COUNTING, SUMS AND SERIES

MATT KERR

– to Bob and Sandy, my first teachers; and to the many students who inspire me –

Abstract. Based on lectures given to high-school students at the Alberta Summer Math Institute, this is a swift but self-contained tour of power sums, zigzag numbers, updown mountains, discrete Fourier transforms, and special values of L-functions.

1. Power sums and Bernoulli numbers

1.1. **Finite sums.** Perhaps anticipating a long quiet morning to read Goethe, the 18th century German schoolmaster charged his pupils with the task of calculating the lengthy sum $1 + 2 + 3 + \cdots + 100$. To his amazement (and, one imagines, vast chagrin), in a matter of seconds an 8-year old student named Carl Friedrich Gauss walked up with $5,050$ written on his tablet. Keeping in mind that in 1785, tablets were made out of slate, the teacher could not plausibly accuse the kid of Googling the answer. (The lesser-known part of the story is that when they went out for lunch, little Carl hit on his idea of curvature while trying to hold a large slice of pizza.)

So how did Gauss do it? He simply made 50 pairs

$$\underbrace{1 + \overbrace{2 + \underbrace{3 + \cdots + \overbrace{50 + 51} + \cdots + 98} + 99} + 100}$$

out of the terms of the sum, each of which summed to 101, resulting in a grand total of $50 \times 101 = 5050$. We can rewrite this more compactly in sigma notation:[1]

$$\sum_{i=1}^{100} i = 5050.$$

[1]For a sequence $\{a_k\}$, we will need the basic properties:

- $c\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} ca_k$;
- $\sum_{k=1}^{n} 1 = n$; and
- $\sum_{k=1}^{n}(a_k \pm b_k) = \sum_{k=1}^{n} a_k \pm \sum_{k=1}^{n} b_k$.

More generally, suppose we would like to find

$$S(n) := \sum_{i=1}^{n} i.$$

To proceed, observe (bear with me now) that

$$\begin{aligned}
(i+1)^2 - i^2 &= 2i+1 \\
\Longrightarrow \sum_{i=1}^{n}\{(i+1)^2 - i^2\} &= \sum_{i=1}^{n}(2i+1) \\
\Longrightarrow \sum_{i=1}^{n}(i+1)^2 - \sum_{i=1}^{n} i^2 &= 2\sum_{i=1}^{n} i + \sum_{i=1}^{n} 1 \\
\Longrightarrow (n+1)^2 - 1^2 &= 2S(n) + n \\
\Longrightarrow n^2 + \cancel{2}n + \cancel{1-1} &= 2S(n) + \cancel{n} \\
\Longrightarrow S(n) = \frac{n^2+n}{2} &= \frac{n(n+1)}{2}.
\end{aligned}$$

The final answer is (in retrospect) exactly what you would get from Gauss's trick, and this argument seems more complicated and less conceptual (a common complaint about algebra). So why did I use this approach? Because it *generalizes* well:

Exercise 1.1. Find formulas for $\sum_{i=1}^{n} i^2$ and $\sum_{i=1}^{n} i^3$.

How does this sequence continue? That is, what is the *general formula* for the k^{th} **power sum** $\sum_{i=1}^{n} i^k$? The surprisingly deep answer will take us back before Gauss (1777-1855), to the time of Leonhard Euler (1707-1783), Jakob Bernoulli (1655-1705) and Johann Faulhaber (1580-1635).

1.2. **Binomial numbers.**

Definition 1.2. The binomial number

$$\binom{n}{k} := \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1} = \frac{n!}{(n-k)!k!} = \binom{n}{n-k}$$

counts how many ways you can choose k objects from a set of n.

The numerator of the fraction has k factors, reflecting your k choices; we divide by $k! = k(k-1)\cdots 1$ because there are that many ways of choosing the *same* k elements in a different *order*, and we don't care about order. Now

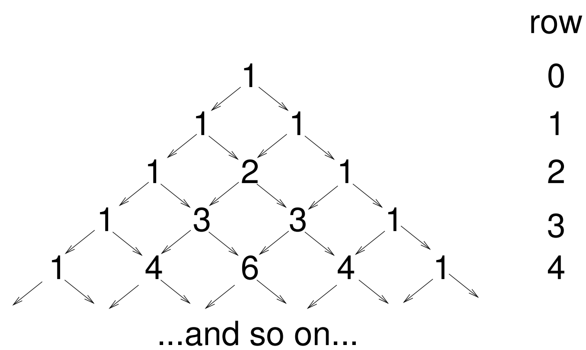
$$\begin{aligned}
(x+y)^n &= \overbrace{(x+y)\cdots(x+y)}^{n \text{ times}} \\
&= \cdots + Cx^k y^{n-k} + \cdots
\end{aligned}$$

where C is the number of ways to choose k of the $(x+y)$'s to extract an x from – that is, $\binom{n}{k}$. So we get the binomial formula:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Exercise 1.3. Find $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}$.

Pascal's triangle facilitates the quick computation of binomial numbers:



Each number in the triangle is (by definition) the sum of the numbers pointing to it.

Exercise 1.4. Write out three more rows.

Why does this work? That is, why does the k^{th} entry in the n^{th} row[2] equal $\binom{n}{k}$? This is tied to the binomial formula:

$$\begin{array}{rcccccccccc}
(x+y)^4 = (x+y)(x+y)^3 & & & & & & & & & & \\
= & x(x^3 & + & 3x^2y & + & 3xy^2 & + & y^3) & & & \\
+ & & & y(x^3 & + & 3x^2y & + & 3xy^2 & + & y^3) & \\
 & 1 & & 3 & & 3 & & 1 & & & \\
 & & & 1 & & 3 & & 3 & & 1 & \\
\hline
 & 1 & & 4 & & 6 & & 4 & & 1 &
\end{array}$$

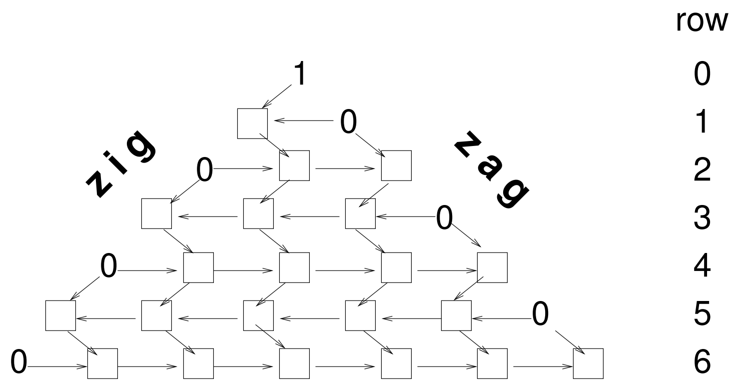
Another way of saying this is the identity

$$\binom{n}{i} + \binom{n}{i+1} = \binom{n+1}{i+1}.$$

Exercise 1.5. Can you show that $\binom{n}{k}$ equals the number of different paths, following the arrows, from the top of Pascal's triangle to the k^{th} position in the n^{th} row?

[2]The entries are numbered 0 to n.

1.3. **Zigzag numbers.** In the same spirit, drawing *Seidel's triangle*



we have the

Definition 1.6. The zigzag number $\begin{bmatrix} n \\ k \end{bmatrix}$ is the number of paths (along the arrows) from the top to the k^{th} entry in the n^{th} row. (Entries are counted in the direction of the arrows. The 0 in each row is considered the 0^{th} entry; that is, for $n > 0$, $\begin{bmatrix} n \\ 0 \end{bmatrix} = 0$.)

Exercise 1.7. Find the first few rows worth of zigzag numbers. What do you notice? Convince yourself that the identity

$$\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n \\ k-1 \end{bmatrix} + \begin{bmatrix} n-1 \\ n-k \end{bmatrix}$$

follows from the observation that paths not passing through the $(k-1)^{\text{st}}$ spot in the n^{th} row, pass through the $(n-k)^{\text{th}}$ place in the $(n-1)^{st}$ row.

We'll be especially keen on the numbers $\tau_n := \begin{bmatrix} n \\ n \end{bmatrix}$, which we'll call *zig numbers* for n odd and *zag numbers* for n even; the first few are:

n	0	1	2	3	4	5	6	7	8	9
τ_n	1	1	1	2	5	16	61	272	1385	7936

Proposition 1.8. *For $0 < k < n$ (on the interior of the triangle),*

$$\begin{bmatrix} n \\ k \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{k-1}{2} \rfloor} (-1)^i \binom{k}{2i+1} \tau_{n-2i-1} \tag{1.1}$$

$$\begin{bmatrix} n \\ k \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{n-k}{2} \rfloor} (-1)^i \binom{n-k}{2i} \tau_{n-2i} \tag{1.2}$$

Proof. We will use induction: assume that we know both identities for all the zigzag numbers coming before $\begin{bmatrix} n \\ k \end{bmatrix}$ in row n, and in all previous rows, and try to prove for $\begin{bmatrix} n \\ k \end{bmatrix}$. Now, there is a small problem with this strategy: since (for example) $\begin{bmatrix} n \\ 1 \end{bmatrix} = \begin{bmatrix} n-1 \\ n-1 \end{bmatrix} + \begin{bmatrix} n \\ 0 \end{bmatrix}$, we'll have to know something about the outer (noninterior) terms to get started in each row. So do (1.1) and (1.2) hold for these terms too?

As you may have guessed, the answer is NO: (1.1) fails for $\begin{bmatrix} n \\ n \end{bmatrix}$ for n even, and (1.2) fails for $\begin{bmatrix} n \\ 0 \end{bmatrix}$ for n odd. These are the only failures, and they are "connected", as they all happen on the "ZAG" edge of the triangle. So we have to correct the formulas in those two cases: for n even and $k = n$, replace (1.1) by

$$\begin{bmatrix} n \\ n \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^i \binom{n}{2i+1} \tau_{n-2i-1} + (-1)^{\frac{n}{2}}; \tag{1.3}$$

and for n odd and $k = 0$, replace (1.2) by

$$\begin{bmatrix} n \\ 0 \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^i \binom{n}{2i} \tau_{n-2i} + (-1)^{\frac{n+1}{2}}. \tag{1.4}$$

Now we are ready to proceed with the induction. By assumption, we have (from (1.1)/(1.3))

$$\begin{bmatrix} n \\ k-1 \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{k-2}{2} \rfloor} (-1)^i \binom{k-1}{2i+1} \tau_{n-2i-1} \ \Big(+(-1)^{\#} \text{ if needed}\Big) \tag{1.5}$$

and (from (1.2)/(1.4))

$$\begin{bmatrix} n-1 \\ n-k \end{bmatrix} = \sum_{i=0}^{\lfloor \frac{k-1}{2} \rfloor} (-1)^i \binom{k-1}{2i} \tau_{n-2i-1} \ \Big(+(-1)^{\#} \text{ if needed}\Big). \tag{1.6}$$

Recalling that $\begin{bmatrix} n \\ k-1 \end{bmatrix} + \begin{bmatrix} n-1 \\ n-k \end{bmatrix} = \begin{bmatrix} n \\ k \end{bmatrix}$ and $\binom{k-1}{2i+1} + \binom{k-1}{2i} = \binom{k}{2i+1}$, we now simply add (1.5) and (1.6) (ignoring the $(-1)^{\#}$ terms); this immediately gives (1.1) for $k < n$ or $k = n$ odd. For n even and $k = n$, (1.6) contributes a $(-1)^{\frac{n}{2}}$, and so we obtain (1.3). $\square$

Exercise 1.9. Complete the proof (i.e. deduce (1.2)/(1.4)) by using $\left[\begin{array}{c} n \\ k-1 \end{array}\right] = \left[\begin{array}{c} n \\ k \end{array}\right] - \left[\begin{array}{c} n-1 \\ n-k \end{array}\right]$ and inducing the other way.

