Intro to linear algebra

Mathcamp 2026

Glenn Sun

These notes may contain errors,
and should not be considered an authoritative source.

Contents

1 Linear equations and transformations

1.1 The Gaussian elimination algorithm

You likely already learned how to solve systems of equations with two or three variables in middle school or high school classes. However, you likely solved systems in ad-hoc ways, with every system requiring a little creativity to get started. Real-life applications often involve systems of equations of millions of variables, and thus we have to find a way to allow computers to solve them for us. To make clear what our focus is, consider the following definition.

Definition 1.1. A *linear equation* in variables $x_1, \dots, x_n$ is of the form

$$a_1x_1 + \cdots + a_nx_n = b$$

for some numbers $a_1, \dots, a_n, b \in \mathbb{R}$. ⌟

We will always write linear equations in this form, with all variables on the left-hand side and the constant term on the right-hand side.

To illustrate the method of Gaussian elimination, consider the following system. The first step is to eliminate x from all equations other than the first, by subtracting copies of the first equation from all other equations.

$$\begin{array}{rcrcrcr} x&+&y&+&z&=&3\\ x&+&2y&+&2z&=&2\\ x&+&3y&+&5z&=&-5 \end{array} \quad\to\quad \begin{array}{lrcrcrcrl} & x&+&y&+&z&=&3&\\ & x&+&2y&+&2z&=&2&\\ -(& x&+&y&+&z&=&3&) \\ \hline & & &y&+&z&=&-1&\\ & x&+&3y&+&5z&=&-5&\\ -(& x&+&y&+&z&=&3&) \\ \hline & & &2y&+&4z&=&-8& \end{array} \quad\to\quad \begin{array}{rcrcrcr} x&+&y&+&z&=&3\\ &&y&+&z&=&-1\\ &&2y&+&4z&=&-8 \end{array}$$

Now that the second equation doesn't have x, we can use it to eliminate y from all other equations.

$$\begin{array}{rcrcrcr} x&+&y&+&z&=&3\\ &&y&+&z&=&-1\\ &&2y&+&4z&=&-8 \end{array} \quad\to\quad \begin{array}{lrcrcrcrl} & x&+&y&+&z&=&3&\\ -(& & &y&+&z&=&-1&) \\ \hline & x& & & & &=&4&\\ & & &y&+&z&=&-1&\\ & & &2y&+&4z&=&-8&\\ -2(& & &y&+&z&=&-1&) \\ \hline & & & & &2z&=&-6& \end{array} \quad\to\quad \begin{array}{rcrcrcr} x&&&&&=&4\\ &&y&+&z&=&-1\\ &&&&2z&=&-6 \end{array}$$

We see that the last equation has only z. To make things easier for ourselves, first divide the last equation by 2, so that the coefficient of z is 1. (We didn't have to do this step for previous variables because their coefficients were already 1.)

$$\begin{array}{rcrcrcr} x&&&&&=&4\\ &&y&+&z&=&-1\\ &&&&2z&=&-6 \end{array} \quad\to\quad \begin{array}{lrcrcrcrl} & x&&&&&=&4&\\ & &&y&+&z&=&-1&\\ \frac{1}{2}(& &&&&2z&=&-6&) \end{array} \quad\to\quad \begin{array}{rcrcrcr} x&&&&&=&4\\ &&y&+&z&=&-1\\ &&&&z&=&-3 \end{array}$$

Lastly, eliminate z from the second equation to get $x = 4$, $y = 2$, and $z = -3$ as the solution.

It is common to use a more efficient notation called an *augmented matrix* when performing Gaussian elimination. This notation writes down all the coefficients in a grid, leaving the variable names implicit, and places a vertical bar where the = sign should go, for legibility. In augmented matrix notation, the computation on the previous page would go as follows.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & 2 & 2 \\ 1 & 3 & 5 & -5 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & -1 \\ 0 & 2 & 4 & -8 \end{array}\right] \to \cdots \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3 \end{array}\right]$$

The word "augmented" refers to the existence of the vertical bar, and a *matrix* in general is just a grid of numbers. Though it might not be clear why yet, we'll also want to perform Gaussian elimination on general matrices, so we'll state the algorithm generally.

Definition 1.2. Given a matrix with m rows, the *Gaussian elimination* algorithm is the following procedure.

1. For $i = 1, \dots, m$,
   (a) If every number in the left part of the ith row is 0, skip this row.
   (b) Otherwise, let a be the first non-zero entry in the ith row, and let j be the column that this entry is in. Multiply the whole ith row by $1/a$, so the first non-zero entry is now 1.
   (c) Subtract copies of the ith row from all other rows to make the rest of the jth column equal to 0.
2. Optionally permute the final rows and columns, if desired. ⌟

Let us illustrate a few more examples of Gaussian elimination to see how it helps us solve different kinds of systems. For example, the algorithm could end in the following state. (The first non-zero entry of each row is 1, and those columns are otherwise clear, so there is nothing left to do.)

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 2 \\ 0 & 1 & 1 & 3 \end{array}\right]$$

In equation form, we have $x + 2z = 2$ and $y + z = 3$. In this case, z could not be eliminated, and is called a *free variable*. For *any* choice of z, setting $x = 2 - 2z$ and $y = 3 - z$ is a solution. There are infinitely many solutions, and they trace out a line in 3D space.

It is also possible for every number in the left part of a row to be 0. There are two subcases here: if the corresponding right-hand side is zero or non-zero, shown below.

$$\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 4 \end{array}\right] \qquad \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{array}\right]$$

In the first case, when the bottom-right corner is not 0, there is no solution because we have proven that 0 equals something non-zero. In the second case, when the bottom-right corner is 0, this asserts $0 = 0$ and can be ignored, so there is a unique solution.

1.2 Linear transformations and matrices

While linear equations are clearly useful and widely applicable, a slightly more general notion of a linear *transformation* is actually more central to linear algebra.

Definition 1.3. A *linear transformation* is a function $f : \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$\begin{aligned} f(x_1, \dots, x_n) = (&a_{11}x_1 + \cdots + a_{1n}x_n, \\ &\cdots, \\ &a_{m1}x_1 + \cdots + a_{mn}x_n) \end{aligned}$$

for some $a_{11}, \dots, a_{mn} \in \mathbb{R}$. ⌟

This generalizes systems of equations because a system such as $x + y = 1, x - y = 3$ can be separated into the linear transformation $f(x, y) = (x + y, x - y)$ and the question "What x and y satisfy $f(x, y) = (1, 3)$?" Note that the transformation f itself does not include any constants. For example, $f(x, y) = (x + y - 1, x - y - 3)$ is *not* a linear transformation.

Linear transformations are useful beyond systems of equations, and this is the whole point of studying them. Here are a few examples:

- In an image processing system, rotating an image is accomplished by a linear transformation. For example, $f(x, y) = (-y, x)$ is a linear transformation that rotates points 90° counterclockwise.
- The linear transformation $f(x, y) = (x + y, x)$ iterates through the Fibonacci sequence as follows.

$$(1, 1) \xrightarrow{f} (2, 1) \xrightarrow{f} (3, 2) \xrightarrow{f} (5, 3) \xrightarrow{f} (8, 5)$$

- Make a recording of a musical chord on your computer, and suppose you wanted to find what notes were being played. You can do this by applying a linear transformation: a particular one called the Fourier transform, where the variables $x_1, \dots, x_n$ represent the long list of audio wave samples, and each coordinate outputs how much of a certain frequency occurs in the audio.
- In quantum mechanics and quantum computing, the state of a system can only be modified by linear transformations (in particular, a certain kind of linear transformation known as *unitary*). This is essentially the superposition principle, but we'll discuss the details later.

When discussing linear transformations, we typically use the more compact notation of a (unaugmented) *matrix*, just like with systems of equations. We call the matrix below an $m \times n$ matrix, stating the number of rows first. We use capital letters like A, B, and C to denote matrices.

$$\begin{aligned} f(x_1, \dots, x_n) = (&a_{11}x_1 + \cdots + a_{1n}x_n, \\ &\cdots, \\ &a_{m1}x_1 + \cdots + a_{mn}x_n) \end{aligned} \quad \Leftrightarrow \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

The matrix A corresponds to the function f, but we also need a notation for the output value of f given inputs $x_1, \dots, x_n$. We use the following notation, where symbols like $\vec{x}$, $\vec{y}$, and $\vec{z}$ denote points.

$$A\vec{x} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}$$

In other words, points may be written with their list of coordinates presented vertically, rather than horizontally. Then we omit the parentheses and just write $A\vec{x}$ to mean "apply the linear transformation corresponding to A on $\vec{x}$." This is sometimes known as *matrix-vector multiplication*, but we won't use this phrase yet, since we have not yet defined vectors nor explained why this should be considered a multiplication.

The vertical presentation of the point $\vec{x}$ is not just an aesthetic choice—it is intended to be consistent with how we wrote A. In particular, each coordinate in the output of A is a new row, so each coordinate in $\vec{x}$ should also be a new row.

1.3 Composition of linear transformations

Functions in general can be *composed*. For example, if $f(x) = x^2$ and $g(x) = x + 1$, we can compose them to get $f(g(x)) = (x + 1)^2$, or compose them the other way to get $g(f(x)) = x^2 + 1$. The new function obtained by doing $f(g(x))$ is often denoted $f \circ g$, meaning $f(g(x)) = (f \circ g)(x)$. An important note is that while function composition is clearly not commutative, it *is* associative: $(f \circ g) \circ h = f \circ (g \circ h)$.

