with independent arguments.

Then $\alpha \otimes \beta$ is bilinear,

$$(\alpha \otimes \beta)(v + \lambda v', w) = \alpha(v + \lambda v')\beta(w) = \alpha(v)\beta(w) + \lambda\alpha(v')\beta(w)$$

and represented in coordinates by the tensor product of the representations of α and β:

$$(\alpha \otimes \beta)(v, w) = (\alpha \otimes \beta)(\sum_i v_i e_i, \sum_j w_j e_j) = \sum_{ij} v_i w_j \, \alpha(e_i)\beta(e_j)$$

and hence

$$\underbrace{\phi_{\mathcal{B},\mathcal{B}}(\alpha \otimes \beta)}_{\text{basis rep. of tensor product of forms}} = \underbrace{\phi_{\mathcal{B}}(\alpha) \otimes \phi_{\mathcal{B}}(\beta)}_{\text{tensor product of basis rep. of forms}}.$$

More generally, if given a multilinear form K on V with k arguments, and another one, L with l arguments. Then one defines a multilinear form with $k + l$ arguments by

$$(K \otimes L)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+1}) = K(v_1, \ldots, v_k)\, L(v_{k+1}, \ldots, v_{k+l}).$$

Tensor product of dual spaces

Define the tensor product of dual spaces by

$$V^* \otimes V^* := \operatorname{span}\{\alpha \otimes \beta \,|\, \alpha \in V^*, \beta \in V^*\}.$$

By the above, the elements are bilinear forms. In fact, all bilinear forms arise this way!

Indeed, let K be a bilinear form. Choose a basis $\{e_i\}$ on V and let $\{\epsilon_i\} \subset V^*$ be the dual basis, defined by $\epsilon_i(e_j) = \delta_{ij}$. Then

$$\Big(\sum_{ij} K(e_i, e_j)\,(\epsilon_i \otimes \epsilon_j)\Big)(e_k, e_l) = \sum_{ij} K(e_i, e_j)\,\epsilon_i(e_k)\epsilon_j(e_l) = K(e_k, e_l).$$

Thus, because

$$\sum_{ij} K(e_i, e_j)\,(\epsilon_i \otimes \epsilon_j) \qquad \text{and} \qquad K \tag{4.2}$$

take the same values on a basis, they must be the same, and hence $K \in V^* \otimes V^*$.

More generally: For vector spaces $V_1, \ldots, V_N$,

$$V_1^* \otimes \cdots \otimes V_N^* := \operatorname{span}\{\alpha_1 \otimes \cdots \otimes \alpha_N \,|\, \alpha_i \in V_i^*\}.$$

The tensor product of dual spaces is the linear span of the products of linear forms.

4.3.2 General geometric tensor products

We have worked out a coordinate-free description of tensor products of multilinear forms. That was comparatively easy, because the product $\alpha \otimes \beta$ has a natural interpretation (as the product of functions with independent arguments).

Elements of a vector space V aren't generally functions, so this simple definition doesn't work... ...or does it?

We now employ a trick that is conceptually simple, standard, and infamously confusing on first encounter. The trick is to realize that elements of a space V *are also linear functionals* – on V^*. That's right! For $v \in V$ and $\alpha \in V^*$, define $v(\alpha) = \alpha(v)$.

Really??
Q.: So, in terms of highschool math, for a point x and a function f, we have defined $x(f)$ to mean $f(x)$?
A.: You got it!

The resulting function is indeed linear:

$$v(\alpha + \lambda\beta) = (\alpha + \lambda\beta)(v) = \alpha(v) + \lambda\beta(v) = v(\alpha) + \lambda v(\beta).$$

In other words, we have shown that $V \subset (V^*)^*$.

Remark. In fact, every linear functional on V^* is realized by some element of V, i.e. $V = (V^*)^*$. Can you show it? (Hint: Use dual bases).

Now take two vectors $v, w \in V$. We can define their tensor product by interpreting them as linear functionals. That is, the tensor product $v \otimes w$ is a bilinear function on V^* acting as

$$(v \otimes w)(\alpha, \beta) = \alpha(v)\beta(w).$$

In general, we can mix and match primal and dual spaces. If $V_1 \ldots, V_N, W_1, \ldots, W_M$ are vector spaces, $v_i \in V_i$ vectors, and $\alpha_i \in W_i^*$ linear functionals, then

$$v_1 \otimes \cdots \otimes v_N \otimes \alpha_1 \otimes \cdots \otimes \alpha_M \tag{4.3}$$

is the multilinear functional that maps arguments $\beta_i \in V_i^*, w_i \in W_i$ to

$$\beta_1(v_1) \ldots \beta_N(v_N)\alpha_1(w_1) \ldots \alpha_M(w_M).$$

Then tensor product of spaces

$$V_1 \otimes \cdots \otimes V_N \otimes W_1^* \otimes \cdots \otimes W_M^*$$

is the linear hull of product tensors of the form (4.3).