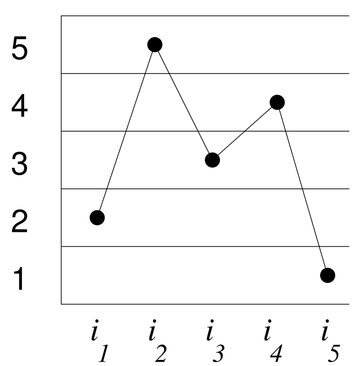
As an immediate consequence of the proof (specifically, (1.4)), we obtain a result which will be very useful later:

Corollary 1.10. $\sum_{i=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^i \binom{n}{2i} \tau_{n-2i} = \begin{cases} (-1)^{\frac{n-1}{2}}, & n \text{ odd} \\ 0, & n \text{ even} \end{cases}.$

1.4. **"Alternating" permutations.**

Definition 1.11. An **updown sequence** of length n is a sequence $i_1, i_2, \ldots, i_n$ such that $\{i_1, i_2, \ldots, i_n\} = \{1, 2, \ldots, n\}$ (it is a permutation, i.e. it contains each number once) and $i_1 < i_2 > i_3 < \cdots i_n$ (the "updown" or "alternating" feature). Let ε_n be the number of these.

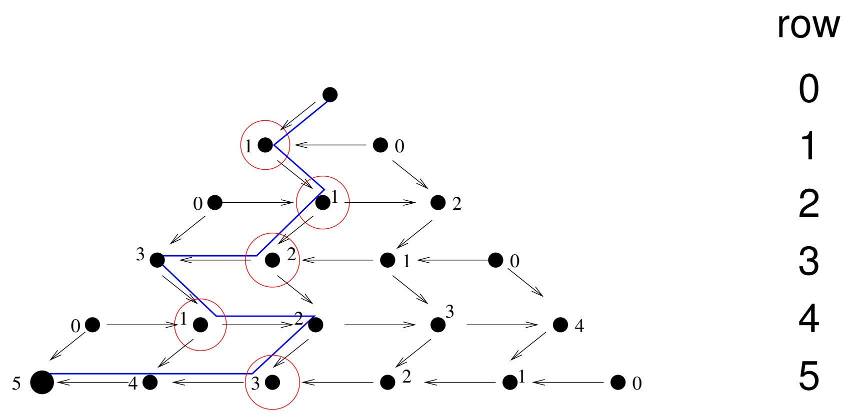
Exercise 1.12. Compute a few values of ε_n (say, for $n = 1$ to 5 or 6). It is helpful to draw "updown mountains" to depict the sequences: for example,



visualizes $\underset{i_1}{2} < \underset{i_2}{5} > \underset{i_3}{3} < \underset{i_4}{4} > \underset{i_5}{1}$

You should notice a rather striking coincidence!

So, what is the correspondence between updown sequences and zigzag numbers? Recall that the latter count paths in Seidel's triangle. For instance, $\tau_5 = \left[\begin{array}{c} 5 \\ 5 \end{array}\right]$ is the number of paths from the top of the triangle to the last (enlarged) dot in the 5th row, like this path:

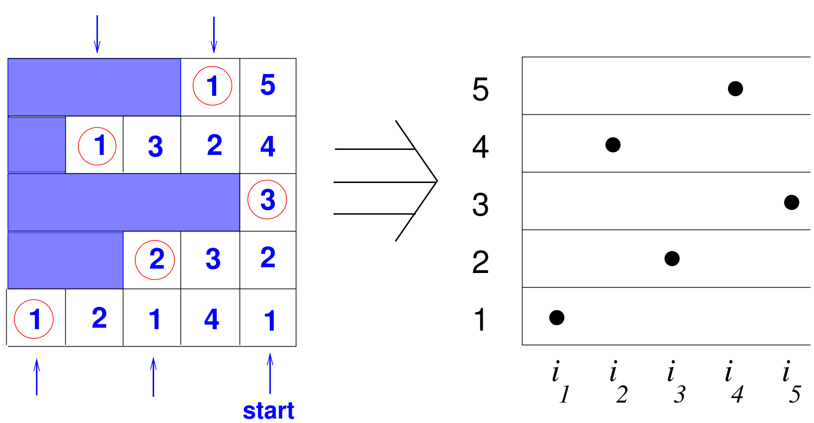


Here the numbers just indicate the values of n and k, and the circles indicate where the path *enters* each row. We have circled those values of k because they are the key to making an updown sequence out of the path.

Here's how: the list of circled values, from bottom to top, is

$$3,\ 1,\ 2,\ 1,\ 1. \tag{1.7}$$

Now draw a vertical stack of 5 boxes, and label them (from bottom to top) 1 thru 5. The first entry of (1.7) is 3, and we circle that label in the stack. Now draw new boxes to the left of each box *except for the one with the circle*, and label these from top to bottom (1 thru 4). The second entry of (1.7) is 1, and we circle that label in the new column. Continuing this process produces the figure below on the left:



The circles now give the graph of an updown sequence, as shown on the right: $1 < 4 > 2 < 5 > 3$.

The algorithm just described yields a "map" from paths in the Seidel triangle (following the arrows) to updown sequences. You may wish to check that the sequence so obtained is always "updown"!

Exercise 1.13. (a) Can you find a reverse correspondence? (This will confirm that the map is 1-to-1 and onto, so that $\tau_n = \varepsilon_n$.)

(b) More generally, can you prove that $\begin{bmatrix} n \\ k \end{bmatrix}$ is the number of updown sequences of length n in which $i_2 - i_1 \leq k$? (It's also the number of updown sequences with $i_1 \leq k$.)

1.5. **Bernoulli numbers.**

Definition 1.14. The **Bernoulli numbers** are an infinite sequence $B_0, B_1, B_2, \ldots$ of rational numbers, starting with $B_0 = 1$, and thereafter defined recursively by the formulas

$$\sum_{k=0}^{n} \binom{n+1}{k} B_k = 0 \tag{1.8}$$

for $n \geq 1$. They come up all over the place, in analysis, combinatorics, number theory, physics, etc.

It's a bit more convenient to add B_{n+1} to both sides of (1.8) to get

$$\sum_{k=0}^{n+1} \binom{n+1}{k} B_k = B_{n+1}.$$

This can be written more compactly (in a formal[3] shorthand known as "umbral calculus") as

$$(\mathbb{B} + 1)^{n+1} = \mathbb{B}^{n+1}, \tag{1.9}$$

which means "expand the left-hand side formally" via the binomial formula, then replace $\mathbb{B}^k$ everywhere by B_k.

Example 1.15. To start you off, we have:

$$\cancel{B_2} + 2B_1 + 1 = \cancel{B_2} \quad \Longrightarrow \quad B_1 = -\tfrac{1}{2},$$
$$\cancel{B_3} + 3B_2 + 3B_1 + 1 = \cancel{B_3} \quad \Longrightarrow \quad B_2 = \tfrac{1}{6}.$$

Exercise 1.16. Find the next few. What pattern(s) can you see? What can you say about the denominators? [Hint: induction!] When you're done, turn the page.

[3]The word "formal" sounds as if it should mean "very rigorous and careful", perhaps pedantically so. In math, it often means (as it does here) the opposite: going through the motions without regard as to meaning or correctness. You see, we mathematicians have craftily encoded our disdain for formalities in the language itself!

So here are some more Bernoulli numbers:

n	B_n
3	0
4	$-\frac{1}{30}$
5	0
6	$\frac{1}{42}$
7	0
8	$-\frac{1}{30}$
9	0
10	$\frac{5}{66}$
11	0
12	$-\frac{691}{2730}$
13	0
14	$\frac{7}{6}$
$\vdots$	$\vdots$

The odd ones (except B_1) are evidently zero, and the signs on the even ones apparently *alternate*. (A uniform proof of these facts will come later.) By induction, denominators (of $B_{n=2m}$) are clearly bounded by $(n+1)!$; the even better bound of $2 \cdot 3 \cdot 5 \cdots (2m-1)(2m+1)$ follows *if* you assume the $B_{2k+1} = 0$.

As you can see, the rational arithmetic makes them get quickly awful to compute. Bernoulli himself (~1690) did the first 10 or so. Euler lists the first 30 in his *Calculi Differentialis* (1755). Perhaps he found an alternative to brute force?

Exercise 1.17. Try multiplying B_{2m} $(m \geq 1)$ by $2^{2m-1}(4^m - 1)$. Notice anything? (Turn the page when you're ready.)

The first few values are

$m =$	1	2	3	4
$2^{2m-1}(4^m - 1)B_{2m} =$	1	-4	48	-1088

from which you might have noticed that they are divisible by m. So dividing by m (and multiplying by $(-1)^{m-1}$) gives $1, 2, 16, 272, \ldots$ which should look quite familiar, as they are none other than the odd τ's!!!! (I did say the Bernoulli numbers are ubiquitous!)

Though these coincidences just *can't* be coincidental, actually *proving* the following is much harder:

Theorem 1.18. $B_{2m} = \dfrac{(-1)^{m-1}2m}{4^{2m} - 4^m}\tau_{2m-1}$.

This gives a much more efficient way to compute the Bernoulli numbers. It's unlikely that this particular method was known to Euler, however, since Seidel's triangle dates from 1877.

