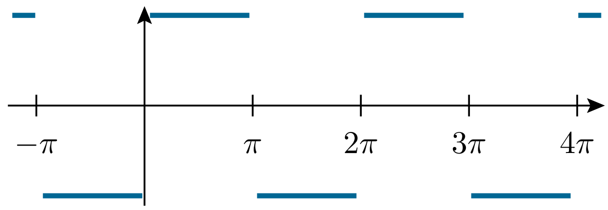
Example 194. Find the Fourier series of the 2π-periodic function $f(t)$ defined by

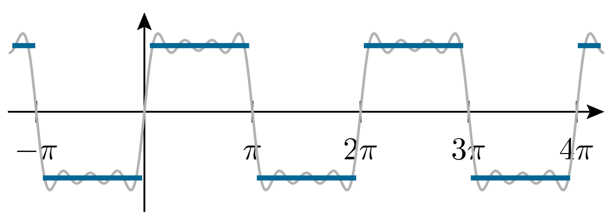
$$f(t) = \begin{cases} -1, & \text{for } t \in (-\pi, 0), \\ +1, & \text{for } t \in (0, \pi), \\ 0, & \text{for } t = -\pi, 0, \pi. \end{cases}$$



Solution. We compute $a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\mathrm{d}t = 0$, as well as

$$\begin{aligned} a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(nt)\mathrm{d}t = \frac{1}{\pi}\left[-\int_{-\pi}^{0}\cos(nt)\mathrm{d}t + \int_{0}^{\pi}\cos(nt)\mathrm{d}t\right] = 0 \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(nt)\mathrm{d}t = \frac{1}{\pi}\left[-\int_{-\pi}^{0}\sin(nt)\mathrm{d}t + \int_{0}^{\pi}\sin(nt)\mathrm{d}t\right] = \frac{2}{\pi n}[1 - \cos(n\pi)] \\ &= \frac{2}{\pi n}[1 - (-1)^n] = \begin{cases} \frac{4}{\pi n} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}. \end{aligned}$$

In conclusion, $f(t) = \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \frac{4}{\pi n}\sin(nt) = \frac{4}{\pi}\left(\sin(t) + \frac{1}{3}\sin(3t) + \frac{1}{5}\sin(5t) + \ldots\right)$.



Observation. The coefficients a_n are zero for all n if and only if $f(t)$ is odd.

Comment. The value of $f(t)$ for $t = -\pi, 0, \pi$ is irrelevant to the computation of the Fourier series. They are chosen so that $f(t)$ is equal to the Fourier series for all t (recall that, at a jump discontinuity t, the Fourier series converges to the average $\frac{f(t^-) + f(t^+)}{2}$).

Comment. Plot the (sum of the) first few terms of the Fourier series. What do you observe? The "overshooting" is known as the **Gibbs phenomenon**: `https://en.wikipedia.org/wiki/Gibbs_phenomenon`

Comment. Set $t = \frac{\pi}{2}$ in the Fourier series we just computed, to get Leibniz' series $\pi = 4[1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots]$. For such an alternating series, the error made by stopping at the term $1/n$ is on the order of $1/n$. To compute the 768 digits of π to get to the Feynman point ($3.14159265\ldots721134999999\ldots$), we would (roughly) need $1/n < 10^{-768}$, or $n > 10^{768}$. That's a lot of terms! (Roger Penrose, for instance, estimates that there are about 10^{80} atoms in the observable universe.)

Remark. Convergence of such series is not completely obvious! Recall, for instance, that the (odd part of) the harmonic series $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots$ diverges. (On the other hand, do you remember the alternating sign test from Calculus II?)

Throughout, V and W are vector spaces.

Just like we went from column vectors to abstract vectors (such as polynomials), the concept of a matrix leads to abstract linear transformations.

In the other direction, picking a basis, abstract vectors can be represented as column vectors (see Lecture 35). Correspondingly, linear transformations can then be represented as matrices.

Definition 195. A map $T: V \to W$ is a **linear transformation** if

$$T(c\boldsymbol{x} + d\boldsymbol{y}) = cT(\boldsymbol{x}) + dT(\boldsymbol{y}) \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \text{ in } V \text{ and all } c, d \text{ in } \mathbb{R}.$$

In other words, a linear transformation respects addition and scaling:

- $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$
- $T(c\boldsymbol{x}) = cT(\boldsymbol{x})$

It necessarily sends the zero vector in V to the zero vector in W:

- $T(\mathbf{0}) = \mathbf{0}$ [because $T(\mathbf{0}) = T(0 \cdot \mathbf{0}) = 0 \cdot T(\mathbf{0}) = \mathbf{0}$]

Comment. Linear transformations are special functions and, hence, can be composed. For instance, if $T: V \to W$ and $S: U \to V$ are linear transformations, then $T \circ S$ is a linear transformation $U \to W$ (sending $\boldsymbol{x}$ to $T(S(\boldsymbol{x}))$). If S, T are represented by matrices A, B, then $T \circ S$ is represented by the matrix BA. In other words, matrix multiplication arises as the composition of (linear) functions.

Example 196. The **derivative** you know from Calculus I is linear.

Indeed, the map $D: \left\{ \begin{array}{c} \text{space of all} \\ \text{differentiable} \\ \text{functions} \end{array} \right\} \to \left\{ \begin{array}{c} \text{space of all} \\ \text{functions} \end{array} \right\}$ defined by $f(x) \mapsto f'(x)$ is a linear transformation:

- $\underbrace{D(f(x) + g(x))}_{(f(x)+g(x))'} = \underbrace{D(f(x))}_{f'(x)} + \underbrace{D(g(x))}_{g'(x)}$
- $\underbrace{D(cf(x))}_{(cf(x))'} = \underbrace{cD(f(x))}_{cf'(x)}$

These are among the first properties you learned about the derivative.