In applications, we are mostly concerned with the case where all spaces are the same: $V_i = W_i = V$. In this case, it is mildly annoying, but straight-forward to generalize the argument around (4.2) to find that a tensor

$$K \in \underbrace{V \otimes \cdots \otimes V}_{N\times} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{M\times} =: V^{\otimes N} \otimes (V^*)^{\otimes M}$$

can be expanded with respect to some basis $\{e_i\}$ of V and dual basis $\{\epsilon_j\}$ of V^* as

$$K = \sum_{i_1,\ldots,i_N;j_1,\ldots,j_M} K_{i_1,\ldots,i_N;j_1,\ldots,j_M}\, e_{i_1} \otimes \cdots \otimes e_{i_N} \otimes \epsilon_{j_1} \otimes \cdots \otimes \epsilon_{j_n}. \tag{4.4}$$

4.3.3 Contractions

Student remark from lecture. "Why do we *want* to produce functions with more arguments? Doesn't that make everything worse?"

Consider the tensor product $\alpha \otimes v \in V^* \otimes V$. There's a natural way to turn this tensor into a number. Because α is a linear form, which likes to eat vectors, and v is just such a thing, can define

$$C : \alpha \otimes v \mapsto \alpha(v) \in \mathbb{F}.$$

This operation is called a *contraction* (*Verjüngung*). Contractions reduce the order of a tensor by two.

With respect to a basis, we have

$$C : \sum_{ij} T_{ij}\, \epsilon_i \otimes e_j \mapsto \sum_{ij} T_{ij}\, \epsilon_j(e_i) = \sum_i c_{ii}.$$

Thus, a contraction corresponds in coordinates to the "diagonal sum" over the two indices corresponding to the dual and the primal basis. We'll give many more examples in the next section.

4.4 Covariant notation

Physicists often employ a set of conventions known as *covariant notation* or *Ricci calculus*. It's like a cheat code for tensors, which allows one to perform meaningful calculations without having to parse abstract non-sense like $V \subset (V^*)^*$...

Start with a vector space V and choose a basis $\{e_i\}_i$ of V, with associated dual basis $\{\epsilon_i\}_i$ of V^*. We want to work with tensors in $V^{\otimes N} \otimes (V^*)^{\otimes M}$. As in Eq. (4.4), a general element of this space is of the form

$$K = \sum_{i_1,\ldots,i_N;j_1,\ldots,j_M} K_{i_1,\ldots,i_N;j_1,\ldots,j_M}\, e_{i_1} \otimes \cdots \otimes e_{i_N} \otimes \epsilon_{j_1} \otimes \cdots \otimes \epsilon_{j_n}.$$

From now on, we will

- write indices of expansion coefficients for vectors as superscripts and call them *contravariant*;
- write indices of expansion coefficients for linear forms as subscripts and call them *covariant*;

- use the opposite index placement for primal and dual basis elements; and
- sum over any index that appears both as a superscript and a subscript. (This is the *Einstein summation convention*).

The above equation becomes

$$K = K^{i_1,\ldots,i_N}{}_{j_1,\ldots,j_M} \left(e_{i_1} \otimes \cdots \otimes e_{i_N} \otimes \epsilon^{j_1} \otimes \cdots \otimes \epsilon^{j_n}\right).$$

Next, we will use $K^{i_1,\ldots,i_N}{}_{j_1,\ldots,j_M}$ to refer both

- to the individual elements of the tensor (i.e. treating the indices as concrete numbers), and
- to the entire tensor K (i.e. treating the indices as unspecified variables).

The equation we started with is now reduced to

$$K = K^{i_1,\ldots,i_N}{}_{j_1,\ldots,j_M}.$$

Users of the Ricci calculus will generally avoid talking about V, V^*, or any basis choice explicitly. Instead, they will say "$K^{i_1,\ldots,i_N}{}_{j_1,\ldots,j_M}$ is an *N-fold contravariant and M-fold covariant tensor*" and only ever work with such expansion coefficients.

Let's go through the various operations of linear algebra, and see how they look through this lens.

concept	basis-free notation	Ricci notation	Ricci name
vector	$v \in V$	v^i	contravariant vector
linear form	$\alpha \in V^*$	α_i	covariant vector
evaluation of form	$\alpha(v)$	$\alpha_i u^i$	contraction
bilinear form	$B \in V^* \otimes V^*$	B_{ij}	twice contravariant tensor metric tensor
evaluation of form	$B(v, w)$	$B_{ij} v^i w^j$	contraction

Now let's see how linear maps fit into this story. The tensor

$$\sum_{ij} L^i{}_j \, (e_i \otimes \epsilon_j) \in V \otimes V^*$$

defines a linear map $V \to V$ by letting the form act on an argument:

$$w = \sum_k w_k e_k \mapsto \sum_{ij} L^i{}_j \, e_i \epsilon_j(w) = \sum_i \left(\sum_j L^i{}_j w_j \right) \, e_i.$$

But any linear map $V \to V$ is of this form, for suitable coefficients $L^i{}_j$. Thus, we have seen that the space of linear maps $V \to V$ is the same as the "mixed tensor product" $V \otimes V^*$. Extending the table:

concept	basis-free notation	Ricci notation	Ricci name
linear map $V \to V$	$L \in V \otimes V^*$	$L^i{}_j$	mixed tensor
evaluation of linear map	$L(v)$	$L^i{}_j v^j$	contraction
trace	$\operatorname{tr} L$	$L^i{}_i$	contraction

Finally, let's talk about the operations of "raising and lowering the index", a favorite past-time e.g. of researchers in general relativity.