1.6. **Faulhaber's formula.** We now describe the first main application of Bernoulli numbers, which they were invented to solve: to compute the power sums

$$S_k(n) = \sum_{i=1}^{n} i^k.$$

Working in our funny "umbral" notation, $(i+1+\mathbb{B})^{k+1} - (i+\mathbb{B})^{k+1} =$

$$= \sum_{j=0}^{k+1} \binom{k+1}{j} i^j (1+\mathbb{B})^{k+1-j} - \sum_{j=0}^{k+1} \binom{k+1}{j} i^j \mathbb{B}^{k+1-j}$$
$$= \sum_{j=0}^{k+1} \binom{k+1}{j} i^j \left\{(1+\mathbb{B})^{k+1-j} - \mathbb{B}^{k+1-j}\right\}.$$

But by (1.9), the quantity in braces is zero for $k+1-j \neq 1$ (i.e. $j \neq k$), and so the above

$$= (k+1)i^k \left\{1 + \cancel{\mathbb{B}^1 - \mathbb{B}^1}\right\}$$
$$= (k+1)i^k.$$

So altogether we have

$$i^k = \frac{1}{k+1}\left\{(i+1+\mathbb{B})^{k+1} - (i+\mathbb{B})^{k+1}\right\}. \tag{1.10}$$

That's a curiously complicated formula for i^k. What of it?

Theorem 1.19 (Faulhaber [first few cases], 1631; Bernoulli [general form], 1713; Jacobi [rigorous proof], 1834)**.** *The power sums are given by*

$$S_k(n) = \frac{1}{k+1}\sum_{j=0}^{k}\binom{k+1}{j}B_j(n+1)^{k+1-j}.$$

Proof. Apply $\sum_{i=0}^{n}$ to both sides of (1.10) to obtain

$$S_k(n) = \frac{1}{k+1}\sum_{i=0}^{n}\left\{(i+1+\mathbb{B})^{k+1} - (i+\mathbb{B})^{k+1}\right\},$$

which is a collapsing sum, telescoping to

$$\frac{1}{k+1}\left\{(n+1+\mathbb{B})^{k+1} - \mathbb{B}^{k+1}\right\}.$$

Expanding via the binomial formula and replacing $\mathbb{B}^j$ by B_j yields

$$\frac{1}{k+1}\left\{\sum_{j=0}^{k+1}\binom{k+1}{j}(n+1)^{k+1-j}B_j \; - \; B_{k+1}\right\},$$

where finally the B_{k+1} serves to cancel the $(k+1)^{\text{st}}$ term of the sum. ☐

Exercise 1.20. Use Theorem 1.19 to recover the formula for $k = 1$, and to obtain a formula for $k = 4$.

1.7. **Bernoulli polynomials.** These yield a slightly more elegant form of Faulhaber's formula. We will make more substantial use of them later on.

Definition 1.21. $B_k(x) := \sum_{j=0}^{k}\binom{k}{j}B_j x^{k-j}$.

Example 1.22. We have

$$B_3(x) = x^3 - \tfrac{3}{2}x^2 + \tfrac{1}{2}x\,,$$

$$B_4(x) = x^4 - 2x^3 + x^2 - \tfrac{1}{30}\,,$$

$$B_5(x) = x^5 - \tfrac{5}{2}x^4 + \tfrac{5}{3}x^3 - \tfrac{1}{6}x\,.$$

Corollary 1.23. $S_k(n) = \frac{1}{k+1}\left\{B_{k+1}(n+1) - B_{k+1}\right\}.$

Proof. Recognize $B_{k+1}(n+1) = \sum_{j=0}^{k+1}\binom{k+1}{j}B_j(n+1)^{k+1-j}$ as the right-hand side in Theorem 1.19, plus B_{k+1}. ☐

Exercise 1.24. Give another proof of this using $\binom{\ell}{m} = \binom{\ell+1}{m+1} - \binom{\ell}{m+1}$. [Hint: what is $B_{k+1}(j+1) - B_{k+1}(j)$?]

Remark 1.25. As discovered by Euler, the Bernoulli numbers are also related to the infinite inverse power sums ($k>1$)

$$1+\frac{1}{2^k}+\frac{1}{3^k}+\cdots=:\sum_{\ell=1}^{\infty}\frac{1}{\ell^k}.$$

This is the direction in which we shall now embark.

2. Infinite sums and generating functions

Toward the end of sixth grade, a classmate informed me that he had taught himself trigonometry. "I feel like Columbus", he said. No, actually, that was a speaker at a recent conference, describing his new computation, to which a colleague in the audience flatly exclaimed, "Wow, that's modest." Anyway . . . it's likely that my classmate said something equally self-congratulatory, given that humility is not one of the many virtues of 11-year-olds, so it might as well be Columbus.

Not to be left out, I ransacked my parents' books for a digest of the subject. I came across a thin volume entitled "Quick Calculus", which appeared to match my sixth-grade attention span while appealing to my ambition to reach the New World first. The truth is, a quarter of the way across I went blind from staring at the sun and fell overboard. Which is to say that, after nodding over the pages on sine and cosine, I reached the part on limits of functions, a concept I could barely grasp and whose point eluded me entirely, and gave up.

The first four sections that follow are my idiosyncratic version of "quick calculus". There is a clear motivation, which hopefully circumvents the customary indigestibility of a "digest". Namely, we want to understand connections between infinite sums, functions, and sequences (like the Bernoulli and updown numbers), and I will only develop what is necessary for that. If you have received your Calculus vaccination, you can skip (or skim) to §§2.5-6, which contain the proof of Theorem 1.18.

2.1. **Infinite series.**

Let $\{A_k\}_{k=1}^{\infty}$ (that is, $A_1, A_2, A_3, \ldots$) be a sequence of real numbers.

Definition 2.1. We say that A_k **converges** to the **limit** L (written $\lim_{k\to\infty} A_k = L$, or $A_k \to L$) if for each real $\epsilon>0$ – no matter how small – there exists $K\in\mathbb{N}$ such that

$$k\geq K \quad\Longrightarrow\quad |a_k-L|<\epsilon.$$

That is, one can get as close as desired to L by going out far enough in the sequence.

Suppose we want to sum the **infinite series** $\sum_{k=1}^{\infty} a_k$, for some sequence $a_k \to 0$. Defining the **sequence of partial sums**

$$A_n := \sum_{k=1}^{n} a_k,$$

we shall say that "the series converges" with **sum** $\mathscr{S}$, written $\sum_{k=1}^{\infty} a_k = \mathscr{S}$, if $A_n \to \mathscr{S}$.

Example 2.2. Let $a_k = x^k$, with $|x| < 1$, so that

$$A_n = 1 + x + \cdots + x^n.$$

Subtracting

$$xA_n = \quad x + \cdots + x^n + x^{n+1}$$

yields (after cancellations)

$$(1-x)A_n = 1 - x^{n+1},$$

whence the **geometric series**

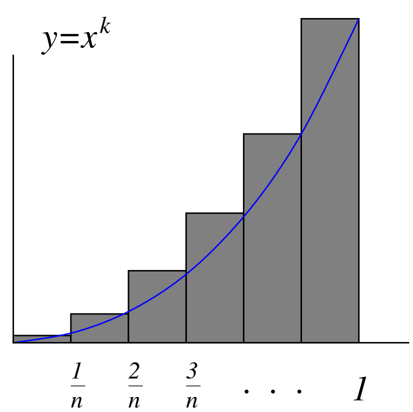
$$\mathscr{S} = \sum_{k=1}^{\infty} x^k = \lim_{n\to\infty} A_n = \lim_{n\to\infty} \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}.$$

Example 2.3. Let $a_k = \frac{1}{k}$. Then the **harmonic series**

$$\sum a_k = 1 + \frac{1}{2} + \underbrace{\left(\frac{1}{3} + \frac{1}{4}\right)}_{>\frac{1}{2}} + \underbrace{\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)}_{>\frac{1}{2}} + \cdots$$

cannot converge, as it grows without bound. (As we shall see, $\sum \frac{1}{k^2}$, $\sum \frac{1}{k^3}$, etc. have no such problem.)

2.2. **Definite integral.** Another game you can play with taking limits of sums is finding the area A under a curve:



Denoting by A_n the sum of the areas under the n "vertical" boxes, we evidently have

$$A_n = \sum_{i=1}^{n} \frac{1}{n} \cdot \left(\frac{i}{n}\right)^k = \frac{1}{n^{k+1}} S_k(n),$$

which by Faulhaber's formula

$$= \frac{n^{k+1} + \text{lower-order terms}}{n^{k+1}(k+1)} = \frac{1}{k+1} + \frac{c_1}{n} + \frac{c_2}{n^2} + \cdots + \frac{c_{k+1}}{n^{k+1}}$$

for some constants c_i. According to the picture, each of these A_n overestimates A, whose correct value is found by shrinking the widths of the boxes to zero:

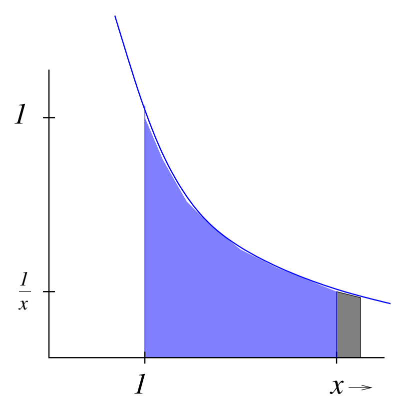
$$A = \lim_{n\to\infty} A_n = \frac{1}{k+1}.$$

This area is called the **definite integral** of x^k over the interval $[0,1]$, denoted

$$\int_0^1 x^k dx.$$

Exercise 2.4. How about the area under $y = x^k$ from 0 to r?

What if we consider $y = \frac{1}{x}$? The area from 0 to 1 (or to anywhere) is ∞, since you can insert "horizontal" boxes between the graph and the y-axis whose sum-of-areas is the harmonic series. So consider instead the area from 1 to x:



and call this $\mathbf{ln(x)}$, the **natural logarithm**. The picture suggests that the rate of change of $\ln(x)$ against x is the height of the graph at that value of x, i.e. $\frac{1}{x}$. So then the rate of change of $\ln(ax)$ is

$$\begin{pmatrix} \text{rate of change of } ax \\ \text{with respect to } x \end{pmatrix} \times \begin{pmatrix} \text{rate of change of } \ln(ax) \\ \text{with respect to } ax \end{pmatrix} = a \cdot \frac{1}{ax} = \frac{1}{x}.$$

Finally, the rate of change of $\ln(a) + \ln(x)$ is again $\frac{1}{x}$, since $\ln(a)$ is constant; while at at $x = 1$, $\ln(a) + \ln(x) = \ln(a) + \ln(1) = \ln(a \cdot 1)$. Conclude that $\ln(a) + \ln(x)$ and $\ln(ax)$ have the same rates of change everywhere and the same values at $x = 1$, so are the same function! Another way to say this is

$$\ln(ab) = \ln(a) + \ln(b). \tag{2.1}$$

Now define the **exponential function $\mathbf{exp(x)}$** to be the inverse function of $\ln(x)$. That is,

$$y = \exp(x) \quad \stackrel{\text{defn.}}{\Longleftrightarrow} \quad x = \ln(y).$$

Then by (2.1),

$$\ln(\exp(a)\exp(b)) = \ln(\exp(a)) + \ln(\exp(b)) = a + b,$$

whereupon taking exp of both sides yields

$$\exp(a)\exp(b) = \exp(a + b). \tag{2.2}$$

So exp behaves like "something raised to a power". That something is

$$e := \exp(1) \simeq 2.71828182845904523536\ldots,$$

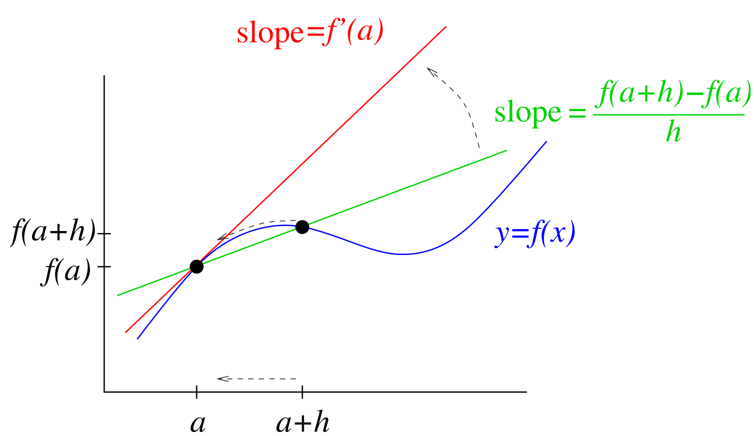
and $\exp(x) = e^x$.

