Unit 2: Linear Algebra

Lectures 11–12 (§3.1)

Differential Equations & Linear Algebra

Course note: where we are going

From here onward:

- **Unit 2:** linear algebra (LA);
- **Unit 3:** differential equations *and* linear algebra together.

Linear algebra is very different from what we have done so far. It does *not* build directly on Calculus 1/2/3, and it is considerably more abstract.

> If you were dissatisfied with your Exam 1 score, let this be a wake-up call — do *not* expect Exam 2 to be easier. Step up your study habits instead.

Recipe for success

- **Skim the book before class** (especially the figures—tough to replicate on a chalkboard).
- **Take notes** — an additional source to study from, beyond the book.
- **Read the homework** early, and start thinking about solution methods.
- **ATTEND OFFICE HOURS** — it is a common mistake to think that attending office hours is only for bad students or students who are behind. Think of it as free tutoring! I would much rather students sit in someone's office hours and just do their homework and then ask the instructor/TA when they get stuck rather than doing it at home and copy solutions from ChatGPT. You can also come to just ask quick clarifying questions if you want.
- **Do practice exams without solutions in front of you** — doing problems while looking at solutions tricks you into thinking you understand things better than you actually might.

1 Lecture 11: Vectors, linear operators, and matrices

1.1 Where we have been, and what comes next

Previously we studied *single* differential equations: mainly the first-order linear equation

$$y' + p(t)\,y = f(t),$$

but we also discussed the general linear equation of order n,

$$a_n(t)\frac{d^n y}{dt^n} + \cdots + a_1(t)\frac{dy}{dt} + a_0(t)y = f(t) \qquad \Longleftrightarrow \qquad \mathrm{L}[y] = f,$$

written compactly in terms of the linear operator L:

$$L = a_n(t)\frac{d^n}{dt^n} + \cdots + a_1(t)\frac{d}{dt} + a_0(t)$$

Last time we introduced *systems* of differential equations, e.g.

$$\frac{dy_1}{dt} = P(y_1, y_2), \qquad \frac{dy_2}{dt} = Q(y_1, y_2).$$

The goal for this entire class is to solve coupled systems of the form

$$\begin{aligned} \mathrm{L}_1[y_1, y_2, \ldots] &= f_1 \\ \mathrm{L}_2[y_1, y_2, \ldots] &= f_2 \\ \mathrm{L}_3[y_1, y_2, \ldots] &= f_3 \\ &\vdots \end{aligned}$$

Such systems are crucial in:

- control systems / control theory;
- quantum mechanics;
- fluids, including aerospace;
- stability of structures (e.g. studying response to seismic activity) and things like vehicle suspensions;
- complex electrical circuits and thermal networks;
- biomedical stuff too (imaging via waves, drug absorption/interactions, etc)

But first we need the language for solving linear systems of algebraic equations — that language is linear algebra. We will first talk about how matrices and vectors work, and then we will abstractify everything and suddenly "vectors" are actually functions, and matrices are linear operators (like $d/dt + p(t)$). The analogy is going to be essentially

$$\underbrace{\text{vectors} : \text{matrices}}_{\text{linear algebraic systems}} \quad :: \quad \underbrace{\text{functions} : \text{operators L}}_{\text{linear systems of DEs}}$$

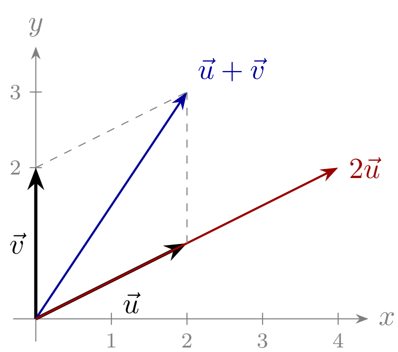
1.2 Scalars and vectors

Scalars (numbers) convey magnitude and sign, like +3 vs −4; we write $c \in \mathbb{R}$ to mean "c is an element of $\mathbb{R}$" or "c is a real number".