The special thing about linear transformations is that the composition of two linear transformations is still a linear transformation! In other words, it is not possible to introduce quadratic terms, constants, or other things by composing linear transformations.

When composing functions, be careful to make sure that the output space of each function matches the input space of the next function. For example, if $f(x, y) = x + y$ (a function $\mathbb{R}^2 \to \mathbb{R}$), it would not make sense to compose f with itself, since f outputs a single number and expects two numbers as input.

Proposition 1.4. *If f and g are linear transformations that can be composed, then their composition $f \circ g$ is a linear transformation.* ⌟

Proof. We'll prove for linear transformations $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^2$. The argument for other linear transformations is very similar.

We'll write the proof using matrices. You could do this directly using the original definition of linear transformation, but this is a little clearer. Write f and g in matrix form as follows.

$$f(x, y) = \begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad g(x, y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Then we just evaluate the linear transformations at the points:

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

$$= \begin{bmatrix} p(ax+by)+q(cx+dy) \\ r(ax+by)+s(cx+dy) \end{bmatrix}$$

$$= \begin{bmatrix} (pa+qc)x+(pb+qd)y \\ (ra+sc)x+(rb+sd)y \end{bmatrix}$$

$$= \begin{bmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

In the end, we see that the result is a linear transformation applied to a point (x, y). $\square$

We take the matrix of the resulting composition above to be the definition of the "matrix composition": writing A and B for the respective matrices, AB denotes the matrix of the composition. It can be computed by taking each (i, j)th entry to combine the ith row of A with the jth column of B as follows.

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{bmatrix}$$

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{bmatrix}$$

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{bmatrix}$$

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} pa+qc & pb+qd \\ ra+sc & rb+sd \end{bmatrix}$$

The composition of linear transformations is often known as *matrix multiplication*. Again, we have not yet explained why this should be considered a kind of multiplication, but you should anyway understand this process as fundamentally composition.

The fully general formula for composing two matrices is as follows. Given a $k \times m$ matrix A with entries a_{ij} and an $m \times n$ matrix B with entries $b_{j\ell}$, the matrix $C = AB$ is a $k \times n$ matrix whose entries are given by:

$$c_{i\ell} = a_{i1}b_{1\ell} + \cdots + a_{im}b_{m\ell}.$$

Homework

1.A Practice

These problems are recommended to practice key concepts from class. You can skip them if you feel comfortable already.

1.A.1 Matrix multiplication

For each item below, perform the computation or say that the question is ill-formed. (Note for part 4: Matrix multiplication is associative just like function composition, so computing left-to-right and right-to-left are both fine. But which order reduces the amount of computation you need to do?)

1. $\begin{bmatrix} 2 & -1 & 0 \\ 3 & 0 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 \\ 0 & -2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} -1 & 3 & 2 \\ 0 & 1 & -2 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 0 \\ -4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & -1 \\ 0 & 2 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

1.A.2 Solving systems of equations

Solve each system using the Gaussian elimination method, and describe the set of solutions.

1. $$\begin{aligned} x + y + z &= 6 \\ 2x - y + z &= 3 \\ x + 2y - z &= 2 \end{aligned}$$

2. $$\begin{aligned} 2x - y - z &= -4 \\ -2x + 2y + z &= 7 \end{aligned}$$

1.A.3 Unambiguous notation

We defined the notation for evaluating a matrix at a point to be:

$$A\vec{x} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}.$$

Then, we used essentially the same notation of placing two things in square brackets next to each other to mean matrix multiplication. Let X be the $n \times 1$ *matrix* with entries $x_1, \dots, x_n$. Perform the matrix multiplication

$$AX = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

to show that the notation really is not ambiguous.

1.B Extensions

These problems will assist you in discovering new ideas, and are recommended. Problems that will be referenced in the future are noted with $(*)$.

1.B.1 Elementary matrices $(*)$

The three key operations used in Gaussian elimination are: multiplying a row by a constant, subtracting some multiple of one row from another row, and swapping two rows. For this problem, for concreteness, let A be a 3×3 matrix.

1. Find the matrix E such that EA is exactly the result of multiplying row 1 of A by a constant c.
2. Find the matrix E such that EA is exactly the result of subtracting c times row 1 of A from row 2 of A.
3. Find the matrix E such that EA is exactly the result of swapping rows 1 and 2 of A.
4. The three kinds of matrices above (for various specific rows) are called *elementary matrices*. Describe generally how one can think of the Gaussian elimination algorithm as finding matrices $E_1, \ldots, E_t$ such that

$$E_t \cdots E_1 A = B,$$

   where B is a matrix satisfying: every row is either all 0 or has 1 as the first non-zero entry, and every column containing such a 1 has all other entries as 0.

1.B.2 Matrix sums $(*)$

Given two matrices A and B of the same size, we can define $A + B$ by adding their corresponding entries. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 4 & 3 \end{bmatrix}.$$

Show that for any matrices A, B, and C of appropriate size, $(A + B)C = AC + BC$ and $A(B + C) = AB + AC$.

2 Saying the same thing in different ways

2.1 A deeper dive into Gaussian elimination

Recall that Gaussian elimination helps us solve systems of linear equations, for which there might be a unique solution, multiple solutions, or no solution. There may be free variables, and there may be rows of zeros. Our first goal is to precisely understand what the output of Gaussian elimination looks like, in order to help us use the same algorithm to solve more problems.

The first observation is that at the end of the algorithm, one can always permute the rows and columns of the matrix to look like the following, where $*$ denotes an entry that can be anything.

$$\left[\begin{array}{cccccc|ccc} 1 & \cdots & 0 & * & \cdots & * & * & \cdots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & * & \cdots & * & * & \cdots & * \\ 0 & \cdots & 0 & 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & * & \cdots & * \end{array}\right]$$

This is often written more compactly as

$$\left[\begin{array}{cc|c} I & * & * \\ 0 & 0 & * \end{array}\right],$$

where I denotes the (unaugmented) matrix with 1 on the diagonal and 0 elsewhere. (I stands for *identity*, it is the matrix satisfying $I\vec{x} = \vec{x}$.) Note that each of these six blocks may be of any size, or even not exist at all.

Note that permuting the rows is completely safe, since the order of the equations doesn't matter. Permuting the columns means reordering the variables, e.g. placing z before y. If this is done, remember the permutation in order to correctly read the solution. Then for the purpose of solving systems of equations, we have the following analysis:

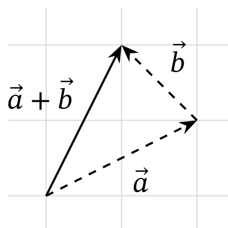
- Suppose there are no free variables (no $*$ block on the left side), so the final matrix looks like
$$\left[\begin{array}{c|c} I & * \\ 0 & * \end{array}\right].$$
(Again, this block notation allows blocks to be of any size, including not existing.) Then it is not possible to have multiple solutions. It is possible to have a unique solution or no solution.
- Suppose there is no row of zeros (no 0 block on the left side), so the final matrix looks like
$$\left[\begin{array}{cc|c} I & * & * \end{array}\right].$$
(Again, this block notation allows blocks to be of any size, including not existing.) Then it is not possible to have no solutions. It is possible to have a unique solution or multiple solutions.
- Putting the two cases together, if you get $\left[\begin{array}{c|c} I & * \end{array}\right]$, you know for sure there is a unique solution.

These cases will be useful to analyze in other applications of Gaussian elimination in the rest of this section.

2.2 Vectors, linear combinations, and span

In the previous section, we were careful to refer to expressions like $(2, -1)$ as *points*, since the only way we used them was by applying transformations to them. However, an important idea in linear algebra is that we can actually add points together. When we think of points as being able to be added together and scaled by constants, we use the word *vector*.

We add and scale vectors in $\mathbb{R}^n$ componentwise. For example, $(2, 1)+(-1, 1) = (1, 2)$ and $3 \cdot (2, 1) = (6, 3)$. To visualize this, the typical idea in $\mathbb{R}^2$ is to treat (x, y) as an arrow that goes x units right and y units up.

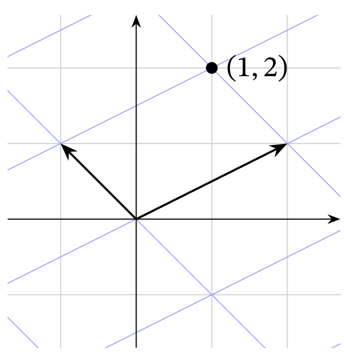


Definition 2.1. A *linear combination* of vectors $\vec{x}_1, \dots, \vec{x}_k$ is an expression of the form $a_1\vec{x}_1+\cdots+a_k\vec{x}_k$ for some constants $a_1, \dots, a_k \in \mathbb{R}$. The *span* of a set of vectors $\vec{x}_1, \dots, \vec{x}_k$ is the set of all possible linear combinations, denoted $\operatorname{span}(\vec{x}_1, \dots, \vec{x}_k)$. We say that $\vec{x}_1, \dots, \vec{x}_k$ is a *spanning set* for $\mathbb{R}^n$ if $\operatorname{span}(\vec{x}_1, \dots, \vec{x}_k) = \mathbb{R}^n$. ⌟

For example, the picture below shows that $(2, 1)$ and $(-1, 1)$ form a spanning set for $\mathbb{R}^2$. Instead of expressing a point in standard coordinates, one could instead express any point as a multiple of $(2, 1)$ plus a multiple of $(-1, 1)$, i.e. a linear combination. For example, the marked point can be expressed as

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

by following the blue gridlines.