2.3. **Slopes of graphs.** Continuing on with the "rate of change" theme, here is a more solid definition for the rate of change of $f(x)$ with respect to x:

Definition 2.5. The **derivative** of f at a is

$$\frac{df}{dx}(a) = f'(a) := \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}.$$

You should visualize this as the slope of the (red) tangent line at $(a, f(a))$, given as the *limit* of slopes of (green) secant lines through $(a, f(a))$ and $(a + h, f(a + h))$:



We develop this through a series of examples, which are needed in the sequel.

Example 2.6. The discussion in §2.2 translates to the calculation

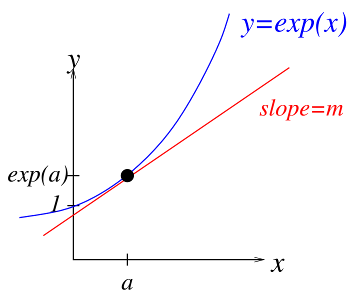
$$\ln'(a) = \lim_{h\to 0} \frac{\ln(a+h) - \ln(a)}{h}$$
$$= \lim_{h\to 0} \frac{\text{area under } y = \frac{1}{x} \text{ from } a \text{ to } a+h}{h} = \frac{1}{a},$$

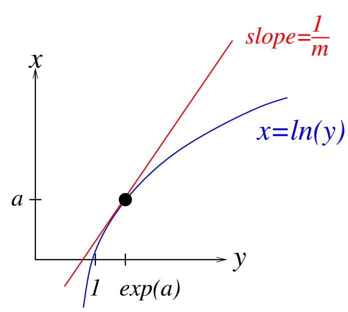
since the area in the numerator is approximated by a box of width h and height $\frac{1}{a}$.

Example 2.7. For $f(x) = x^k$, we have

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^k - x^k}{h} = \lim_{h\to 0} \frac{khx^{k-1} + h^2(\cdots)}{h} = kx^{k-1}.$$

Example 2.8. Since exp and ln are inverses, their graphs are reflections of one another about the line $y = x$:





and so their tangent lines are too, with slope $m = \frac{\text{rise}}{\text{run}}$ reflected to $\frac{\text{run}}{\text{rise}} = \frac{1}{m}$. Thus

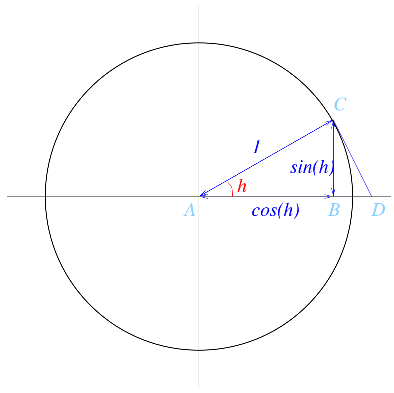
$$\exp'(a) = \frac{1}{\ln'(\exp(a))} = \frac{1}{\left(\frac{1}{\exp(a)}\right)} = \exp(a),$$

i.e. the derivative of e^x is e^x.

Example 2.9. Turning to trigonometric functions, we make use of a trig identity to write

$$\begin{aligned} \sin'(x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h} \\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h) + \sin(h)\cos(x) - \sin(x)}{h} \\ &= (\sin(x)) \lim_{h\to 0} \frac{\cos(h) - 1}{h} + (\cos(x)) \lim_{h\to 0} \frac{\sin(h)}{h}. \end{aligned}$$

To evaluate the last two limits, look at the unit circle diagram



with similar triangles ABC and ACD (so that $|CD| = \frac{\sin(h)}{\cos(h)} = \tan(h)$). The angular sector with angle[4] h contains ABC but is contained by ACD; taking areas, this gives

$$\frac{1}{2} \cdot \cos(h) \cdot \sin(h) < \frac{h}{2} < \frac{1}{2} \cdot 1 \cdot \frac{\sin(h)}{\cos(h)}.$$

Dividing through by $\sin(h)$ and taking inverses gives

$$\cos(h) < \frac{\sin(h)}{h} < \frac{1}{\cos(h)},$$

whose outer terms limit to 1, forcing the inner term to do the same. Then the Pythagorean theorem gives

$$\frac{1-\cos(h)}{h} = \frac{1-\cos^2(h)}{h(1+\cos(h))} = \frac{\sin^2(h)}{h(1+\cos(h)} = \frac{\sin(h)}{h} \cdot \frac{\sin(h)}{1+\cos(h)}$$

which limits to $1 \cdot 0 = 0$. We conclude that $\sin'(x) = \cos(x)$.

Exercise 2.10. Check that $\cos'(x) = -\sin(x)$. Then use the examples below to find $\tan'(x)$ and $\sec'(x)$.

Example 2.11. For a product of functions, $f(x) = F(x)G(x)$, we compute

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{F(x+h)G(x+h) - F(x)G(x)}{h} \\ &= \lim_{h\to 0} \frac{F(x+h)G(x+h) - F(x+h)G(x) + F(x+h)G(x) - F(x)G(x)}{h} \\ &= F(x) \lim_{h\to 0} \frac{G(x+h)-G(x)}{h} + G(x) \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} \\ &= F(x)G'(x) + F'(x)G(x). \end{aligned}$$

[4] All our angles are in radians: i.e. h is just the length of the arc on the circle from the x-axis to C.

Example 2.12. If $f(x) = 1/F(x)$, then

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\frac{1}{F(x+h)} - \frac{1}{F(x)}}{h} \\ &= \lim_{h\to 0} \frac{-1}{F(x+h)F(x)} \cdot \frac{F(x+h)-F(x)}{h} \\ &= \frac{-F'(x)}{(F(x))^2}. \end{aligned}$$

Example 2.13 (Quotient rule)**.** Combining the last two examples, for a quotient $f(x) = \frac{F(x)}{G(x)} = F(x) \cdot \frac{1}{G(x)}$, we get

$$f'(x) = \frac{G(x)F'(x) - F(x)G'(x)}{G(x)^2}.$$

Remark 2.14. One application of derivatives is to computing limits. Suppose that $f(x)$ and $g(x)$ both limit to 0 as $x \to 0$. To what limit does their ratio tend? **L'Hôpital's rule**

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0} \frac{f'(x)}{g'(x)}$$

can sometimes help. For instance, $\lim_{x\to 0} \frac{\sin(x)}{x} = \lim_{x\to 0} \frac{\cos(x)}{1} = 1$.

2.4. **Power series.** So far we have only discussed summing series of numbers. But it's equally natural to consider *function series*: when those numbers depend on x, so will their sum (where it exists). If the functions we are summing are of a particular type, then the series go by that name as well – like *trigonometric series* for sums of sines or cosines, and *power series* for sums of powers of x.

For instance, given a function $f(x)$, we can look at the power series

$$g(x) := \sum_{k\geq 0} \frac{f^{(k)}(0)}{k!} x^k \tag{2.3}$$

for x near 0. (Here $f^{(n)}(x)$ means $f^{\overbrace{\prime\prime\prime\cdots\prime}^{n \text{ times}}}(x)$, and we are assuming that f is a reasonably "nice" function so that all of those successive derivatives exist, at least at 0.) Formally taking the derivative of both sides of (2.3) n times yields

$$g^{(n)}(x) = \sum_{k\geq n} \frac{f^{(k)}(0)}{(k-n)!} x^{k-n},$$

whereupon setting $x = 0$ gives

$$g^{(n)}(0) = f^{(n)}(0) \quad (\forall n). \tag{2.4}$$

Example 2.15. Let's look at $f(x) = \frac{1}{1-x}$: differentiating repeatedly gives

$$\begin{aligned} f'(x) &= \tfrac{1}{(1-x)^2} &&\implies f'(0) = 1 \\ f''(x) &= \tfrac{2}{(1-x)^3} &&\implies f''(0) = 2 \\ &\vdots && \qquad\vdots \\ f^{(k)}(x) &= \tfrac{k!}{(1-x)^{k+1}} &&\implies f^{(k)}(0) = k!, \end{aligned}$$

and so $g(x) = \sum_{k\geq 0} \frac{k!}{k!} x^k = \sum_{k\geq 0} x^k$ . . . which, by Example 2.2, recovers $f(x)$!

Indeed, for any analytic function (i.e. all the functions you know), (2.4) means that (where the sum of the series exists, i.e. at least for small x) $f(x) = g(x)$ are the same function.

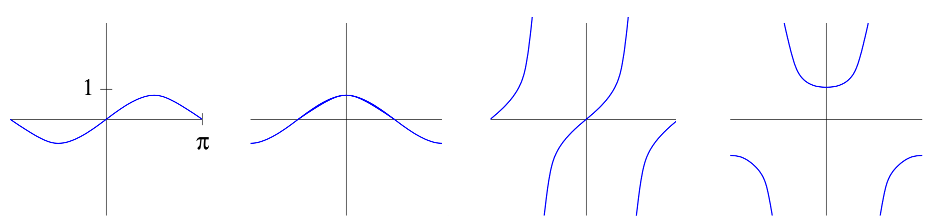
Example 2.16. If $f(x) = e^x$, then $f^{(n)}(x) = e^x \implies f^{(n)}(0) = 1$ $(\forall n)$ $\implies e^x = \sum_{k\geq 0} \frac{x^k}{k!}$, which holds for all x.