Vectors are (for now) just arrows—they convey magnitude and direction. For example,

$$\vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad \vec{v} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \qquad \vec{u}, \vec{v} \in \mathbb{R}^2.$$

We write "$\vec{u} \in \mathbb{R}^2$" to say "$\vec{u}$ is a two-dimensional vector whose entries are real numbers". By two-dimensional vector, I mean (graphically) that it's an arrow that can be drawn on the two-dimensional x-y plane (as opposed to an arrow that you would draw on a 1D number line, or in a three-dimensional visualization on a computer or something), or (numerically) that it can be represented as an array with two entries. The vector $\vec{u}$ above has two entries, which we might write as $u_1 = 2$ and $u_2 = 1$ (so, in general, u_j represents the jth entry of the vector $\vec{u}$).



Any two vectors of the same length can be added together, and they can also be "dotted" together; one can also multiply vectors by scalars.

Adding vectors ⇒ add the entries:

$$\vec{u} + \vec{v} = \begin{bmatrix} 2+0 \\ 1+2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Multiplication by a scalar ⇒ multiply the entries:

$$2\vec{u} = \begin{bmatrix} 2 \cdot 2 \\ 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

The dot product. For $\vec{u}, \vec{v} \in \mathbb{R}^2$,

$$\vec{u} \cdot \vec{v} = |\vec{u}|\,|\vec{v}| \cos\theta = u_1 v_1 + u_2 v_2,$$

(where θ is the angle between the two vectors) or, more generally, for $\vec{u}, \vec{v} \in \mathbb{R}^n$,

$$\vec{u} \cdot \vec{v} = |\vec{u}|\,|\vec{v}| \cos\theta = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Note that the dot product of a vector with itself is how we measure its length, since (since $\theta = 0 \Rightarrow \cos\theta = 1$ for the angle between a vector and itself)

$$\vec{u} \cdot \vec{u} = |\vec{u}|^2 = u_1 u_1 + u_2 u_2 = u_1^2 + u_2^2.$$

Note that the dot product of two vectors is a *scalar*.

1.3 Linear operators and matrices

Recall: a *linear operator* is any function that takes in vectors and outputs vectors, which also satisfies

$$\mathrm{L}[\vec{u}+\vec{v}] = \mathrm{L}[\vec{u}] + \mathrm{L}[\vec{v}], \quad \text{and} \quad \mathrm{L}[c\,\vec{u}] = c\ \mathrm{L}[\vec{u}].$$

The reason we're doing all this linear algebra is because of the following fundamentally important & useful fact about linear operators:

> Any linear operator L that maps vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$ (which we write as $\mathrm{L} : \mathbb{R}^n \to \mathbb{R}^m$) can be represented as an $m \times n$ **matrix**
>
> $$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1n} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2n} \\ \vdots & & \ddots & \vdots \\ \ell_{m1} & \ell_{m2} & \cdots & \ell_{mn} \end{bmatrix} = [\ell_{ij}].$$

In the box above, $[\ell_{ij}]$ is just a quick shorthand way to write down "**L** is a matrix, and the entry in the ith row and jth column is ℓ_{ij}".

If $\vec{u} \in \mathbb{R}^n$, i.e.,

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix},$$

and $\vec{v}$ is the vector in $\mathbb{R}^m$ that you get when plugging $\vec{u}$ into **L** (like plugging some number x into some function $f(x)$ back in algebra class), then we write $\vec{v} = \mathbf{L}\vec{u}$. If we know all the entries of **L** and of $\vec{u}$, then we can calculate the entries of $\vec{v}$ by doing matrix multiplication, as follows:

$$\mathbf{L}\vec{u} = \begin{bmatrix} \ell_{11}u_1 + \ell_{12}u_2 + \cdots + \ell_{1n}u_n \\ \ell_{21}u_1 + \ell_{22}u_2 + \cdots + \ell_{2n}u_n \\ \vdots \\ \ell_{m1}u_1 + \ell_{m2}u_2 + \cdots + \ell_{mn}u_n \end{bmatrix}.$$

So the jth entry of the vector $\vec{v}$ is the result of taking the dot product between the jth row of **L** and the vector $\vec{u}$.