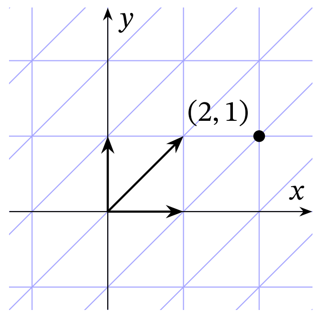


Not every set of vectors is spanning. For instance, the span of a single (nonzero) vector is just a line through the origin. Two vectors can span at most a plane through the origin, which is not the whole space in $\mathbb{R}^3$ or higher dimensions.

It's also possible that a single point could be expressed as a linear combination in multiple ways. For example, $(1, 0)$, $(0, 1)$ and $(1, 1)$ also span $\mathbb{R}^2$. Intuitively, since we have 3 vectors in 2 dimensions, one of them is redundant. We will formalize the idea of redundancy shortly, but for now, just observe that this allows us to express

$$\begin{bmatrix}2\\1\end{bmatrix} = 1\cdot\begin{bmatrix}1\\0\end{bmatrix} + 0\cdot\begin{bmatrix}0\\1\end{bmatrix} + 1\cdot\begin{bmatrix}1\\1\end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix}2\\1\end{bmatrix} = 2\cdot\begin{bmatrix}1\\0\end{bmatrix} + 1\cdot\begin{bmatrix}0\\1\end{bmatrix} + 0\cdot\begin{bmatrix}1\\1\end{bmatrix}$$

or many other ways.



This dance between unique solutions, no solutions, and many solutions should feel familiar by now, and indeed there is a strong connection. View the definition of linear combination as a matrix expression instead.

$$a_1\vec{x}_1 + \cdots + a_k\vec{x}_k = \vec{b} \qquad \Leftrightarrow \qquad \begin{bmatrix}\vec{x}_1 & \cdots & \vec{x}_k\end{bmatrix}\begin{bmatrix}a_1\\ \vdots \\ a_k\end{bmatrix} = \vec{b}$$

In this notation, we are considering a matrix whose columns are $\vec{x}_1, \dots, \vec{x}_k$. The matrix has as many rows as each $\vec{x}_i$ has entries. But now it is clear: expressing $\vec{b}$ as a linear combination of vectors is equivalent to finding solutions to a system of equations! We read the rows of any matrix as coefficients for equations, and we can also read the columns of any matrix as vectors with which to take linear combinations.

Proposition 2.2. *Let $\vec{a}_1, \dots, \vec{a}_k \in \mathbb{R}^n$ be the columns of A. Then the following are equivalent:*

- *After Gaussian elimination on A, the resulting matrix has no rows of zeros.*
- *The system of equations $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b} \in \mathbb{R}^n$.*
- *The vectors $\vec{a}_1, \dots, \vec{a}_k$ form a spanning set for $\mathbb{R}^n$.* ⌟

2.3 Linear independence and basis

In the example with three vectors $(1,0)$, $(0,1)$, and $(1,1)$ spanning $\mathbb{R}^2$, there is clearly some redundancy here, and not all three vectors are needed to span the space. That is also why every vector in $\mathbb{R}^2$ could be expressed as a linear combination of these three in multiple ways. Let's formalize what we mean by redundancy.

Definition 2.3. A set of vectors $\vec{x}_1, \dots, \vec{x}_k \in \mathbb{R}^n$ is *linearly independent* if no vector can be removed without changing the span. The vectors are *linearly dependent* otherwise. ⌟

For example, take the vectors $(1,0,0)$, $(0,1,0)$, $(1,1,0)$, and $(0,0,1)$. The first three vectors live on the x-y plane, and the fourth points in the positive z direction. These vectors are linearly dependent. One reason is because the vector $(1,1,0)$ can be removed without changing the span. The remaining vectors $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ are linearly independent.

Proposition 2.4. *Vectors $\vec{x}_1, \dots, \vec{x}_k$ are linearly dependent if and only if there exist $a_1, \dots, a_k$, not all 0, such that $a_1\vec{x}_1 + \cdots + a_k\vec{x}_k = \vec{0}$.* ⌟

Proof. Suppose $a_1\vec{x}_1 + \cdots + a_k\vec{x}_k = \vec{0}$ with coefficients not all 0. Then for any $a_i \neq 0$, this can be rearranged into

$$\vec{x}_i = -\sum_{j\neq i} \frac{a_j}{a_i}\vec{x}_j.$$

Since $\vec{x}_i$ can be written as a linear combination of the other vectors, it can be removed without changing the span. Conversely, the same argument applies in reverse. □

As a corollary, we get the counterpart to Proposition 2.2. In particular, $A\vec{x} = \vec{0}$ has non-trivial solutions if and only if there are free variables, because otherwise Gaussian elimination keeps the right-hand side always 0.

Proposition 2.5. *Let $\vec{a}_1, \dots, \vec{a}_k \in \mathbb{R}^n$ be the columns of A. Then the following are equivalent:*

- *After Gaussian elimination on A, the resulting matrix has no free variables.*
- *The system of equations $A\vec{x} = \vec{b}$ has at most one solution for each $\vec{b} \in \mathbb{R}^n$.*
- *The system of equations $A\vec{x} = \vec{0}$ has no non-trivial solutions.*
- *The vectors $\vec{a}_1, \dots, \vec{a}_k$ are linearly independent.* ⌟

Putting this together, we can now discuss the perfect case in the middle, corresponding to unique solutions.

Definition 2.6. A *basis* of $\mathbb{R}^n$ is a set of vectors that is both a spanning set of $\mathbb{R}^n$ and linearly independent. ⌟

Proposition 2.7. *Every spanning set of $\mathbb{R}^n$ has $\geq n$ vectors, and every linearly independent set of $\mathbb{R}^n$ has $\leq n$ vectors. (Thus, every basis of $\mathbb{R}^n$ must have exactly n vectors.)* ⌟

Proof. Let A be the matrix whose columns are the set of vectors in question. (Note that A is not augmented here.) By the previous propositions, spanning implies that Gaussian elimination produces the form

$$\begin{bmatrix} I & * \end{bmatrix}.$$

Since I is square, this implies there are $\geq n$ vectors. Likewise, linear independence implies that Gaussian elimination produces the form

$$\begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Since I is square, this implies there are $\leq n$ vectors. □

We conclude with a mega-theorem that puts everything together.

Theorem 2.8. *Let $\vec{a}_1, \dots, \vec{a}_n \in \mathbb{R}^n$ be a list of n vectors forming the columns of A. Then the following are equivalent:*

- *After Gaussian elimination on A, the resulting matrix has no rows of zeros.*
- *After Gaussian elimination on A, the resulting matrix has no free variables.*
- *After Gaussian elimination on A, the resulting matrix is exactly I on the left.*
- *The system of equations $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b} \in \mathbb{R}^n$.*
- *The system of equations $A\vec{x} = \vec{b}$ has at most one solution for each $\vec{b} \in \mathbb{R}^n$.*
- *The system of equations $A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{R}^n$.*
- *The system of equations $A\vec{x} = \vec{0}$ has no non-trivial solutions.*
- *The vectors $\vec{a}_1, \dots, \vec{a}_n$ form a spanning set for $\mathbb{R}^n$.*
- *The vectors $\vec{a}_1, \dots, \vec{a}_n$ are linearly independent.*
- *The vectors $\vec{a}_1, \dots, \vec{a}_n$ form a basis.* ⌟

The *standard basis* of $\mathbb{R}^n$ is given by vectors $\vec{e}_i$, where the ith component of $\vec{e}_i$ is 1 and the other components are 0. For example, over $\mathbb{R}^3$, we have

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

2.4 Surjective, injective, and invertible

One final view that we'll discuss uses the concepts of surjective, injective, and invertible (bijective). These are properties that can generally apply to all functions.

- A function $f : X \to Y$ is *surjective* if for all $y \in Y$, there is at least one x such that $f(x) = y$.
- A function $f : X \to Y$ is *injective* if for all $y \in Y$, there is at most one x such that $f(x) = y$. (Sometimes stated equivalently, if $f(x) = f(x')$ implies $x = x'$.)
- A function $f : X \to Y$ is *invertible* or *bijective* if for all $y \in Y$ there is exactly one x such that $f(x) = y$. We use this to define the *inverse* $f^{-1}(y) = x$. For example, the inverse of $f : [0, \infty) \to [0, \infty)$ given by $f(x) = x^2$ is $f^{-1}(x) = \sqrt{x}$.

Applied to a linear transformation $f(\vec{x}) = A\vec{x}$, these are exactly equivalent to the conditions referring to solutions to the system of equations $A\vec{x} = \vec{b}$.

Proposition 2.9. *Let A be a matrix. The following are equivalent:*

- *The items in Proposition 2.2.*
- *The linear transformation $f(\vec{x}) = A\vec{x}$ is surjective.* ⌟

Proposition 2.10. *Let A be a matrix. The following are equivalent:*

- *The items in Proposition 2.5.*

- *The linear transformation $f(\vec{x}) = A\vec{x}$ is injective.* ⌟

Theorem 2.11. *Let A be an $n \times n$ matrix. The following are equivalent:*

- *The items in Theorem 2.8.*
- *The linear transformation $f(\vec{x}) = A\vec{x}$ is surjective.*
- *The linear transformation $f(\vec{x}) = A\vec{x}$ is injective.*
- *The linear transformation $f(\vec{x}) = A\vec{x}$ is invertible (bijective).* ⌟

An important fact about functions in general is that the inverse of a function f is the unique function g satisfying $f \circ g = g \circ f = \mathrm{id}$, where id denotes the identity function $\mathrm{id}(x) = x$. (Try to prove it if you haven't seen this fact before!)