Example 2.17. For $f(x) = \sin(x)$, we find a repeating pattern (of length four):

$$\begin{aligned} f'(x) &= \cos(x) &&\implies f'(0) = 1 \\ f''(x) &= -\sin(x) &&\implies f''(0) = 0 \\ f^{(3)}(x) &= -\cos(x) &&\implies f^{(3)}(0) = -1 \\ f^{(4)}(x) &= \sin(x) &&\implies f^{(4)}(0) = 0, \end{aligned}$$

so that $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$. Taking derivatives on both sides gives $\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$. Both sums hold for all x.

Example 2.18. What about the functions $\tan(x) = \frac{\sin(x)}{\cos(x)}$ and $\sec(x) = \frac{1}{\cos(x)}$?



y=sin(x) *y=cos(x)* *y=tan(x)* *y=sec(x)*

Though one can compute successive derivatives by "brute force" via the quotient rule, the resulting functions quickly become horrible. Still, we can get the first few terms of the power series representations this way:

$$\tan(x) = x + \frac{2x^3}{3!} + \frac{16x^5}{5!} + \frac{272x^7}{7!} + \cdots$$

$$\text{and} \quad \sec(x) = 1 + \frac{x^2}{2!} + \frac{5x^4}{4!} + \frac{61x^6}{6!} + \cdots \quad ,$$

where the presence of only odd (resp. even) powers of x in tan (resp. sec) reflects the fact that the function is odd (resp. even).

2.5. **Generating functions.** Whenever you have an infinite sequence $\{a_k\}_{k=0}^{\infty}$, a natural problem is to determine what sort of functions are given by the sums of

$$\sum_{k\geq 0} a_k x^k \qquad \textbf{(generating function)} \qquad \text{or}$$

$$\sum_{k\geq 0)} a_k \frac{x^k}{k!} \qquad \textbf{(exponential generating function)}.$$

The solution can be used to prove *relations* between different sequences by manipulating the *functions*, a standard tool in combinatorics.

Example 2.19. We consider the exponential generating function of the Bernoulli sequence,

$$f(x) = \sum_{k\geq 0} B_k \frac{x^k}{k!}.$$

By (1.8) we have $\sum_{k=0}^{n-1} \binom{n}{k} B_k = 0$, hence $\sum_{k=0}^{n-1} \frac{B_k}{k!(n-k)!} = 0$, for $n \geq 2$. So

$$(e^x - 1)f(x) = \sum_{\ell\geq 1} \frac{x^\ell}{\ell!} \sum_{k\geq 0} B_k \frac{x^k}{k!} = \sum_{n\geq 1} \left(\sum_{k=0}^{n-1} \frac{B_k}{k!(n-k)!} \right) x^n = x,$$

and $f(x) = \frac{x}{e^x - 1}$.

Example 2.20. What if we try the same thing for the zigzag/updown sequence τ_k? Dividing the result of Corollary 1.10 by $n!$ gives

$$\sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^j \frac{\tau_{n-2j}}{(2j)!(n-2j)!} = \begin{cases} \frac{(-1)^{\frac{n-1}{2}}}{n!}, & n \text{ odd} \\ 0, & n > 0 \text{ even} \\ 1, & n = 0. \end{cases}$$

The exponential generating function

$$g(x) = \sum_{k\geq 0} \tau_k \frac{x^k}{k!}$$

then satisfies

$$\begin{aligned}
\cos(x)g(x) &= \sum_{j\geq 0}(-1)^j\frac{x^{2j}}{(2j)!}\sum_{k\geq 0}\tau_k\frac{x^k}{k!} \\
&= \sum_{n\geq 0}\left(\sum_{j=0}^{\lfloor\frac{n}{2}\rfloor}(-1)^j\frac{\tau_{n-2j}}{(2j)!(n-2j)!}\right)x^n \\
&\underset{n=2m+1}{=} 1+\sum_{m\geq 0}\frac{(-1)^m}{(2m+1)!}x^{2m+1} \\
&= 1+\sin(x).
\end{aligned}$$

Dividing through by $\cos(x)$ now gives

$$g(x) = \sec(x) + \tan(x), \tag{2.5}$$

so that taking account of the even-ness of $\sec(x)$ and odd-ness of $\tan(x)$, we have

$$\sec(x) = \underset{\textbf{ZAG}}{\sum_{m\geq 0}\tau_{2m}\frac{x^{2m}}{(2m)!}} \quad \text{and} \quad \cos(x) = \underset{\textbf{ZIG}}{\sum_{m\geq 1}\tau_{2m+1}\frac{x^{2m+1}}{(2m+1)!}}$$

so that

$$\sec^{(2m)}(0) = \tau_{2m} \quad \text{while} \quad \tan^{(2m+1)}(0) = \tau_{2m+1}.$$

You may have guessed this from the first few terms of the power series in Example 2.18, but it took a bit more than the quotient rule to prove it!

Remark 2.21. Equation (2.5) was discovered by D. André in 1881.

2.6. **Euler's identity.** A more serious test of this "generating functionology" is whether it can give us the desired relation (Theorem 1.18) between the even Bernoulli and "zig" numbers, B_{2m} and τ_{2m-1}. With $f(x)$ as in Example 2.19, first note that

$$\sum_{m\geq 1}B_{2m}\frac{x^{2m}}{(2m)!} = f(x) - B_0\frac{x^0}{0!} - B_1\frac{x^1}{1!} = \frac{x}{e^x-1} - 1 + \frac{x}{2}. \tag{2.6}$$

We will also make use of an important relation between Examples 2.16 and 2.17: writing[5] $\mathbf{i} := \sqrt{-1}$,

$$\begin{aligned} e^{\mathbf{i}x} &= \sum_{k\geq 0} \frac{(\mathbf{i}x)^k}{k!} \\ &= \sum_{m\geq 0} \frac{(\mathbf{i}x)^{2m}}{(2m)!} + \sum_{m\geq 0} \frac{(\mathbf{i}x)^{2m+1}}{(2m+1)!} \\ &= \sum_{m\geq 0} (-1)^m \frac{x^{2m}}{(2m)!} + \mathbf{i} \sum_{m\geq 0} (-1)^m \frac{x^{2m+1}}{(2m+1)!} \\ &= \cos(x) + \mathbf{i}\sin(x), \end{aligned}$$

which (using $\cos(-x) = \cos(x)$ and $\sin(-x) = -\sin(x)$) yields

$$e^{\mathbf{i}x} + e^{-\mathbf{i}x} = 2\cos(x) \quad \text{and} \quad e^{\mathbf{i}x} - e^{-\mathbf{i}x} = 2\mathbf{i}\sin(x). \tag{2.7}$$

The result

$$e^{\mathbf{i}x} = \cos(x) + \mathbf{i}\sin(x) \tag{2.8}$$

is called "Euler's identity".

Now we use (2.6)[6] and (2.7) to compute

$$\begin{aligned} &\sum_{m\geq 1} B_{2m} \left(2^{2m} - 4^{2m}\right) \overbrace{(-1)^m}^{\mathbf{i}^{2m}} \frac{x^{2m}}{(2m)!} \\ &\quad = \sum_{m\geq 1} B_{2m} \frac{(2\mathbf{i}x)^{2m}}{(2m)!} - \sum_{m\geq 1} B_{2m} \frac{(4\mathbf{i}x)^{2m}}{(2m)!} \\ &\quad = \left(\frac{2\mathbf{i}x}{e^{2\mathbf{i}x} - 1} - \cancel{1} + \frac{2\mathbf{i}x}{2} \right) - \left(\frac{4\mathbf{i}x}{e^{4\mathbf{i}x} - 1} - \cancel{1} + \frac{4\mathbf{i}x}{2} \right) \\ &\quad = \frac{2\mathbf{i}x(e^{2\mathbf{i}x} + 1) - 4\mathbf{i}x}{(e^{2\mathbf{i}x} - 1)(e^{2\mathbf{i}x} + 1)} - \mathbf{i}x \quad = \quad \frac{2\mathbf{i}x(e^{2\mathbf{i}x} - 1)}{(e^{2\mathbf{i}x} - 1)(e^{2\mathbf{i}x} + 1)} - \mathbf{i}x \\ &\quad = \frac{2\mathbf{i}x}{e^{2\mathbf{i}x} + 1} - \mathbf{i}x \quad = \quad \frac{\cancel{2}\mathbf{i}x - e^{2\mathbf{i}x}\mathbf{i}x - \cancel{\mathbf{i}x}}{e^{2\mathbf{i}x} + 1} \\ &\quad = \frac{\mathbf{i}x(1 - e^{2\mathbf{i}x})}{e^{2\mathbf{i}x} + 1} \quad = \quad \frac{-\mathbf{i}x(e^{\mathbf{i}x} - e^{-\mathbf{i}x})}{e^{\mathbf{i}x} + e^{-\mathbf{i}x}} \\ &\quad = \mathbf{i}x \frac{2\mathbf{i}\sin(x)}{2\cos(x)} \quad = \quad x\tan(x). \end{aligned}$$

[5] I'll say a bit about complex numbers in the next subsection; for now, you can just interpret this "formally", as a symbol whose "square" is -1.

[6] with $2\mathbf{i}x$ and $4\mathbf{i}x$ formally substituted for x

On the other hand, by Example 2.20,

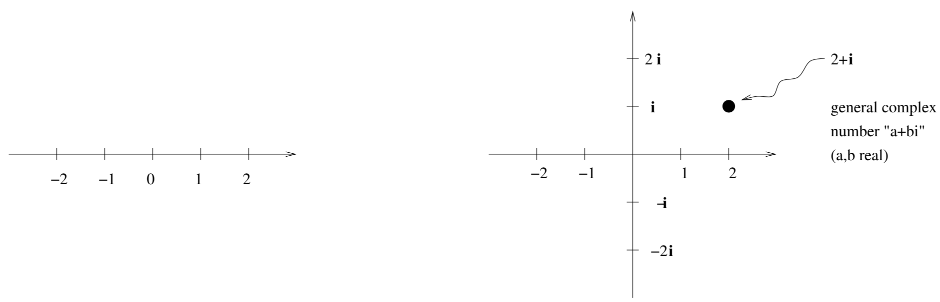
$$\sum_{m\geq 1} \tau_{2m-1} \frac{x^{2m}}{(2m-1)!} = x \sum_{m\geq 1} \tau_{2m-1} \frac{x^{2m-1}}{(2m-1)!} = x\tan(x)$$

as well. Given two power series (as we have here) both summing to $x\tan(x)$, the coefficients of powers of x must be equal: by (2.4), they are determined by the derivatives of $x\tan(x)$ at zero. So we conclude that

$$B_{2m} = \frac{(-1)^{m-1} 2m}{4^{2m} - 4^m} \tau_{2m-1},$$

proving Theorem 1.18.