Let $B \in V^* \otimes V^*$ be a bilinear map, $v \in V$ a vector. Then one can define a linear form

$$w \mapsto B(v, w).$$

If B is a Euclidean scalar product, then this form is the projection onto v. In the Ricci calculus, the above reads

$$w^i \mapsto B_{ij} v^j w^i.$$

One then introduces the short-hand $v_i := B_{ij} v^j$, so that $B_{ij} v^j w^i$ can be written as $v_i w^j$. In reference to its appearance in the Ricci calculus, the operation that sends a vector v^i to the linear functional v_i projecting onto it is referred to as *lowering the index*.

If B is symmetric, as is often the case, then

$$B(v, w) = v_i w^i = v^i w_i.$$

A bilinear form is *non-degenerate* if the map $v^i \mapsto v_i$ is invertible. In this case, one defines B^{ij} as the coefficients of the inverse matrix of B_{ij}. It then follows that one can recover the original vector from the associated linear form as

$$v^i = B^{ij} v_j,$$

a procedure unsurprisingly called *raising the index*.

> ***Remark.*** No self-respecting mathematician would use such a terse notation as the Ricci calculus. However, in what passes for humor around here, some mathematicians have taken to using the symbol "$\flat$" for the isomorphism $V \to V^*, v \mapsto B(v, \cdot)$ that is furnished by a non-degenerate bilinear form. Likewise, "$\sharp$" is used for the inverse, and the two are referred to as the *musical isomorphisms*. Of course, this alludes to their original meaning of, uh, lowering or raising the pitch of a musical note.

concept	basis-free notation	Ricci notation	Ricci name
map $V \to V^*$, (from bilinear form)	$\flat : v \mapsto B(v, \cdot)$	$v^i \mapsto v_i = B_{ij} v^j$	lowering index
isomorphism $V^* \to V$, (from non-deg. bilinear form)	$\sharp = \flat^{-1}$	$\alpha_i \mapsto \alpha^i = B^{ij} \alpha_j$	raising index
evaluation of symmetric form	$B(v, w)$	$v_i w^j = v^i w_j$	

4.4.1 Transformation laws

Now choose a new basis $\{\tilde{e}_i\}$ of V and associated dual basis $\{\tilde{e}^j\}$ of V^*. By Sec. 2.7.2, there exists a matrix $\boldsymbol{T}$ that maps the old coordinates of a vector to the new coordinates as in $\tilde{\boldsymbol{x}} = \boldsymbol{T}\boldsymbol{x}$. Recall also that the coordinate change for linear forms is achieved by right-multiplication with the inverse matrix: $\tilde{\boldsymbol{\alpha}} = \boldsymbol{T}^{-1}\boldsymbol{\alpha}$.

In Ricci calculus, the coordinate change matrix is represented by its matrix elements ${T^i}_j$, so that:

concept	prior notation	Ricci notation	Ricci name
coordinate change of vector	$\tilde{\boldsymbol{x}} = \boldsymbol{T}\boldsymbol{x}$	$\tilde{x}^i = {T^i}_j x^j$	contravariant transformation law
coordinate change of functional	$\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\boldsymbol{T}^{-1}$	$\tilde{\alpha}_i = \alpha_j {(T^{-1})^j}_i$	covariant transformation law

Consistency check:

$$\alpha(x) = \tilde{\alpha}_i \tilde{x}^i = \alpha_j {(T^{-1})^j}_i {T^i}_k x^k = \alpha_j {(\boldsymbol{T}^{-1}\boldsymbol{T})^j}_k x^k = \alpha_j {\delta^j}_k x^k = \alpha_j x^j = \alpha(x),$$

so the contraction between a covariant and a contravariant vector is indeed invariant under coordinate change.

The same way, one verifies that the new coordinates of an N-times contravariant and M-times covariant tensor are given in terms of the old ones as

$$\tilde{K}^{i_1,\ldots,i_N}{}_{j_1,\ldots,j_M} = {T^{i_1}}_{k_1} \ldots {T^{i_N}}_{k_N} {(T^{-1})^{l_1}}_{j_1} \ldots {(T^{-1})^{l_M}}_{j_M} K^{k_1,\ldots,k_N}{}_{l_1,\ldots,l_M}$$

(I never claimed it would be fun...).

Chapter 5

Fourier analysis