A priori, it is not clear whether the inverse of a linear transformation is itself linear. This uniqueness fact suggests that we make the following definition.

Definition 2.12. The *inverse* of a matrix A, if one exists, is the unique matrix A^{-1} satisfying $AA^{-1} = A^{-1}A = I$, where I denotes the identity matrix with 1 on the diagonal and 0 elsewhere. ⌟

For example, as mentioned earlier, $f(x, y) = (-y, x)$ is a 90° counterclockwise rotation. Its inverse is the clockwise rotation $f^{-1}(x, y) = (y, -x)$. In matrix form,

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Two questions remain. First, though A^{-1} is a matrix by definition, it is not clear given $f(\vec{x}) = A\vec{x}$, whether or not f^{-1} can exist and be non-linear so that A^{-1} does not exist. Second, and perhaps more pressing, is that we have not yet given a method to calculate the inverse. Both of these questions will be resolved on the homework.

Homework

2.A Practice

These problems are recommended to practice key concepts from class. You can skip them if you feel comfortable already.

2.A.1 Independence, span, basis, dimension

For each of the following sets of vectors, determine if they are linearly independent, spanning $\mathbb{R}^n$, and/or a basis.

1. $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}$
2. $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}3\\1\\2\end{bmatrix}$
3. $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}2\\3\\1\end{bmatrix}$
4. $\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}$
5. $\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}2\\1\\3\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}$
6. $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$

2.A.2 Creating a basis

1. Given a linearly independent set of vectors in $\mathbb{R}^n$, suggest a method to add vectors to extend the set to a basis of $\mathbb{R}^n$.
2. Given a spanning set of vectors for $\mathbb{R}^n$, suggest a method to remove vectors until the set is a basis of $\mathbb{R}^n$.

2.A.3 Left and right inverse

Given a matrix A, a matrix B such that $AB = I$ is known as a *right inverse*, and a matrix C such that $CA = I$ is known as a *left inverse*. Prove that:

1. A has a right inverse if and only if the items in Proposition 2.2 hold.
2. A has a left inverse if and only if the items in Proposition 2.5 hold.
3. If A is $n \times n$, then A has a right inverse if and only if A has a left inverse, if and only if the items in Theorem 2.8 hold.

(Notably, while the equivalence between left inverse and injective is easy to prove as a basic property of functions, the equivalence between right inverse and surjective for functions in general requires the Axiom of Choice! This issue will not happen in our case focusing on matrices and linear transformations.)

2.B Extensions

These problems will assist you in discovering new ideas, and are recommended. Problems that will be referenced in the future are noted with ($*$).

2.B.1 Properties of the inverse

1. Let A be an $n \times n$ matrix. Suppose $AB = I$. Prove that $BA = I$. Show that this may not be true for non-square matrices.
2. Suppose A and B are both invertible $n \times n$ matrices. Prove that AB is invertible by giving a formula for the inverse, and checking it.

2.B.2 Computing inverses

Let $A = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ and denote $A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. In this problem, you will compute A^{-1}.

1. Using $AA^{-1} = I$, write and solve a system of linear equations for a and c. Use Gaussian elimination.
2. Using $AA^{-1} = I$, write and solve a system of linear equations for b and d. Use Gaussian elimination.
3. Come up with a way to do parts (1) and (2) above at the same time with Gaussian elimination, avoiding duplicate work. State your algorithm generally for an $n \times n$ matrix. (You should *not* have to solve a system of n^2 equations.)
4. Analyze the 4 output cases (the bijective, surjective, injective, and general cases, again remarking that we have not explained these names fully yet). How would it detect when a matrix is not invertible?
5. Prove that if a linear transformation $f(\vec{x}) = A\vec{x}$ is invertible as a function, then its inverse must be linear (in particular, be A^{-1}).

2.B.3 Subspaces

A subset $U \subseteq \mathbb{R}^n$ is a *subspace* if it is nonempty and closed under addition and scalar multiplication: for all $\vec{x}, \vec{y} \in U$ and all $c \in \mathbb{R}$, we have $\vec{x} + \vec{y} \in U$ and $c \cdot \vec{x} \in U$. Intuitively, these are spaces such as the set $\{\vec{0}\}$, a line through the origin, a plane through the origin, or higher dimensional analogues.

1. Prove that for all vectors $\vec{x}_1, \dots, \vec{x}_k$, the set $\operatorname{span}(\vec{x}_1, \dots, \vec{x}_k)$ is a subspace.
2. Provide reasonable definitions that extend the definition of spanning set and basis to all subspaces of $\mathbb{R}^n$, not just $\mathbb{R}^n$ itself as we originally defined it. (Linear independence already does not depend on the subspace.)
3. Prove that every subspace U has a basis. (Hint: Consider a maximal independent set contained within U, and prove that it is a spanning set.)
4. The *dimension* of a subspace is the number of vectors in a basis. Compute the dimensions of all the spans of Problem 2.A.1.

2.B.4 Rank-nullity theorem

The rank-nullity theorem says the following. Let A be an $m \times n$ matrix, so it takes in n-dimensional vectors and produces m-dimensional ones. The *rank* of A is the dimension of the span of the columns of A. The *nullity* of A is the dimension of $\{\vec{x} \mid A\vec{x} = \vec{0}\}$. The rank-nullity theorem says that the dimensions always add up to n.

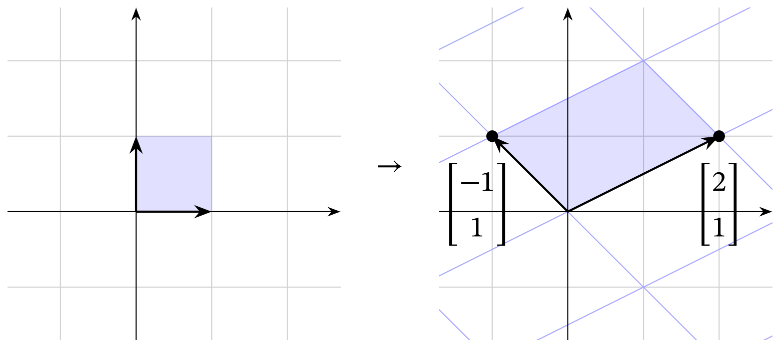
1. Prove that $\{\vec{x} \mid A\vec{x} = \vec{0}\}$ is a subspace of $\mathbb{R}^n$. (Thus the nullity is well-defined.)
2. Prove the theorem in the special case where the rank of A is n.
3. Prove the theorem in general. Hint: consider solving $A\vec{x} = \vec{0}$ using Gaussian elimination. How would you write down a basis of the nullity using the free variables?

3 Visualizing matrices

3.1 Matrices transforming the standard basis

The standard way to visualize a linear transformation is by considering how it transforms the standard basis. Check that if $\vec{a}_1, \dots, \vec{a}_n$ are the columns of A, then $A\vec{e}_i = \vec{a}_i$. In words, the columns of a matrix are exactly the images of the standard basis. Here is an example.

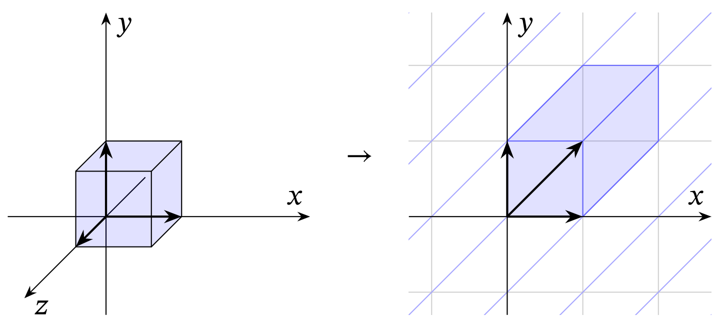
$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$:



In this picture, the original gridlines have also been transformed. In general, this is what a linear transformation does: stretch, skew, and rotate the space around the origin, taking everything along with it.

Consider a second example, which compresses 3 dimensions into 2 dimensions. The following transformation keeps the first two basis vectors of $\mathbb{R}^3$ fixed, and maps the basis vector corresponding to the z direction to $(1, 1)$.

$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$:



As a first application of this idea, we will prove an important alternative characterization of linear transformations. In fact, many people take the following to be the *definition* of a linear transformation.

Proposition 3.1. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if for all $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $f(\vec{x} + \vec{y}) = f(\vec{x}) + f(\vec{y})$ and $f(c \cdot \vec{x}) = c \cdot f(\vec{x})$.*

Proof. If f is linear, plug in the definition and apply distributivity to conclude the result. The converse direction is more interesting. Suppose f satisfies the two properties. We guess the matrix of f to be $\begin{bmatrix} f(\vec{e}_1) & \cdots & f(\vec{e}_n) \end{bmatrix}$. To prove this, we simply compute that

$$\begin{aligned} \begin{bmatrix} f(\vec{e}_1) & \cdots & f(\vec{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} &= x_1 f(\vec{e}_1) + \cdots + x_n f(\vec{e}_n) \\ &= f(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) \\ &= f(\vec{x}). \end{aligned}$$

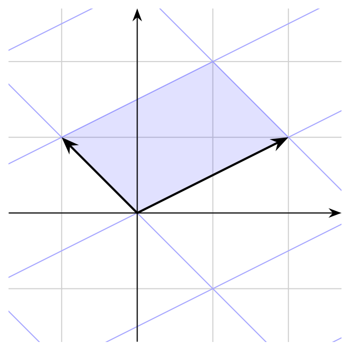
□

3.2 Determinant

When you read the columns of a matrix as the images of the standard basis, it is common to shade in the parallelogram (or higher dimensional analogue) given by these columns. The area of this parallelogram represents how much "blow-up" the matrix performs, and is called the *determinant*.