2.7. **Complex numbers.** So what is this "**i**"? We can think of it as a first step in extending the real number line to the "complex plane":

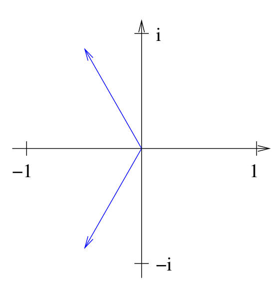


The resulting **complex numbers** $a + b\mathbf{i}$ (points on this plane) were invented in the late 1700's to solve equations.

Example 2.22. $0 = x^2 + 3x + 2 = (x+2)(x+1)$ has two real solutions $x = -2, -1$; while $0 = x^2 + x + 1$ has (by the quadratic formula) solutions

$$x_\pm = \frac{-1 \pm \sqrt{1^2 - 4\cdot 1\cdot 1}}{2\cdot 1} = \frac{-1\pm\sqrt{-3}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}\mathbf{i}.$$

We can represent these solutions by points in the complex number plane, or better, by arrows:

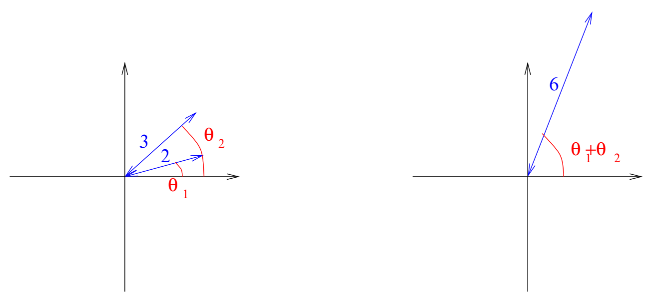


Notice that these (points or arrows) are the flips of each other about the "real axis". This flip is called complex conjugation and is written with a bar: $\overline{a+b\mathbf{i}} = a - b\mathbf{i}$; complex solutions to real quadratic equations are always conjugate.

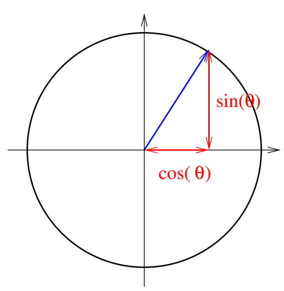
What about arrows is "better"?

- Adding complex numbers can be visualized by adding them:

- Multiplying complex numbers can be visualized by multiplying the lengths of the arrows and adding the angles they make with the x-axis:



- By the Pythagorean theorem, $e^{\mathbf{i}\theta} = \cos(\theta) + \mathbf{i}\sin(\theta)$ is a point on the unit circle:



  so we can write any complex number $a+b\mathbf{i} = \sqrt{a^2+b^2}e^{\mathbf{i}\theta}$ where $\theta = \arctan(\frac{b}{a})$ (and $\sqrt{a^2+b^2}$ is the length of the arrow).

Exercise 2.23. What are $e^{2\pi\mathbf{i}}$, $e^{\pi\mathbf{i}}$, $e^{\frac{\pi\mathbf{i}}{2}}$? Convince yourself, more generally, that $e^{2\pi\mathbf{i}/n} =: \zeta_n$ is an "n^{th} root of 1", i.e. $(\zeta_n)^n = 1$.

If this seems bizarre, just remember that once people scorned the idea of -5 apples and $\frac{2}{3}$ of a person, but now we speak of negative bank balances and the average household with 2.3 kids with alacrity. The bottom line is that all numbers are theoretical constructs, including the "real" ones!

3. Discrete Fourier transforms and special values of L-functions

My junior year in high-school, I worked in a photonics laboratory at a nearby university,[7] using lasers, lenses, and LCD filters to produce holograms. The underlying mathematics was a "fast" 2-D version (implemented on MATLAB and used to program the filter) of the 1-D discrete Fourier transform described in this section. Almost two decades later, I was shocked to discover that the 2-variable DFT played an equally central role in the study of algebraic cycles and number theory! Below I'll sketch how Fourier transforms arose and how they relate to signal processing and Bernoulli numbers, and some interesting infinite series.

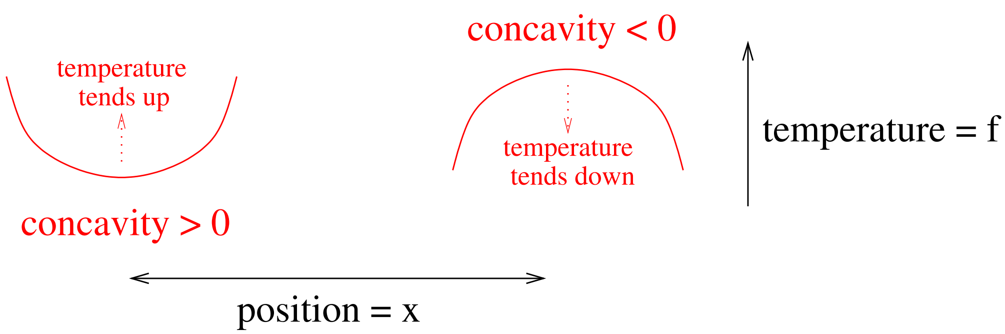
3.1. **Fourier series.** Here is a good example of revolutionary mathematics motivated by physics.[8] Fourier had introduced the **heat equation**

$$\underbrace{\frac{\partial f}{\partial t}(x,t)}_{\substack{\text{rate of change}\\ \text{of temperature}\\ \text{with resp. to time}}} = \underbrace{\alpha^2 \frac{\partial^2 f}{\partial x^2}(x,t)}_{\substack{\text{concavity of}\\ \text{temp. distribution}}}$$

[7]George Mason University, under the inspiring direction of Dr. Ravi Athale

[8]Though mathematics and physics are now more separate as disciplines than they were in the 19th Century, this phenomenon continues unabated, with the unexpected and profound unifying influence of quantum field theory and string dualities on algebraic geometry, symplectic topology, and number theory over the past three decades.

for a periodic[9] temperature distribution $f(x,t)$ with initial conditions $f(x,0) =: f(x)$:



He noticed that if you could represent the initial conditions by trigonometric[10] series

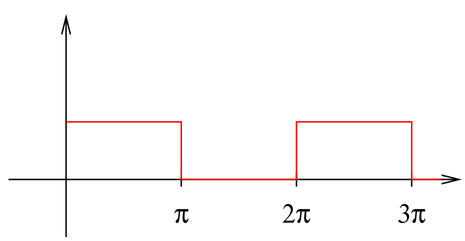
$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{\mathbf{i}nx} \qquad (a_n \in \mathbb{C}) \tag{3.1}$$

then

$$f(x,t) = \sum_{n\in\mathbb{Z}} \left(a_n e^{-\alpha^2 n^2 t}\right) e^{\mathbf{i}nx}$$

solved it. This reveals that *the highest frequencies are smoothed most rapidly.*

Fourier's idea that (3.1) should be possible even for discontinuous $f(x)$ — e.g. the square wave function



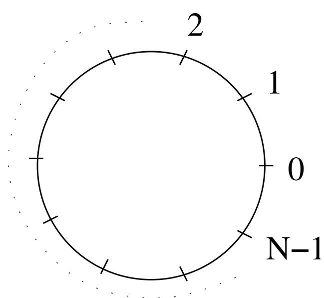
obtained by "shoving an iron ring halfway into a fire" — caused controversy and led to the rejection of his 1807 paper. Eventually, he became president of the professional society that rejected his paper, and had it published in their prestigious journal. (In fact, the so-called **Fourier series** of a periodic function — even a continuous one — need not converge everywhere to the function, unless the function is everywhere differentiable.)

[9]Period 2π: that is, think of it as the temperature function on a circle.
[10]Recall that $e^{\mathbf{i}nx} = \cos(nx) + \mathbf{i}\sin(nx)$. (Also, $\sum_{n\in\mathbb{Z}}$ means $\sum_{n=-\infty}^{\infty}$.)

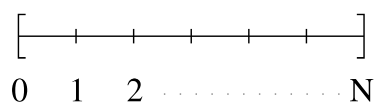
3.2. **Discrete Fourier transform.** This is the finite version, typically programmed on computers via the "fast Fourier transform" (FFT) algorithm. (Unlike Fourier series and Fourier transforms, this requires no calculus.) Again we will be dealing with periodic functions, but on the integers:

$$f(0),\ f(1),\ f(2),\ \ldots,\ f(N-1);\ \underbrace{f(N)}_{=f(0)},\ \underbrace{f(N+1)}_{=f(1),\,\text{etc.}},\ \ldots$$

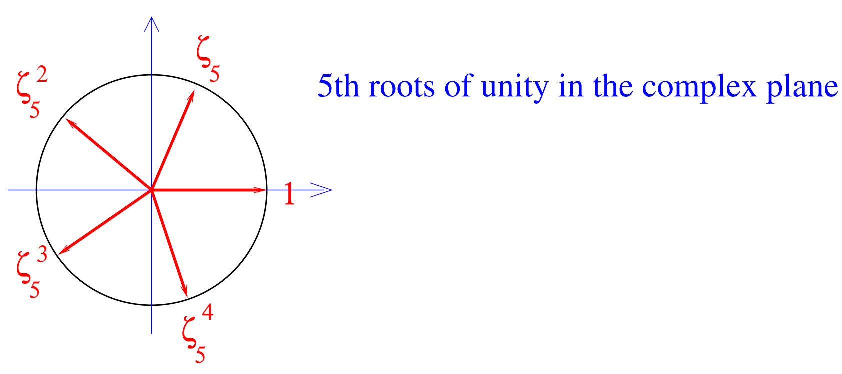
You can think of this as an N-point sample of a function on a circle,



or perhaps just of a function on an interval:



Now the $\mathbf{N^{th}}$ **root of unity**[11] $\zeta_N = e^{\frac{2\pi \mathbf{i}}{N}}$ and its powers



will play an important role.

Exercise 3.1. Show that $\sum_{n=0}^{N-1} \zeta_N^n = 0$. [Hint: multiply by $(1-\zeta_N)$.] Can you show that this also holds when ζ_N is replaced by a power ζ_N^k (so long as this power is not 1)?

[11]"unity" here is a fancy way of saying "1".

The function ζ_N^{km} (of m) cycles through 1 k times as m goes from 0 to N. So it has "frequency k".