Example 1.1 (90° counterclockwise rotation)**.** One can show that the act of rotating a vector counterclockwise by 90° satisfies all the defining properties of a linear operator: it takes vectors as input, spits out vectors as output, and you can show it satisfies linearity. Call the matrix corresponding to such rotations **R**. Here's what happens when you plug in a couple simple vectors.

$$\mathbf{R}\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix}, \qquad \mathbf{R}\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}-1\\0\end{bmatrix}.$$

(if you draw the input and output vectors from these two examples, you can see that they are at right angles to each other; or you can use inner products to show that they are orthogonal).

It turns out that the entries of this matrix are specifically

$$\mathbf{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \qquad \Longrightarrow \qquad \mathbf{R}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -u_2 \\ u_1 \end{bmatrix}.$$

Let's see what happens when we rotate a vector twice. Applying $\mathbf{R}$ twice to a hypothetical $\vec{u}$ gives

$$\mathbf{R}\mathbf{R}\vec{u} = \mathbf{R}\big[\mathbf{R}\vec{u}\big] = \mathbf{R}\begin{bmatrix} -u_2 \\ u_1 \end{bmatrix} = \begin{bmatrix} -u_1 \\ -u_2 \end{bmatrix}.$$

We've said that rotation by 90 degrees is a linear operator, but rotation by any amount is also (there's nothing special about 90). So rotation by 180 degrees is also a linear operator. So It must have a matrix representation also. This motivates the idea of matrix multiplication, which we'll save for next lecture.

2 Lecture 12 (§3.1–3.2): Matrix multiplication

2.1 Recap

- An $m \times n$ matrix (m rows, n columns) represents a linear operator mapping an input $\vec{u} \in \mathbb{R}^n$ to an output $\vec{v} \in \mathbb{R}^m$. We write $\mathbf{A}\vec{u} = \vec{v}$ to mean "$\mathbf{A}$ acting on $\vec{u}$ returns $\vec{v}$" (just like $y = f(x)$ is like saying "f acting on x returns y"; matrix multiplication is just functional evaluation, but for linear functions of vectors)
- Given

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \\ a_{m1} & & a_{mn} \end{bmatrix}, \qquad \vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix},$$

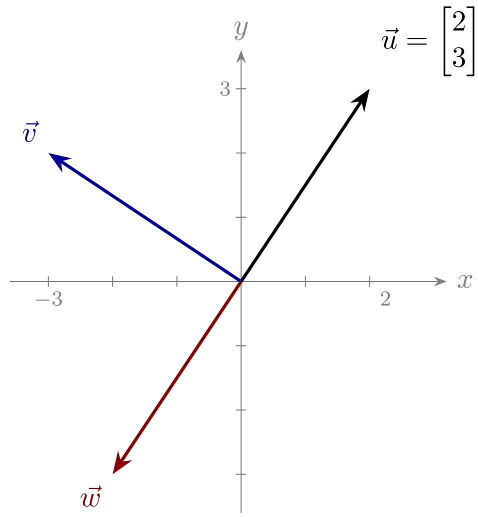
we can compute $\vec{v}$ by "dotting" $\vec{u}$ into rows 1 through m of $\mathbf{A}$:

$$v_1 = \vec{u} \cdot \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix}, \qquad v_2 = \vec{u} \cdot \begin{bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2n} \end{bmatrix}, \qquad \cdots$$

2.2 Worked example: rotation, revisited

Take the 90° counterclockwise rotation

$$\mathbf{R} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad \vec{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$



$$\vec{v} = \mathbf{R}\vec{u} = \begin{bmatrix} 2 \cdot 0 + 3 \cdot (-1) \\ 2 \cdot 1 + 3 \cdot 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}.$$

Check the angle:

$$\underbrace{\vec{u} \cdot \vec{v}}_{= |\vec{u}||\vec{v}| \cos\theta} = 2 \cdot (-3) + 3 \cdot 2 = 0 \qquad \Longrightarrow \qquad \theta = 90^\circ.\checkmark$$

What is $\mathbf{R}\vec{v}$? A 90° rotation of $\vec{v}$ is a 180° rotation of $\vec{u}$; call it $\vec{w}$:

$$\vec{w} = \mathbf{R}\vec{v} = \begin{bmatrix} -3 \cdot 0 + 2 \cdot (-1) \\ -3 \cdot 1 + 2 \cdot 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \end{bmatrix}.$$