Similarly, the **integral** you love from Calculus II is linear:

$$\int_a^b (f(x) + g(x))\mathrm{d}x = \int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x, \qquad \int_a^b cf(x)\mathrm{d}x = c\int_a^b f(x)\mathrm{d}x$$

In this form, we are looking at a map $T: \left\{ \begin{array}{c} \text{space of all} \\ \text{continuous} \\ \text{functions} \end{array} \right\} \to \mathbb{R}$ defined by $T(f(x)) = \int_a^b f(x)\mathrm{d}x$.

Example 197. Consider the space V of all polynomials $p(x)$ of degree 3 or less. The map $D: V \to V$ given by $p(x) \mapsto p'(x)$ is a linear. Write down the matrix M for this linear map with respect to the basis $1, x, x^2, x^3$.

Solution. $M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

For instance, the 3rd column says that x^2 (the 3rd basis element) gets sent to $0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 = 2x$.

Example 198. Consider the map

$$D\colon \begin{Bmatrix}\text{space of poly's} \\ \text{of degree } \leqslant 3\end{Bmatrix} \rightarrow \begin{Bmatrix}\text{space of poly's} \\ \text{of degree } \leqslant 2\end{Bmatrix}, \quad p(x) \mapsto p'(x).$$

Write down the matrix M for this linear map with respect to the bases $1, x, x^2, x^3$ and $1, x, x^2$.

Solution. $M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

For instance, the 3rd column says that x^2 (the 3rd basis element) gets sent to $0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 = 2x$.

Example 199. What is the pseudo-inverse of the matrix M from the previous example? Interpret your finding.

Solution. (final answer only) The pseudo-inverse of $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ is $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$.

The corresponding linear map sends 1 to x, x to $\frac{1}{2}x^2$ and x^2 to $\frac{1}{3}x^3$. That is, the pseudo-inverse computes the antiderivative of each monomial.

Comment. This is not surprising, since we are familiar from Calculus with the concepts of derivatives and antiderivatives (or integrals), and that these are "pseudo" inverse to each other.

Comment. Similarly, the pseudo-inverse of $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}$.

Now, the corresponding linear map sends 1 to x, x to $\frac{1}{2}x^2$, x^2 to $\frac{1}{3}x^3$, and x^3 to 0. That is, the pseudo-inverse computes the antiderivative of each monomial, with the exception of x^3 which gets send to 0 (its antiderivative does not live in the space of polynomials of degree 3).

Example 200. (The April Fools' Day "proof" that $\pi = 4$, cont'd)

In that "proof", we are constructing curves c_n with the property that $c_n \rightarrow c$ where c is the circle. This convergence can be understood, for instance, in the same sense $\|c_n - c\| \rightarrow 0$ with the norm introduced as we did for functions.

Since $c_n \rightarrow c$ we then wanted to conclude that $\text{perimeter}(c_n) \rightarrow \text{perimeter}(c)$, leading to $4 \rightarrow \pi$.

However, in order to conclude from $x_n \rightarrow x$ that $f(x_n) \rightarrow f(x)$ we need that f is continuous (at x)!!

The "function" perimeter, however, is not continuous. In words, this means that (as we see in this example) curves can be arbitrarily close, yet have very different arc length.

We can dig a little deeper: as you learned in Calculus II, the arc length of a function $y = f(x)$ for $x \in [a, b]$ is

$$\int_a^b \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} = \int_a^b \sqrt{1 + f'(x)^2}\mathrm{d}x.$$

Observe that this involves f'. Try to see why the operator D that sends f to f' is not continuous with respect to the distance induced by the norm

$$\|f\| = \left(\int_a^b f(x)^2 \mathrm{d}x\right)^{1/2}.$$

In words, two functions f and g can be arbitrarily close, yet have very different derivatives f' and g'.

That's a huge issue in **functional analysis**, which is the generalization of linear algebra to infinite dimensional spaces (like the space of all differentiable functions). The linear operators ("matrices") on these spaces frequently fail to be continuous.

How little we actually know!

Q: *How fast can we solve N linear equations in N unknowns?*

Estimated cost of Gaussian elimination:

$$\begin{bmatrix} \blacksquare & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & & & \vdots \\ 0 & * & * & \cdots & * \end{bmatrix}$$

- to create the zeros below the first pivot:
  $\Longrightarrow$ on the order of N^2 operations
- if there are N pivots total:
  $\Longrightarrow$ on the order of $N \cdot N^2 = N^3$ operations

- A more careful count places the cost at $\sim\frac{1}{3}N^3$ operations.
- For large N, it is only the N^3 that matters.

  It says that if $N \to 10N$ then we have to work 1000 times as hard.

That's not optimal! We can do better than Gaussian elimination:

- Strassen algorithm (1969): $N^{\log_2 7} = N^{2.807}$
- Coppersmith–Winograd algorithm (1990): $N^{2.375}$
- ... Stothers–Williams–Le Gall (2014): $N^{2.373}$ (If $N \to 10N$ then we have to work 229 times as hard.)

Is $N^{2+(\text{a tiny bit})}$ possible? **We don't know!** (People increasingly suspect so.) (Better than N^2 is impossible; why?)

Comment. The above algorithms actually are for computing matrix products. It can be shown that, if $M(N)$ is the cost for multiplying two $N \times N$ matrices, then $N \times N$ systems can also be solved for cost on the order of $M(N)$. In other words, we don't even know how costly it is to multiply two matrices.

Good news for applications:

- Matrices typically have lots of structure and zeros

  which makes solving so much faster.