Now set[12]

$$\begin{aligned}\Phi(N) &:= \text{ set of functions } f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C} \\ \Phi(N)^\circ &:= \text{ the functions with } f(0) = 0 \\ \Phi(N)_\circ &:= \text{ the functions with } \sum_{n=0}^{N-1} f(n) = 0.\end{aligned}$$

Definition 3.2. The **discrete Fourier transform (DFT)** is the operation (on functions from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$) sending f to $\hat{f}$, where

$$\hat{f}(k) := \sum_{n=0}^{N-1} f(n) e^{-\frac{2\pi \mathbf{i} kn}{N}} = \sum_{n=0}^{N-1} f(n)\zeta_N^{-kn}.$$

The idea is that $\hat{f}(k)$ measures how much stuff "of frequency k" $f(n)$ has. The following formula for the **inverse DFT** makes this plain:

Proposition 3.3. *We may recover f from $\hat{f}$ by*

$$f(m) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{f}(k)\zeta_N^{km}. \tag{3.2}$$

Proof. $\frac{1}{N}\sum_k \hat{f}(k)\zeta_N^{km} = \frac{1}{N}\sum_k \left(\sum_n f(n)\zeta_N^{-kn}\right)\zeta_N^{km} = \frac{1}{N}\sum_{k,n} f(n)\zeta_N^{k(m-n)}$ $= \sum_n f(n)\left\{\frac{1}{N}\sum_{k=0}^{N-1}(\zeta_N^{m-n})^k\right\}$. By Exercise 3.1, the bracketed expression is 0 if $m \neq n$ in $\mathbb{Z}/N\mathbb{Z}$; if $m = n$ it is $\frac{N}{N} = 1$. Therefore we get $f(m)$. □

Accordingly, $\hat{f}(k)$ is sometimes called the k^{th} **Fourier coefficient** of f.

3.3. Properties of the DFT.

3.3.1. *Duality of* Φ° *and* $\Phi_\circ$. I claim that $f \in \Phi^\circ \iff \hat{f} \in \Phi_\circ$, and $f \in \Phi_\circ \iff \hat{f} \in \Phi^\circ$.

Exercise 3.4. Prove it! Say $f(0) = 0$. Can you show that $\sum_k \hat{f}(k) = 0$? [Hint: use the definition of the DFT together with Exercise 3.1.]

[12] Notation: $\mathbb{C}$ = complex numbers; $\mathbb{Z}/N\mathbb{Z}$ = integers modulo N (i.e. on the circle, with $N \equiv 0$).

3.3.2. *Parseval's formula.* Write $|a+b\mathbf{i}| := \sqrt{a^2+b^2}$ for the "length" of a complex number. Notice that

$$|a+b\mathbf{i}|^2 = a^2+b^2 = (a+b\mathbf{i})(a-b\mathbf{i}) = (a+b\mathbf{i})\overline{(a+b\mathbf{i})}.$$

Parseval says that

$$\sum_n |f(n)|^2 = \frac{1}{N}\sum_k |\hat{f}(k)|^2, \tag{3.3}$$

which a physicist might read as "f and $\hat{f}$ have the same energy".

Exercise 3.5. Using Euler's identity (2.8), check that $\overline{e^{\mathbf{i}\theta}} = e^{-\mathbf{i}\theta}$.

Proof of Parseval.

$$\begin{aligned}\sum_k |\hat{f}(k)|^2 = \sum_k \hat{f}(k)\overline{\hat{f}(k)} &= \sum_k\sum_n f(n)\zeta_N^{-nk}\sum_m \overline{f(m)}\zeta_N^{mk}\\ &= \sum_{m,n} f(n)\overline{f(m)}\underbrace{\sum_k (\zeta_N^{m-n})^k}_{\substack{0 \text{ if } m\neq n\\ N \text{ if } m=n}} = N\sum_n f(n)\overline{f(n)},\end{aligned}$$

so that dividing by N yields the assertion. □

3.3.3. *The convolution.* The **convolution** $f\star g$ of f and g is defined by

$$(f\star g)(n) := \sum_{\ell=0}^{N-1}\overline{f(\ell)}g(n+\ell).$$

If f is a shift of g – say, $f(m)=g(m+a)$ – then $f\star g$ tends to develop a "peak" at $n=a$ (where f and g match in the sum). The remainder of its behavior is . . . well, convoluted – using one function to smear the other around.

Exercise 3.6. Show that $(\widehat{f\star g})(k) = \overline{\hat{f}(k)}\hat{g}(k)$, i.e. that convolution is the same as Fourier transforming both functions, multiplying them,[13] then Fourier transforming back.

3.3.4. *Differentiation.* In view of (3.2), $f(m)$ can be extended to a continuous function (interpolating the values of f at integers) by

$$f(x) := \frac{1}{N}\sum_{k=0}^{N-1}\hat{f}(k)e^{\frac{2\pi\mathbf{i}kx}{N}}.$$

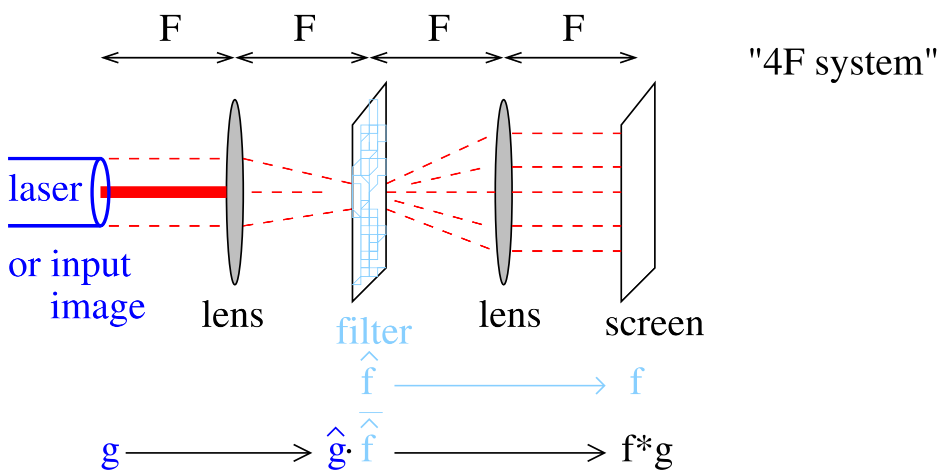
[13] To be precise, multiplying one by the conjugate of the other. Details . . .

Exercise 3.7. Show that "differentiation becomes multiplication" under Fourier transform:

$$f'(x) = \frac{1}{N}\sum_{k=0}^{N-1} \frac{2\pi\mathbf{i}k}{N}\hat{f}(k)e^{\frac{2\pi\mathbf{i}kx}{N}},$$

so that $\frac{2\pi\mathbf{i}k}{N}\hat{f}(k)$ is the k^{th} Fourier coefficient of f'.

3.4. **Applications of Fourier transforms (and series).** Since they decompose a function into constituent frequencies, one expects that they should be very useful for studying waves – or, in engineering jargon, *signal processing*. Computer generated holography is a good example: in the photonics lab of my youth, we set up a *4-F system* (the "F" means focal length) with two identical lenses, a laser, and an LCD filter controlled by the computer:



The idea was to "program the filter with $\hat{f}$", so that the first lens converted the laser beam into a broad sheet of light, which then was converted to $\hat{f}$ by the filter, and Fourier transformed to the desired image f by the remaining lens. Since the filter was pixellated, this was done via the DFT. Unfortunately, a filter just blocks light, and doesn't delay it, so you can only alter the amplitude of the light wave and not its phase; that is, you can't "program the filter with" $\hat{f}$ (a complex-valued function), but only with a positive-valued function. Finding the best such replacement for $\hat{f}$ was part of the object of study.

A 4-F system can also be used for image recognition. If we want to recognize f, then we should construct a filter that "multiplies by $\overline{\hat{f}}$"; in view of §3.3.3, the input g is then converted to $f \star g$ by the system. If the screen records a function with a sharp peak, then we've spotted f; and the location of the peak even guides our "smart" missile toward its target. For while there may be *actually smarter* uses of image recognition than to blow stuff up, the ethos[14] of "defense" as a driving force for STEM seems ineradicable, dating back (at least) to Archimedes's creative use of mathematics to liquidate Roman soldiers.

More peaceable applications to signal processing include the removal of "noise" from old sound recordings and aerial photographs, as well as data compression – which might be achieved, for example, by applying the DFT to 8×8 pixel sets and throwing out the unimportant "frequencies". (This is why your nice digital photos are only 1 megabyte and not 10.) In point of fact, it is the more flexible *wavelet transform* that is preferred for these applications, as well as in CT and MRI scans. The point is that replacing "sine waves" by (translates and dilations of) a "wavelet" tailored to the specific problem (removal of coughs? storage of fingerprints?) yields better results and greater efficiency.

To give an idea of its scope, here are just a few more uses of Fourier series/transforms:

- Solving partial differential equations, like the heat equation (this makes use of the property in §3.3.4);
- Designing earthquake-proof buildings, by making sure that their vibration modes avoid the domain of the Fourier transform of the local seismograph's output;
- Finding the structure of large biochemical molecules (e.g. DNA) by X-ray diffraction;
- Radio reception, sonar, astronomy;
- Music, voice (used in vocal labs), birdsong study;
- Number theory: e.g., the efficient multiplication of large numbers (via a variant of §3.3.3, applied to polynomial multiplication).

In what follows, we'll see how the DFT can be used to evaluate certain infinite series in terms of Bernoulli numbers and powers of π.

3.5. **Infinite inverse power sums.** Recall the Bernoulli numbers

$$B_k = 1, -\frac{1}{2}, \frac{1}{6}, 0, -\frac{1}{30}, \ldots$$

and the generating function

$$\sum_{n\geq 0} B_k \frac{t^k}{k!} = \frac{t}{e^t - 1}. \tag{3.4}$$

[14] or maybe it's just the money

Now by applying L'Hôpital (cf. Remark 2.14) twice, we have

$$\lim_{t\to 0}\left(\frac{1}{e^t-1}-\frac{1}{t}\right)=\lim_{t\to 0}\frac{t-e^t+1}{te^t-t}=\lim_{t\to 0}\frac{1-e^t}{te^t+e^t-1}=\lim_{t\to 0}\frac{-e^t}{2e^t+te^t}$$
$$=-\frac{1}{2}.$$

So near $t=0$, $\frac{1}{e^t-1}$ looks approximately like $\frac{1}{t}-\frac{1}{2}$. Moreover, it has poles whenever $e^t=1$, which happens when t is $2\pi\mathbf{i}$ times an integer. More precisely, using complex analysis, one can show that

$$\frac{1}{e^t-1}=\frac{1}{t}-\frac{1}{2}+\sum_{n\neq 0}\left(\frac{1}{t-2\pi\mathbf{i}n}+\frac{1}{2\pi\mathbf{i}n}\right) \tag{3.5}$$

since the difference of the two sides is bounded and free of poles. Now the sum on the right-hand side of (3.5)

$$=\sum_{n>0}\frac{2t}{t^2-(2\pi\mathbf{i}n)^2}$$
$$=-\sum_{n>0}\frac{2t}{(2\pi\mathbf{i}n)^2}\sum_{k\geq 0}\frac{t^{2k}}{(2\pi\mathbf{i})^{2k}n^{2k}}$$
$$\underset{m=k+1}{=}-\sum_{m\geq 1}\frac{t^{2m-1}}{(2\pi\mathbf{i})^{2m}}\left(2\sum_{n>0}\frac{1}{n^{2m}}\right)$$
$$=-\sum_{m\geq 1}\frac{(-1)^m t^{2m-1}}{(2\pi)^{2m}}2\zeta(2m).$$

Here

$$\zeta(x):=\sum_{n\geq 1}\frac{1}{n^x}$$

is the famous **Riemann zeta-function**, subject of perhaps the most celebrated unsolved problem in number theory.[15] Since the left-hand side of (3.5) is given by (3.4), we conclude that

$$\sum_{n\geq 0}B_k\frac{t^k}{k!}=\frac{1}{t}-\frac{1}{2}-\sum_{m\geq 1}\frac{(-1)^m t^{2m-1}}{(2\pi)^{2m}}2\zeta(2m),$$

which immediately yields

[15]The **Riemann Hypothesis** is one of the six unsolved million-dollar Clay Millenium Problems: in 1859, Riemann conjectured that if we extend ζ to a function of a complex variable $x+iy$, then its zeroes that are not at negative even integers lie on the line $x=\frac{1}{2}$. It has profound implications for the distribution of the prime numbers.