Notice that

$$\vec{w} = \mathbf{R}\vec{v} = \underbrace{\mathbf{R}\Big(\underbrace{\mathbf{R}\vec{u}}_{\text{1st mult.}} \Big)}_{\text{2nd mult.}}.$$

Rotation by 180° is *also* a linear operator, so it has its own matrix. To find it, we use:

Composition of operators $\Longleftrightarrow$ matrix multiplication.

$$\mathbf{C}(\vec{u}) = \mathbf{A}\big(\mathbf{B}(\vec{u})\big) \qquad \Longleftrightarrow \qquad \mathbf{C}\vec{u} = \mathbf{A}\mathbf{B}\,\vec{u}.$$

2.3 Defining matrix multiplication

Let $\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$ with $\mathbf{C}\vec{u} = \mathbf{A}\mathbf{B}\vec{u}$, where $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$. Then we calculate the entries of $\mathbf{C}$ by doing dot products between rows of $\mathbf{A}$ and columns of $\mathbf{B}$:

$$\begin{aligned}
c_{11} &= \{\text{1st } \textit{row} \text{ of } \mathbf{A}\} \cdot \{\text{1st } \textit{column} \text{ of } \mathbf{B}\}, \\
c_{12} &= \{\text{1st row of } \mathbf{A}\} \cdot \{\text{2nd column of } \mathbf{B}\}, \\
c_{21} &= \{\text{2nd row of } \mathbf{A}\} \cdot \{\text{1st column of } \mathbf{B}\}, \\
c_{22} &= \{\text{2nd row of } \mathbf{A}\} \cdot \{\text{2nd column of } \mathbf{B}\}.
\end{aligned}$$

$$\begin{bmatrix} \boxed{c_{11}} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \overset{\text{row of } \mathbf{A}}{\begin{bmatrix} \boxed{a_{11} \quad a_{12}} \\ a_{21} \quad\; a_{22} \end{bmatrix}} \underset{\text{col. of } \mathbf{B}}{\begin{bmatrix} \boxed{\begin{matrix} b_{11} \\ b_{21} \end{matrix}} & \begin{matrix} b_{12} \\ b_{22} \end{matrix} \end{bmatrix}}$$

Back to the rotations:

$$\mathbf{R}_{180^\circ} = \mathbf{R}_{90^\circ}\mathbf{R}_{90^\circ} = \mathbf{R}_{90^\circ}^2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \cdot 0 - 1 \cdot 1 & 0 \cdot (-1) - 1 \cdot 0 \\ 1 \cdot 0 + 0 \cdot 1 & -1 \cdot 1 + 0 \cdot 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Now we have the matrix representation of the 180 degree rotation. So if we want to rotate any vector by 180 degrees, we can just multiply it by $\mathbf{R}_{180^\circ}$ once, whereas before we had to multiply it by $\mathbf{R}_{90^\circ}$ twice.

2.4 Dimensions of a product

Generally, if $\mathbf{A}$ is $m \times n$, so $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$, and $\mathbf{B}$ is $n \times k$, so $\mathbf{B} : \mathbb{R}^k \to \mathbb{R}^n$, then

$$\mathbf{C} = \mathbf{AB} : \mathbb{R}^k \to \mathbb{R}^m \quad \text{is } m \times k.$$

$$\begin{tikzcd}
\mathbb{R}^k \arrow[d, "\mathbf{B}"] \arrow[dr, bend left, "\mathbf{C} = \mathbf{AB}"] & \\
\mathbb{R}^n \arrow[r, "\mathbf{A}"'] & \mathbb{R}^m
\end{tikzcd}$$

Example 2.1. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ -2 & 4 \\ 3 & 0 \end{bmatrix}.$$

Here $\mathbf{B} : \mathbb{R}^2 \to \mathbb{R}^3$ and $\mathbf{A} : \mathbb{R}^3 \to \mathbb{R}^2$, so $\mathbf{C} = \mathbf{AB}$ must go from/to $\mathbb{R}^2 \to \mathbb{R}^2$. Therefore, $\mathbf{C}$ is 2×2:

$$\mathbf{C} = \begin{bmatrix} 1 \cdot 0 + 0 \cdot (-2) + 2 \cdot 3 & 1 \cdot 1 + 0 \cdot 4 + 2 \cdot 0 \\ 0 \cdot 0 + 1 \cdot (-2) + 2 \cdot 3 & 0 \cdot 1 + 1 \cdot 4 + 2 \cdot 0 \end{bmatrix} = \begin{bmatrix} 6 & 1 \\ 4 & 4 \end{bmatrix}.$$