As a technical detail, the determinant is actually the *signed area*, meaning the order of the vectors matters. In 2 dimensions, if the counterclockwise angle from the first vector to the second is less than π, the determinant is positive, and otherwise the determinant is negative. In general, swapping the order of two vectors multiplies the determinant by -1. Here is an example.

$$\det\left(\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}\right) = 3$$



$$\det\left(\begin{bmatrix} -1 & 2 \\ 1 & 1 \end{bmatrix}\right) = -3$$

Definition 3.2. Let A be an $n \times n$ matrix with columns $\vec{a}_1, \ldots, \vec{a}_n$. The *determinant* of A is the signed volume of the parallelepiped given by $\vec{a}_1, \ldots, \vec{a}_n$. More specifically,

- $\det(I) = 1$.
- Suppose A' is the same as A, except with $c \cdot \vec{a}_i$ instead of $\vec{a}_i$, for some $c \in \mathbb{R}$. Then $\det(A') = c \cdot \det(A)$.
- Suppose A' is the same as A, except with $\vec{a}_i + c \cdot \vec{a}_j$ instead of $\vec{a}_i$, for some $c \in \mathbb{R}$ and $j \neq i$. Then $\det(A') = \det(A)$.
- Suppose A' is the same as A, except two columns are swapped. Then $\det(A') = -\det(A)$. This is what is meant by *signed* area. ⌟

As a first note, the definition immediately implies that any matrix with a column of zeros has determinant 0. Intuitively, that means there are only $n-1$ vectors left, and the n-dimensional volume of any $(n-1)$-dimensional thing is 0 (e.g. a line has zero area, a plane has zero volume). Formally, if a column is $\vec{0}$, scale this column by 0, which doesn't change the matrix. But the definition says that the determinant gets multiplied by 0 during this scaling, so it must have been 0.

This definition also almost exactly corresponds to the operations allowed in Gaussian elimination, except that it refers to columns instead of rows. For ease of computation and compatibility with our previous theory, we would have really preferred rows. However, this is not a problem at all. To prove this, we'll also use an important lemma that the determinant is multiplicative. In any case, the statement of the proposition is much more important than the proof.

Definition 3.3. The *transpose* of a matrix A is denoted A^T, and refers to the matrix whose (i,j)th entry is the (j,i)th entry in A. In other words, A^T is the same as A except flipped across its diagonal. ⌟

An important identity is that $(AB)^T = B^T A^T$. To understand this intuitively, recall that AB matches rows of A with columns of B. Now, $B^T A^T$ matches rows of B^T (columns of B) with columns of A^T (rows of A), so the result is the same, up to another transpose. It is not difficult but tedious to comb through the index notation to prove this formally.

Proposition 3.4. *Let A and B be $n \times n$ matrices.*

1. $\det(AB) = \det(A)\det(B)$.
2. $\det(A) = \det(A^T)$.

(Thus, the rules to compute the determinant could have equivalently discussed the rows.) ⌟

Proof sketch. Both of these facts can be proven using the same idea, that invertible matrices can be expressed as products of elementary matrices. In particular, first treat separately the case of non-invertible matrices and determinant 0, left to the reader. Then,

1. To prove $\det(AB) = \det(A)\det(B)$, first check that $\det(AE) = \det(A)\det(E)$ for all elementary matrices E by just computing it. (Note that right-multiplication by E corresponds to column operations, whereas a previous homework discussed left-multiplication by E corresponding to row operations.) Then, express $B = E_1 \cdots E_t$ and induct on t to conclude $\det(AB) = \det(A)\det(B)$.
2. To prove $\det(A) = \det(A^T)$, first compute explicitly that $\det(E) = \det(E^T)$ for all elementary matrices E. Then,

$$\det(A) = \det(E_1 \cdots E_t) = \det(E_1)\cdots\det(E_t) = \det(E_t^T \cdots E_1^T) = \det(A^T). \qquad \square$$

Now, we can compute the determinant using our familiar Gaussian elimination technique. Here is one example.

$$\begin{aligned}
\det\left(\begin{bmatrix} 2 & 4 \\ 1 & 5 \end{bmatrix}\right) &= 2 \cdot \det\left(\begin{bmatrix} 1 & 2 \\ 1 & 5 \end{bmatrix}\right) && \text{(scale first row by } 1/2\text{)} \\
&= 2 \cdot \det\left(\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}\right) && \text{(subtract first row from second)} \\
&= 6 \cdot \det\left(\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\right) && \text{(scale second row by } 1/3\text{)} \\
&= 6 \cdot \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) && \text{(subtract } 2\times \text{ second row from first)} \\
&= 6
\end{aligned}$$

We note a trick to reduce the amount of computation needed for the determinant. It suffices as a first step to subtract $0.5\times$ the first row from the second, resulting in

$$\det\left(\begin{bmatrix} 2 & 4 \\ 1 & 5 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix}\right)$$

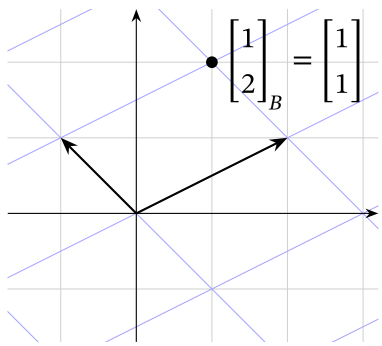
Now that the matrix has no entries below the main diagonal, we know it suffices to scale the rows and use them to clear out the entries above. Thus, the determinant is the product of the entries on the diagonal, here $2 \cdot 3 = 6$.

3.3 Vectors and matrices in other bases

The last way we'll visualize matrices is by considering what they look like in bases other than the standard basis. To clarify what this means, let us first visualize vectors.

Definition 3.5. Given an invertible matrix B and a vector $\vec{x}$, the notation $[\vec{x}]_B$ refers to the vector of coefficients used to express $\vec{x}$ in the basis of the columns of B. ⌟

This is a picture we've seen before, but as a reminder, if B is the matrix with columns $(2,1)$ and $(-1,1)$, then the vector $\vec{x} = (1,2)$ can be expressed as $1\cdot(2,1)+1\cdot(-1,1)$. Thus $[\vec{x}]_B = (1,1)$. (Note that this is unfortunately the opposite of the standard convention in number theory, where the subscript x_b means that x is already written in base b, e.g. $13_7 = 10$. Here, $[\vec{x}]_B$ means that $\vec{x}$ is currently written in standard coordinates, and the result $[\vec{x}]_B$ is the vector of coefficients.)



As a general formula, recall that a linear combination with coefficients $[\vec{x}]_B$ can be written as $B[\vec{x}]_B = \vec{x}$. Thus, we get the formula

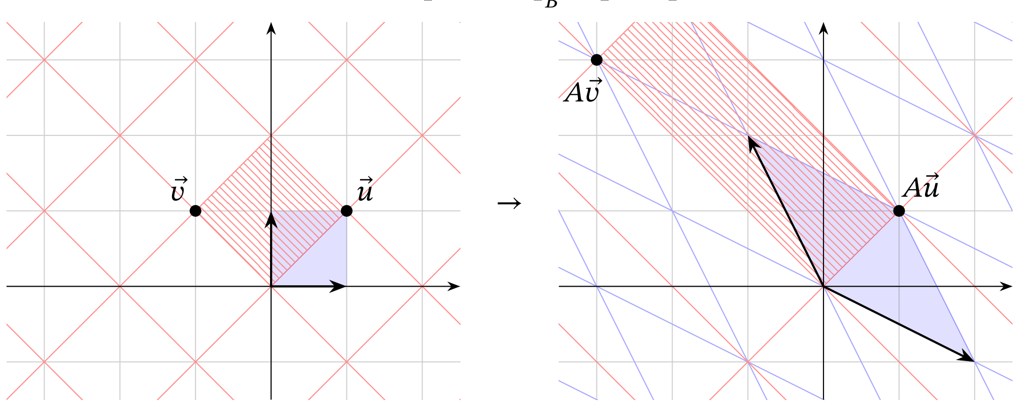
$$[\vec{x}]_B = B^{-1}\vec{x}.$$

Remember that B sends coefficients of columns of B to the vectors' actual positions. B^{-1} sends actual positions to coefficients. These directions are easy to confuse.

Definition 3.6. Given an invertible matrix B and a matrix A, the notation $[A]_B$ refers to the matrix given by $[A]_B[\vec{x}]_B = [A\vec{x}]_B$. In other words, $[A]_B$ is the same transformation as A, except working entirely in B-coordinates for both input and output. ⌟

In the picture below, let B be the matrix with columns $\vec{u}$ and $\vec{v}$. The picture shows a matrix A being applied, with the standard coordinate view in blue and the B-basis view in red. We see that

$$[A]_B = \begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}_B = \begin{bmatrix}1 & 0\\0 & 3\end{bmatrix}.$$



To explain the picture a bit more, the standard coordinate view is exactly what we are familiar with. The standard basis vectors go to the columns of A. Now tilt your head and look at the same transformation, except now with the basis vectors given by B as the starting point. The transformation looks exactly like the matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix},$$

meaning the first basis vector $\vec{u}$ stays put and the second basis vector $\vec{v}$ is scaled up by 3.