Theorem 3.8. *$B_{2m+1} = 0$ for $m \geq 1$, and*

$$B_{2m} = (-1)^{m-1}(2m)!\frac{2\zeta(2m)}{(2\pi)^{2m}}.$$

So we finally get our *proof* that the odd Bernoulli numbers vanish, and we can also evaluate the "even" Riemann zeta-values in terms of even Bernoulli numbers and powers of π:

$$\zeta(2m) = 1 + \frac{1}{2^{2m}} + \frac{1}{3^{2m}} + \cdots = \frac{(2\pi)^{2m}|B_{2m}|}{2\,(2m)!},$$

as proved by Euler in 1739. The first couple of examples are

$$\zeta(2) = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{(2\pi)^2\frac{1}{6}}{2\cdot 2} = \frac{\pi^2}{6}$$

and

$$\zeta(4) = 1 + \frac{1}{16} + \frac{1}{81} + \frac{1}{256} + \cdots = \frac{(2\pi)^4\frac{1}{30}}{2\cdot 4!} = \frac{\pi^4}{90}.$$

We learn nothing about the odd zeta-values $\zeta(3)$, $\zeta(5)$, and so forth, which are much more mysterious.[16] But perhaps we can shed some light on some series with odd powers in the denominator, like $1 - \frac{1}{3^3} + \frac{1}{5^3} - \frac{1}{7^3} + \cdots$? How?

3.6. **Bernoulli polynomials revisited.** Since the generating function of the Bernoulli numbers was so helpful, what happens if we play the same game with Bernoulli *polynomials*? (Recall that these are given by $B_k(x) := \sum_{j=0}^{k}\binom{k}{j}B_j x^{k-j}$.) We compute

$$\begin{aligned}
\sum_{k\geq 0} B_k(x)\frac{t^{k-1}}{k!} &= \sum_{k\geq 0}\sum_{j=0}^{k}\binom{k}{j}B_j x^{k-j}\frac{t^{k-1}}{k!} \\
&\underset{\ell=k-j}{=} \sum_{\ell\geq 0}\sum_{j\geq 0}\binom{j+\ell}{j}B_j x^{\ell}\frac{t^{j+\ell-1}}{(j+\ell)!} \\
&= \sum_{j\geq 0}\frac{B_j t^{j-1}}{j!}\sum_{\ell\geq 0}\frac{(tx)^{\ell}}{\ell!} \\
&= \frac{e^{tx}}{e^t - 1},
\end{aligned}$$

[16]That $\zeta(3)$ is not a rational number was only discovered (by R. Apéry) in 1978, and its transcendentality is (while expected) not known. We don't yet have a proof that $\zeta(5)$ is irrational! On the other hand, being rational multiples of powers of π, the even zeta values are all transcendental, that is, satisfy no polynomial equation with rational coefficients.

where we have used $\frac{\binom{j+\ell}{j}}{(j+\ell)!} = \frac{1}{j!}\frac{1}{\ell!}$. Again appealing to complex analysis, the last expression

$$
\begin{aligned}
&= \left(\frac{1}{t} - \frac{1}{2} + x\right) + \sum_{n\neq 0}\left(\frac{e^{-2\pi \mathbf{i} nx}}{t+2\pi \mathbf{i} n} - \frac{e^{-2\pi \mathbf{i} nx}}{2\pi \mathbf{i} n}\right) \\
&= \left(\frac{1}{t} - \frac{1}{2} + x\right) + \sum_{n\neq 0}\frac{e^{-2\pi \mathbf{i} nx}}{2\pi \mathbf{i} n}\left(\sum_{\substack{\cancel{j\geq 0} \\ j\geq 1}} \frac{(-1)^j t^j}{(2\pi \mathbf{i} n)^j} \cancel{-1}\right) \\
&\underset{j=k-1}{=} \left(\frac{1}{t} - \frac{1}{2} + x\right) + \sum_{k\geq 2}\frac{(-1)^{k-1}t^{k-1}}{(2\pi \mathbf{i})^k}\sum_{n\neq 0}\frac{e^{-2\pi \mathbf{i} nx}}{n^k}.
\end{aligned}
$$

Comparing coefficients of t^{k-1}, we conclude that

$$B_k(x) = \frac{(-1)^{k-1}k!}{(2\pi \mathbf{i})^k}\sum_{n\neq 0}\frac{e^{2\pi \mathbf{i} nx}}{n^k} \tag{3.6}$$

for all integers $k \geq 2$.

3.7. ***L*-functions.** These generalizations of Riemann's zeta function are very important in number theory. The basic examples are obtained by taking $F \in \Phi(N)$ and defining[17]

$$L(F, x) := \sum_{n>0}\frac{F(n)}{n^x}.$$

Certain of their "special values" $L(F, k)$ at positive integers have a beautiful connection to the DFT and Bernoulli polynomials. For computational convenience, we define "symmetrized" L-values

$$\tilde{L}(F, k) := \sum_{n>0}\frac{F(n) + (-1)^k F(-n)}{n^k} = \sum_{n\neq 0}\frac{F(n)}{n^k}.$$

Now consider $f \in \Phi(N)$. For $\ell \geq 0$, (3.6) yields

$$
\begin{aligned}
\sum_{a=0}^{N-1} f(a) B_{\ell+2}\left(\tfrac{a}{N}\right) &= \frac{(-1)^{\ell+1}(\ell+2)!}{(2\pi \mathbf{i})^{\ell+2}}\sum_{a=0}^{N-1} f(a)\sum_{m\neq 0}\frac{e^{-2\pi \mathbf{i} ma/N}}{m^{\ell+2}} \\
&= \frac{(-1)^{\ell+2}(\ell+2)!}{(2\pi \mathbf{i})^{\ell+2}}\sum_{m\neq 0}\frac{1}{m^{\ell+2}}\underbrace{\sum_{a=0}^{N-1} f(a) e^{-2\pi \mathbf{i} ma/N}}_{\hat{f}(m)}
\end{aligned}
$$

hence

[17] This gives $\zeta(x)$ if F is identically 1.

Theorem 3.9. *The symmetrized L-values are given by*

$$\tilde{L}(\hat{f}, \ell+2) = \frac{(2\pi \mathbf{i})^{\ell+2}}{(-1)^{\ell+2}(\ell+2)!} \sum_{a=0}^{N-1} f(a) B_{\ell+2}\left(\tfrac{a}{N}\right).$$

In particular, if f is $\mathbb{Q}$-valued, then (although $\hat{f}$ usually will *not* be) $\tilde{L}(\hat{f}, \ell+2) \in (2\pi i)^{\ell+2}\mathbb{Q}$.

Exercise 3.10. Take $N = 4$, and $f(0) = 0$, $f(1) = 1$, $f(2) = 0$, $f(3) = -1$.

(a) What is $\hat{f}$?
(b) Write $\tilde{L}(\hat{f}, 3)$ as a sum over $m \geq 0$.
(c) Evaluate the sum using the Theorem.
(d) Conclude that $\sum_{m\geq 0} \frac{(-1)^m}{(2m+1)^3} = \frac{\pi^3}{32}$.

So what about Catalan's constant $G := \sum_{m\geq 0} \frac{(-1)^m}{(2m+1)^2}$? or $\sum_{m\geq 1} \frac{1}{m^3}$ ($= \zeta(3)$)? The problem is that the "normalization" $\tilde{L}$ in both cases is zero, due to cancellation of $F(n)$ and $(-1)^k F(-n)$. So the Theorem does nothing to demystify these numbers; and indeed, we don't know whether G is irrational.

Our Theorem 3.9 is the "one-dimensional" analogue of a result for functions on $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, which allows one to define something called the *horospherical map*. With a significant amount of work, it led to the proof of an analogue of the Hodge Conjecture for modular curves by A. Beilinson, one of the great achievements in arithmetic geometry of the last half-century.

4. Further reading

I conclude with a short list of articles and monographs which further develop several of the themes in this essay. The combinatorial angle is represented by Stanley's survey article [3] and the paper [1], in which you will find many other interpretations (topological, probabilistic, geometric) and generalizations of updown numbers, as well as ample references. If you would like to learn more about Fourier series and transforms, the first volume [4] of Stein and Shakarchi's "Lectures in Analysis" assumes only an acquaintance with the Riemann integral from high-school calculus; to get an idea of how wavelets are different, see Strang's short note [6]. As for number theory and Riemann's hypothesis in particular, look no further than the beautiful book [2]. Finally, if you would like to learn more about the lives of mathematicians like Euler, Fourier, Gauss and Riemann, pick up the incomparable Ian Stewart's new book [5].

References

[1] F. Brown, T. Fink, and K. Willbrand, *Arithmetic and asymptotic properties of up-down numbers*, Discrete Mathematics 307 (2007), 1722-1736, available at http://arxiv.org/pdf/math/0607763.pdf.

[2] B. Mazur and W. Stein, "Prime numbers and the Riemann hypothesis", Cambridge University Press, 2016.

[3] R. Stanley, *A survey of alternating permutations*, Contemporary Mathematics 531 (2010), 165-196, available at http://www-math.mit.edu/~rstan/papers.altperm.pdf.

[4] E. Stein and R. Shakarchi, "Fourier analysis: an introduction", Princeton University Press, 2003.

[5] I. Stewart, "Significant figures: the lives and work of great mathematicians", Basic Books, 2017.

[6] G. Strang, *Wavelets*, American Scientist 82 (1994), 250-255, available at http://www.jstor.org/stable/pdf/29775194.pdf.

Department of Mathematics, Campus Box 1146
Washington University in St. Louis
St. Louis, MO 63130, USA

e-mail: matkerr@math.wustl.edu