Exercise for you to try: pick any $\vec{u} \in \mathbb{R}^2$. Calculate $\vec{v} = \mathbf{B}\vec{u}$ and then $\vec{w} = \mathbf{A}\vec{v}$. Then verify that $\mathbf{C}\vec{u}$ gives the same $\vec{w}$.

Remark. With $\mathbf{A}$ and $\mathbf{B}$ as above, $\mathbf{C} = \mathbf{AB} : \mathbb{R}^2 \to \mathbb{R}^2$ is 2×2, but $\mathbf{D} = \mathbf{BA} : \mathbb{R}^3 \to \mathbb{R}^3$ is 3×3. Hence, in general (but there are some exceptions!)

$$\boxed{\mathbf{AB} \neq \mathbf{BA}}$$

the two products need not even have the same shape.

3 Important matrices, rules, and properties

3.1 Special matrices and basic rules

(Recall that "$\forall$" is shorthand for "for all" or "for every")

- **The zero matrix.** For $\mathbf{A} = [a_{ij}]$, if every $a_{ij} = 0$, then $\mathbf{A}\vec{u} = \vec{0}$ for all $\vec{u}$.

- **The identity matrix.** The $n \times n$ matrix

$$\mathbb{I}_n \equiv \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{satisfies} \quad \mathbb{I}_n \vec{u} = \vec{u} \quad \forall\, \vec{u}.$$

- **Addition and scalar multiplication.** If $\mathbf{A} = [a_{ij}]$, $\mathbf{B} = [b_{ij}]$, and $c \in \mathbb{R}$, then

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}], \quad \text{and} \quad c\mathbf{A} = [c\, a_{ij}].$$

  This only works if $\mathbf{A}$ and $\mathbf{B}$ are the *same size*. (Note §3.1 of the book lists way more properties, so you should read the book; technically a lot of those properties follow from what I've already written.)

- **Compatibility of products.** $\mathbf{C} = \mathbf{AB}$ only makes sense if $\mathbf{B}$'s output has the same length as $\mathbf{A}$'s input: for instance, if $\mathbf{B}$ is $2 \times n$, then $\mathbf{A}$ must be $m \times 2$.

- **Non-commutativity.** $\mathbf{AB} \neq \mathbf{BA}$, except in special cases. When $\mathbf{AB} = \mathbf{BA}$ we say that "$\mathbf{A}$ **and** $\mathbf{B}$ **commute**" or "$\mathbf{A}$ **commutes** with $\mathbf{B}$".

3.2 Trace, transpose, and diagonal matrices

- **Trace.** For a *square* matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$,

$$\operatorname{Tr} \mathbf{A} \equiv a_{11} + a_{22} + \cdots + a_{nn}$$

  is the **trace** of $\mathbf{A}$.

- **Transpose.** For an $m \times n$ matrix $\mathbf{A} = [a_{ij}]$, the "flip"

$$\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & \ddots & & \vdots \\ \vdots & & & \\ a_{1n} & & \cdots & a_{mn} \end{bmatrix} \equiv \mathbf{A}^\mathsf{T}$$

(which is now $n \times m$ rather than $m \times n$!) is the *transpose* of $\mathbf{A}$. Properties:

$$(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}, \qquad (\mathbf{A}+\mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} + \mathbf{B}^\mathsf{T}, \qquad (\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}.$$

Note the order: $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$, *not* $\mathbf{A}^\mathsf{T}\mathbf{B}^\mathsf{T}$.

- **Diagonal matrices.** If

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

we call $\mathbf{A}$ *diagonal*. Then

$$\mathbf{A}\vec{u} = \begin{bmatrix} a_{11}u_1 \\ a_{22}u_2 \\ \vdots \\ a_{nn}u_n \end{bmatrix},$$

i.e. a diagonal matrix, when multiplying a vector, simply rescales each component of that vector.