To calculate $[A]_B$, simply start with the definition and expand to get a formula.

$$\begin{aligned} [A]_B[\vec{x}]_B = [A\vec{x}]_B \quad &\Leftrightarrow \quad [A]_B B^{-1}\vec{x} = B^{-1}A\vec{x} \\ &\Leftrightarrow \quad [A]_B B^{-1} = B^{-1}A \\ &\Leftrightarrow \quad [A]_B = B^{-1}AB \end{aligned}$$

Recalling that functions apply right-to-left, the diagram corresponding to the above formula is as follows.

$$\begin{tikzcd} \text{coefficients} \arrow[r, "{[A]_B}"] \arrow[d, "B"'] & \text{coefficients} \\ \text{actual position} \arrow[r, "A"'] & \text{actual position} \arrow[u, "B^{-1}"'] \end{tikzcd}$$

Thinking about matrices in other bases is important because much of the nice structure of a linear transformation may be hidden when the wrong basis is used. For instance, the example above showed a matrix that did not appear particularly well-structured in the standard basis, but once we changed to the B basis, we saw that the matrix is simply scaling along the basis directions, with no rotation or shear. Even more striking, the B basis that made this work had two vectors that were perpendicular to each other, making the picture exceptionally clean. These provide a sneak peek at our next two topics, eigenvectors and orthogonality.

Homework

3.A Practice

These problems are recommended to practice key concepts from class. You can skip them if you feel comfortable already.

3.A.1 Find the matrix

Use the fact that the columns of a matrix correspond to the images of the standard basis vectors to solve the following problems.

1. What is the matrix that expresses counterclockwise rotation by θ in $\mathbb{R}^2$?
2. Think of a complex number $x + yi$ as a vector $(x, y) \in \mathbb{R}^2$. What is the matrix that expresses complex multiplication by $a + bi$? In other words, if

$$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} z \\ w \end{bmatrix},$$

   then $(x + yi)(a + bi) = (z + wi)$.
3. You saw that $[A]_B = B^{-1}AB$ is one way to compute $[A]_B$. The other way is to compute each column directly, as we have been doing with the above parts. Use this method with $[A]_B[\vec{x}]_B = [A\vec{x}]_B$ to compute

$$[A]_B = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}_B \qquad \text{where} \qquad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

3.A.2 Computing determinants

Compute the determinants of the following matrices.

1. $\begin{bmatrix} 2 & 1 & -1 \\ 4 & 3 & 1 \\ -2 & 1 & 5 \end{bmatrix}$
2. $\begin{bmatrix} 1 & 0 & 2 \\ 3 & 1 & -1 \\ -1 & -1 & 5 \end{bmatrix}$

3.A.3 Determinant of a 2×2 matrix

1. Find the general formula for the determinant of a 2×2 matrix.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

2. Using basic planar geometry rather than linear algebra, compute the area of the parallelogram given the *rows* of the matrix (i.e. (a, b) and (c, d)), showing that it is the same up to absolute value.
3. Using basic planar geometry rather than linear algebra, compute the area of the parallelogram given the *columns* of the matrix (i.e. (a, c) and (b, d)), showing that it is the same up to absolute value.

3.B Extensions

These problems will assist you in discovering new ideas, and are recommended. Problems that will be referenced in the future are noted with $(*)$.

3.B.1 More determinant properties $(*)$

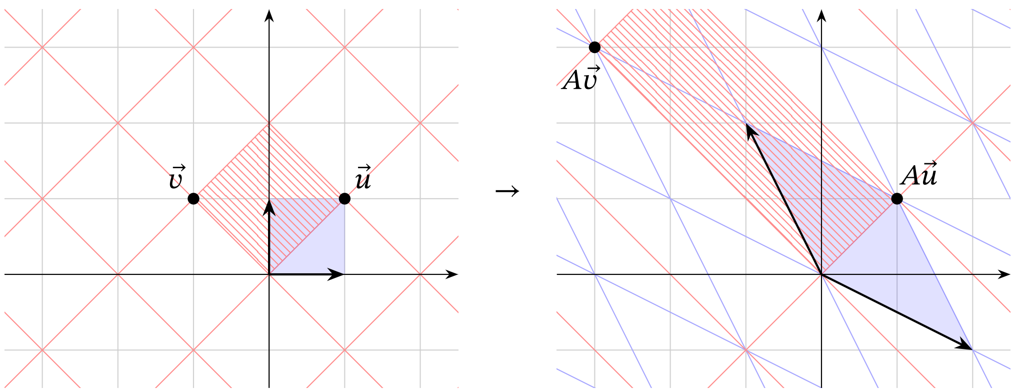
1. Show that $\det(A) \neq 0$ if and only if the conditions in Theorem 2.8 hold. (One-line proof from what we discussed.)
2. (Challenge, but at least read the statement) You saw in Problem 3.A.3 that the determinant of a 2×2 matrix is a degree-2 multivariate polynomial in its matrix entries. Show in general that the determinant of an $n \times n$ matrix is a degree-n multivariate polynomial in its matrix entries. (Hint: What happens during Gaussian elimination?)

4 Eigenvectors and Fibonacci numbers

In the previous section, we discussed how matrices can be expressed in different bases, and we noted specifically how it was nice to have a basis for a linear transformation that is formed only by "pure-scaling directions." In the example

$$[A]_B = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}_B = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

which is drawn again below, the pure-scaling directions for A are exactly $\vec{u} = (1, 1)$ and $\vec{v} = (-1, 1)$. We say that A is *diagonalizable* since $[A]_B$ is a *diagonal matrix*.



Definition 4.1. Given an $n \times n$ matrix A, a vector $\vec{x} \neq \vec{0}$ is an *eigenvector* with associated *eigenvalue* λ if $A\vec{x} = \lambda\vec{x}$. If there exist vectors $\vec{x}_1, \dots, \vec{x}_n$ that are all eigenvectors of A and form a basis, this is known as an *eigenbasis*. ⌟

Eigenbases don't always exist, but when they do, they are very useful. For our one application, consider the matrix of the Fibonacci sequence,

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

Starting at $\vec{x} = (1, 0)$, we are interested in the sequence $\vec{x}, A\vec{x}, A^2\vec{x}, A^3\vec{x}, \dots$. The first component of $A^n\vec{x}$ is the $(n+1)$th Fibonacci number.

We will derive a closed-form formula for the nth Fibonacci number. Here is the plan for doing so:

- First, compute the eigenvalues and eigenvectors. There will be two eigenvalues λ and μ, with corresponding eigenvectors $\vec{u}$ and $\vec{v}$, and they will form a basis.
- Express the starting vector $\vec{x}$ in this basis, as $a \cdot \vec{u} + b \cdot \vec{v}$.
- By linearity, we will get

$$A^n\vec{x} = A^n(a \cdot \vec{u} + b \cdot \vec{v}) = a \cdot A^n\vec{u} + b \cdot A^n\vec{v} = a\lambda^n\vec{u} + b\mu^n\vec{v}.$$

This is the closed form we're looking for, using an exponential instead of recursion!

To get started, we need to solve $A\vec{u} = \lambda\vec{u}$. Unfortunately, this appears a priori to be a non-linear equation, since both λ and $\vec{u}$ are unknowns, and they are multiplied together. However, we can use a neat trick to actually solve it!

Recall that matrices of the same size can be added and subtracted. Then because $\lambda\vec{u} = (\lambda I)\vec{u}$, we can rearrange to get $(\lambda I - A)\vec{u} = \vec{0}$. While this system of equations still has the problem of being non-linear, we know from the previous section's homework that this equation has non-trivial solutions if and only if $\det(\lambda I - A) = 0$! Furthermore, $\det(\lambda I - A)$ must be a polynomial in λ. Therefore, as we compute it, we can feel free to divide by things like λ and $\lambda - 1$, since there is a unique polynomial that fits the infinitely many remaining points.

We begin to compute:

$$\det\left(\begin{bmatrix} \lambda - 1 & -1 \\ -1 & \lambda \end{bmatrix}\right) = \det\left(\begin{bmatrix} \lambda - 1 & -1 \\ 0 & (\lambda(\lambda-1)-1)/(\lambda-1) \end{bmatrix}\right).$$

We get that $\det(\lambda I - A) = \lambda(\lambda - 1) - 1 = \lambda^2 - \lambda - 1$. By the quadratic formula, $\det(\lambda I - A) = 0$ if and only if $\lambda = (1 \pm \sqrt{5})/2$. These are the eigenvalues. To compute the corresponding eigenvectors, the last row in the matrix above is now 0, so the first row gives the following, where $\vec{u} = (u_1, u_2)$, $\vec{v} = (v_1, v_2)$, and u_2 and v_2 are free variables.

$$\left(\frac{1+\sqrt{5}}{2} - 1\right) \cdot u_1 - u_2 = 0$$
$$\left(\frac{1-\sqrt{5}}{2} - 1\right) \cdot v_1 - v_2 = 0$$

Pick arbitrary values for u_2 and v_2, say $u_2 = v_2 = 1$. Then we calculate

$$u_1 = \frac{1}{\frac{1+\sqrt{5}}{2} - 1} = \frac{1}{\frac{-1+\sqrt{5}}{2}} = \frac{2}{-1+\sqrt{5}} = \frac{2(-1-\sqrt{5})}{-4} = \frac{1+\sqrt{5}}{2} = \lambda.$$

The calculation for v is similar, and we end up getting the following eigenvectors.

$$\lambda = \frac{1+\sqrt{5}}{2} \quad \Leftrightarrow \quad \vec{u} = \begin{bmatrix} \lambda \\ 1 \end{bmatrix} \qquad\qquad \mu = \frac{1-\sqrt{5}}{2} \quad \Leftrightarrow \quad \vec{v} = \begin{bmatrix} \mu \\ 1 \end{bmatrix}$$

To express $\vec{x} = (1, 0)$ in this basis, one can check that

$$\vec{x} = \frac{1}{\sqrt{5}} \cdot \vec{u} - \frac{1}{\sqrt{5}} \cdot \vec{v}.$$

Finally, we conclude that the $(n+1)$th Fibonacci number is the first component of $A^n\vec{x}$,

$$F_{n+1} = \frac{1}{\sqrt{5}} \cdot \lambda^n \cdot \lambda - \frac{1}{\sqrt{5}} \cdot \mu^n \cdot \mu = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right).$$

Homework

4.A Practice

These problems are recommended to practice key concepts from class. You can skip them if you feel comfortable already.

4.A.1 Computing eigenvalues and eigenvectors

Compute the eigenvalues and eigenvectors of the following matrices.

1. $\begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$

2. $\begin{bmatrix} 2 & 1 & -1 \\ -1 & 4 & 1 \\ -2 & 2 & 3 \end{bmatrix}$

4.B Extensions

These problems will assist you in discovering new ideas, and are recommended. Problems that will be referenced in the future are noted with $(*)$.

4.B.1 Exploring cases without real eigenbases

1. Think of the entries of the 90° counterclockwise rotation matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

as numbers in $\mathbb{C}$, not $\mathbb{R}$. Compute the complex-valued eigenvalues and complex-valued eigenvectors.

2. Prove that every matrix has at least one eigenvalue when thought of over $\mathbb{C}$.

3. Show that although

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

has an eigenvalue, it does not have 2 distinct eigen*vectors* pointing in different directions. Thus this is a true failure case, where a basis of eigenvectors does not exist, even over $\mathbb{C}$.

5 A brief intro to orthogonality

The last big topic in linear algebra is *orthogonality*: the ability for vectors to point in perpendicular directions. Bases where all the vectors are orthogonal to each other are especially nice, and eigenbases where all the vectors are orthogonal to each other are arguably the nicest bases. To discuss orthogonality, we'll need to discuss a new operation.

Definition 5.1. The dot product of two vectors is defined

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + \cdots + x_n y_n.$$

People also sometimes write $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$, where $\vec{x}^T$ denotes the $1 \times n$ matrix that lays out the entries of $\vec{x}$ in a row. Vectors $\vec{x}$ and $\vec{y}$ are called *orthogonal* if $\vec{x} \cdot \vec{y} = 0$.

The norm of a vector is defined to agree with the standard notation of distance in $\mathbb{R}^n$, and can be computed using dot product by

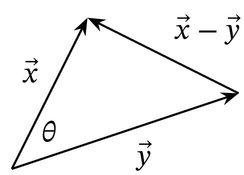
$$||\vec{x}|| = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Vectors $\vec{x}$ and $\vec{y}$ are called *orthonormal* if they are orthogonal and $||\vec{x}|| = ||\vec{y}|| = 1$. ⌟

It is not obvious why an expression like $x_1 y_1 + \cdots + x_n y_n$ should say anything about vectors being perpendicular. To make the connection clearer, here is another view of the dot product.

Proposition 5.2. *For all $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have $\vec{x} \cdot \vec{y} = ||\vec{x}|| ||\vec{y}|| \cos(\theta)$, where θ is the angle between $\vec{x}$ and $\vec{y}$.* ⌟

Proof. No matter what dimensions these vectors live in, $\vec{x}$ and $\vec{y}$ (assuming they are not parallel) span a plane, so consider the below picture in that plane.



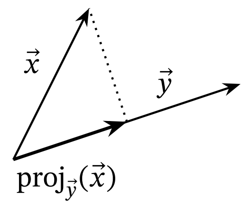
By law of cosines, we have that

$$\begin{aligned} ||\vec{x}||^2 + ||\vec{y}||^2 - 2||\vec{x}|| ||\vec{y}|| \cos(\theta) &= ||\vec{x} - \vec{y}||^2 \\ &= (\vec{x} - \vec{y}) \cdot (\vec{x} - \vec{y}) \\ &= \vec{x} \cdot \vec{x} - 2\vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{y} \\ &= ||\vec{x}||^2 + ||\vec{y}||^2 - 2\vec{x} \cdot \vec{y} \end{aligned}$$

We have used distributive and commutative properties that we did not prove, but are hopefully believable. Cancel to get the desired equation. □

Thus, dot product being 0 corresponds to $\theta = \pi/2$, or in other words the vectors are perpendicular. This is a useful fact to remember! Here are a few more points about orthogonality that make it useful and practical:

- You can compute the projection of one vector onto another



The formula is given by

$$\mathrm{proj}_{\vec{y}}(\vec{x}) = \frac{\vec{x} \cdot \vec{y}}{\vec{y} \cdot \vec{y}} \cdot \vec{y}.$$

- Given a basis of vectors $\vec{x}_1, \dots, \vec{x}_n$, you can always pick an orthonormal basis $\vec{y}_1, \dots, \vec{y}_n$ such that $\mathrm{span}(\vec{x}_1, \dots, \vec{x}_i) = \mathrm{span}(\vec{y}_1, \dots, \vec{y}_i)$ for all i. Just do so iteratively. Start with $\vec{y}_1 = \vec{x}_1 / \|\vec{x}_1\|$. Then $\mathrm{span}(\vec{x}_1, \vec{x}_2)$ is a plane, so choose $\vec{y}_2$ within that plane and perpendicular to $\vec{y}_1$. In particular, you can take

$$\vec{y}_2 = \frac{\vec{x}_2 - \mathrm{proj}_{\vec{y}_1}(\vec{x}_2)}{\|\vec{x}_2 - \mathrm{proj}_{\vec{y}_1}(\vec{x}_2)\|}.$$

Then iterate.

- Once you have an orthonormal basis $\vec{b}_1, \dots, \vec{b}_n$, it is super easy to use. In particular, recall that typically to express a vector as a linear combination of basis vectors, you need to either do Gaussian elimination or compute an inverse matrix. Now, the coefficient of $\vec{b}_i$ is just $\vec{x} \cdot \vec{b}_i$!

$$\vec{x} = (\vec{x} \cdot \vec{b}_1)\vec{b}_1 + \cdots + (\vec{x} \cdot \vec{b}_n)\vec{b}_n$$

To prove this, first write $\vec{x} = a_1 \vec{b}_1 + \cdots + a_n \vec{b}_n$. Then take the dot product of both sides with $\vec{b}_i$. We conclude that

$$\begin{aligned} \vec{x} \cdot \vec{b}_i &= (a_1 \vec{b}_1 + \cdots + a_n \vec{b}_n) \cdot \vec{b}_i \\ &= a_i, \end{aligned}$$

using the fact that $\vec{b}_i \cdot \vec{b}_j = 0$ if $i \neq j$ and 1 if $i = j$.

Theorem 5.3 (Real spectral theorem)**.** *Recall that A^T denotes the matrix obtained from A with the (i, j)th entry swapped with the (j, i)th entry. A matrix is called* symmetric *if $A = A^T$. Every symmetric matrix (of real numbers) has a basis of orthonormal eigenvectors.* ⌟

Symmetric matrices appear in a variety of applications. Here are two:

- The Hessian matrix from multivariable calculus, which contains $\frac{\partial^2 f}{\partial x_i \partial x_j}$ in the (i, j)th entry, is a symmetric matrix. This is a noteworthy result from a theorem in multivariable calculus, called the *equality of mixed partials*.
- In graph theory, one often represents a graph using a square matrix, with rows and columns representing vertices. Each (i, j)th entry is 1 or 0, depending on whether or not there is an edge from i to j. An *undirected* graph corresponds to a symmetric matrix.

The spectral theorem has key applications in both of these domains, but those are beyond the current scope.

6 Linear algebra in not just $\mathbb{R}^n$

6.1 Abstract vector spaces

So far, we have focused on considering $\mathbb{R}^n$. This is the space that we can visualize the best, and it offers the most intuition for all the theorems and properties we discussed. However, many applications of linear algebra rely on other spaces that look a little different, but still satisfy the same core properties of $\mathbb{R}^n$, allowing most of our results to transfer over.

In the below, a *field* is a set that allows addition, subtraction, multiplication, and division, with the standard associativity, commutativity, and distributivity properties. Examples of fields include $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, the integers modulo a prime p, the set $\mathbb{R}(x)$ of rational functions in x over $\mathbb{R}$, etc.

Definition 6.1. A *vector space* is a set V associated with a field of scalars F, with two operations defined, $+ : V^2 \to V$ between two vectors and $\cdot : F \times V \to V$ between a scalar and a vector. We say that V is a vector space *over* F. It is required to satisfy:

- (associativity of addition) $(u+v)+w = u+(v+w)$ for all $u, v, w \in V$.
- (commutativity of addition) $u+v = v+u$ for all $u, v \in V$.
- (existence of additive identity) There exists a special element denoted $\vec{0} \in V$ such that $v + \vec{0} = \vec{0} + v = v$ for all $v \in V$.
- (existence of additive inverse) For all $u \in V$, there exists $v \in V$ such that $u + v = v + u = \vec{0}$.
- (compatibility of 0) $0 \cdot v = \vec{0}$ for all $v \in V$, where 0 denotes the additive identity of F.
- (compatibility of 1) $1 \cdot v = v$ for all $v \in V$, where 1 denotes the multiplicative identity of F.
- (compatibility of field multiplication) $(cd) \cdot v = c \cdot (d \cdot v)$ for all $c, d \in F$ and $v \in V$.
- (distributivity over vector addition) $c \cdot (u+v) = c \cdot u + c \cdot v$ for all $c \in F$ and $u, v \in V$.
- (distributivity over field addition) $(c+d) \cdot v = c \cdot v + d \cdot v$ for all $c, d \in F$ and $v \in V$.

⌟

Usually, the definition of addition and scalar multiplication will be obvious (e.g. just do it componentwise), so people just describe the space V and the field F. Here are examples of vector spaces:

- Over any F, the set F^n for any n.
- Over any F, the set of all $m \times n$ matrices with elements in F, for any m and n.
- Over any F, the set of all polynomials with degree $\leq n$ and coefficients in F, for any n.
- Over $\mathbb{R}$, the set $\mathbb{C}$.
- Over $\mathbb{R}$, the set of all infinite sequences $a_1, a_2, \ldots$ satisfying the Fibonacci recurrence $a_n = a_{n-1} + a_{n-2}$.

Definition 6.2. An *isomorphism* of vector spaces A and B is a way to translate between their objects while preserving the operation structure. Formally, denote $+_A$, $+_B$, $\cdot_A$, and $\cdot_B$ the operations in the two spaces, and suppose A and B are both vector spaces over F. (Isomorphisms are not defined between vector spaces over different fields.) An isomorphism is a bijection $\varphi : A \to B$ such that $\varphi(\vec{x} +_A \vec{y}) = \varphi(\vec{x}) +_B \varphi(\vec{y})$ and $\varphi(c \cdot_A \vec{x}) = c \cdot_B \varphi(\vec{x})$ for all $\vec{x}, \vec{y} \in A$ and $c \in F$. ⌟

Theorem 6.3. *A vector space is called* finite-dimensional *if it has a finite basis (linear independence, span, and basis are defined similarly to before). All finite-dimensional vector spaces over F are isomorphic to F^n for some n.* ⌟

Here are some examples of such isomorphisms. Note that isomorphisms only have to preserve the structure of addition and multiplication by scalars. The sets can have much more structure (e.g. multiplication between vectors, etc.) but that structure is ignored for the purpose of discussing isomorphisms of vector spaces.

- The set of $m \times n$ matrices over F is isomorphic to F^{mn}.
- The set of all polynomials of degree $\leq n$ with coefficients in F is isomorphic to F^{n+1}.
- The vector space $\mathbb{C}$ over $\mathbb{R}$ is isomorphic to $\mathbb{R}^2$. A basis is $\{1, i\}$.
- The set of infinite sequences satisfying the Fibonacci recurrence is isomorphic to $\mathbb{R}^2$, since the sequences are fully determined by the first two entries. A basis is $\{(0, 1, 1, 2, 3, \dots), (1, 0, 1, 1, 2, \dots)\}$

Because of these isomorphisms, our results about $\mathbb{R}^n$ are fully general, at least with regards to topics like Gaussian elimination, linear independence, span, and eigenvalues and eigenvectors.

6.2 Inner product spaces

Orthogonality is the one topic that fails to generalize to some vector spaces. The most apparent failure is over finite fields, e.g. integers modulo a prime. While you can try to compute a dot product, take the following example over $\{0, 1\}^2$, which shows a non-zero vector being "orthogonal" to itself!

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot 1 + 1 \cdot 1 = 1 + 1 = 0$$

In order to satisfy all the properties that we want, inner product spaces only exist over $\mathbb{R}$ and $\mathbb{C}$. Inner products generalize the dot product, and can be defined for all finite-dimensional vector spaces over $\mathbb{R}$ and $\mathbb{C}$: we will see this shortly.

Definition 6.4. An *inner product space* is a vector space V over $F = \mathbb{R}$ or $\mathbb{C}$ with an extra operator called the inner product $\langle \cdot, \cdot \rangle : V^2 \to F$. It satisfies:

- (conjugate symmetry) $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$.
- (linearity in the first argument) $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle c \cdot u, v \rangle = c \cdot \langle u, v \rangle$ for all $u, v, w \in V$ and $c \in F$.
- (positive definiteness) $\langle u, u \rangle \geq 0$ for all $u \in V$, and $\langle u, u \rangle = 0$ if and only if $u = \vec{0}$ (the additive identity of V). ⌟

From here, we can define orthogonality, length, and other notions much like we did for $\mathbb{R}^n$. The standard inner product over $\mathbb{R}^n$ is the dot product, and the standard inner product over $\mathbb{C}^n$ is $\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^n x_i \overline{y_i}$. The construction for vector spaces over $\mathbb{C}$ deserves a few notes.

- First, there is conjugate linearity in the second argument: $\langle u, c \cdot v \rangle = \overline{\langle c \cdot v, u \rangle} = \overline{c \cdot \langle v, u \rangle} = \overline{c} \cdot \langle u, v \rangle$.
- The fuss about conjugation is necessary for positive definiteness, which is necessary to take square roots and define a length. If we had full linearity, then $\langle iu, iu \rangle = i^2 \langle u, u \rangle = -\langle u, u \rangle$. So at least one of $\langle iu, iu \rangle$ and $\langle u, u \rangle$ would have to be negative or 0.
- While there are many similarities between the theory for $\mathbb{C}$ and the theory for $\mathbb{R}$, the conjugation also creates notable differences. For example, over $\mathbb{R}^2$, the matrix A that rotates by 90° counterclockwise satisfies $A\vec{x} \cdot \vec{x} = 0$ for all $\vec{x} \in \mathbb{R}^2$. But with this inner product over $\mathbb{C}^2$, this matrix does not have this property. For example,

$$\left\langle \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix}, \begin{bmatrix} 1 \\ i \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} -i \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ i \end{bmatrix} \right\rangle = -2i \neq 0.$$

Proposition 6.5. *Let $\{e_1, \dots, e_n\}$ be any basis of a vector space V over $F = \mathbb{R}$ or $\mathbb{C}$. We can always define an inner product as follows: for every $u, v \in V$, first write $u = \sum_{i=1}^n a_i e_i$ and $v = \sum_{i=1}^n b_i e_i$. Then*

$$\langle u, v \rangle = \sum_{i=1}^n a_i \overline{b_i}$$

is a valid inner product. ⌟

6.3 Infinite dimensional spaces

Lastly, some vector spaces are infinite dimensional. Here are some examples.

- Over any F, the set of infinite sequences of elements of F, equivalently functions $f : \mathbb{N} \to F$.
- Over any F, the set $F[x]$ of all polynomials with no degree bound and coefficients in F.
- Over $\mathbb{R}$, the set of differentiable functions $f : \mathbb{R} \to \mathbb{R}$.
- Over $\mathbb{R}$, the set of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\int_{-\infty}^{\infty} f(x)^2 \, dx$ is finite.
- Over $\mathbb{Q}$, the set $\mathbb{R}$.

It is much harder to do linear algebra with these, and some of our nice properties no longer hold. For example, recall that any linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$ was injective if and only if it was also surjective. Now let V be the vector space of infinite sequences over $\mathbb{R}$. Then the function $f : V \to V$ given by

$$f(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$$

is only injective and not surjective.

Additionally, it is difficult to write down bases for infinite dimensional vector spaces. In fact, there are even two competing notions of "basis":

- If you don't have a notion of convergence in your space, you can only take finite linear combinations. A *Hamel basis* is an infinite set of vectors for which every finite subset is linearly independent and every vector can be written as a linear combination of a finite number of basis vectors.

  Some spaces have Hamel bases that can be written down explicitly. For example, because every polynomial has a finite degree, the set $\{1, x, x^2, x^3, \dots\}$ is a Hamel basis of the polynomial space $F[x]$. However, many spaces require the Axiom of Choice to get a Hamel basis, such as $\mathbb{R}$ over $\mathbb{Q}$ and the set of infinite sequences. Hamel bases always do exist, assuming Axiom of Choice.
- If you have a notion of convergence in your space, you can take infinite sums of vectors. In this sense, a *Schauder basis* more closely mimics the finite case, and is what Fourier series use. The Axiom of Choice is typically not needed to construct Schauder bases, though some pathological examples of spaces without Schauder bases do exist.

Infinite dimensional vector spaces can be inner product spaces, too. For example, when discussing functions $f : \mathbb{R} \to \mathbb{R}$, a common inner product is

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)\,dx,$$

though some care is needed to ensure this integral is finite and to identify functions that differ on negligible sets. This is the natural analogue of $\langle \vec{u}, \vec{v} \rangle = \sum_{i=1}^{n} u_i v_i$ in the infinite-dimensional world. When dealing with infinite sequences instead of functions, a convergent infinite sum replaces the integral. These are also used extensively in Fourier analysis.