Quadratic Reciprocity and the Langlands Program: the pictorial way through Hidden Symmetries of numbers

Reconnaissance-in-force — with 555 anecdotes at vistapoint reststops

Ilya Zakharevich

Email: math@ilyaz.org

> ... the history of the theory of numbers [...] is dominated by the law of reciprocity.
>
> Letter from Andre Weil to Simone Weil (Bonne-Nouvelle military prison, Rouen, March 1940)

> When I discovered that the sine can be expressed algebraically as a series, a barrier came tumbling down, and mathematics became one. To this day I see the various branches of mathematics, together with mathematical physics, as a unified whole.
>
> I. M. Gelfand, Interview with *Quantum* (Jan–Feb 1991)

These notes grew up from a brief discussion about the Langlands program we had at our Math Circles (in sections for grades 3–4). The format of our Math Circles includes detailed reports about every meeting sent to the parents (who are assumed to discuss them with the kids). At the meeting in question, we were intentionally vague about most of technical details, painting only the rudimentary outline in very coarse strokes. However, it turned out that to make a meaningful written exposition, we needed to fill these holes in the report. This resulted in a huge appendix[1] to the report to the parents; it became the bulk of these notes.

All that these notes require from the reader is a fluent working knowledge of "engineering-grade math" — and a lot of stamina. We also had an ulterior motive: the way we wrote the "Langlands part" puts it out of reach for all but a handful of most advanced high-schoolers. So we hope that these notes demonstrate how accessible this beautiful landscape turns out to be, and that this may inspire one of the readers to find further simplification which would allow detailed discussions of the Langlands Program in Math Circles for high-schoolers.

Moreover, we already know how to discuss the first segment of these notes (one dedicated to Quadratic Reciprocity) at Math Circles. To reflect this, certain parts of this segment are written in a particular form to match what we did with kids in our circles (Grades 1–4). We put such parts between the signs Ⓜ Ⓜ̸.

Essentially, our aim is to expose a few simplest cases (of those not covered by Class Field Theory) for which the Langlands program *works as a bridge* between two "almost" completely elementary contexts — "almost" since one of them needs Fourier series...[2] (In general, the Langlands program needs to be stated in terms of two "representations" — and both are quite high-brow topics. In contrast, we do not mention representations until almost the end of these notes.)

In the appendix, we highlight features of Euler's approach to quadratic reciprocity. This approach ("look for symmetries") is more suitable to generalizations than Legendre's approach (the "reciprocity"). (However, it is the Legendre's approach which the "popular math" movement made better known.)

---

[1] We hoped to make it short, and the first versions were — but they turned out to be unreadable.

[2] Only a very minimal knowledge of Fourier series is required. In this context, the most important feature is the Fourier transform being a *bijection* between sequences and periodic functions — so it *recodes* the information contained in the series into a function (and back).

It is also useful to understand that the rate of decay of Fourier coefficients corresponds to the smoothness of the sum of Fourier series. (In particular, taking derivative — which makes a C^k-function "less smooth" — corresponds to multiplication by n — which makes the Fourier coefficients decay slower. Likewise for integration: it makes a function "smoother", and makes the Fourier coefficients to decay quicker.)

Finally, it may help to know some particular cases of the preceding connection. In particular, if coefficents are in ℓ_1 (hence "do not decay too slow") then the sum of the series is a continuous function. (In the opposite direction one can get only a much weaker estimate: continuity implies that the coefficients are bounded.[3])

[3] Recall that it is possible to get a 1-to-1 match between "degrees of the growth of coefficients" and "degrees of smoothness of the sum" — but one needs a bit more complicated gauges of these degrees, such as *the Sobolev classes*

In the electronic copy there is a lot of clickable crossreferences and links to Web resources.[4]
The plots there allow a deep zooming in.

Contents

[4] The paper copy has them dot-underlined, as above.

Digest: Meetings on Quadratic Reciprocity etc. (Grades 1–4, Ilya 2018-05): the "hidden symmetries" in $\deg = 2$ are periodicity and mirroring

This chapter is very special: these (very mathematically rich and important for contemporary math!) themes can be discussed — on a rather deep level — with a first-grader who is well advanced in math; nevertheless, even many working mathematicians have a rather skewed impression what the quadratic reciprocity "is really about". As a workaround, here we try to intertwine three different stories:

- What are the "hidden symmetries" in the simplest non-trivial cases;
- How a not-very-experienced-in-math person can "discover" these symmetries (and how to teach this in Math Circles; these parts are bracketed by Ⓜ/~~Ⓜ~~ signs);
- Remarks for experienced mathematicians on the general framework (mostly in the last two sections).

This chapter is based on a digest of what we did in our Math Circles. An experienced reader who is not interested in teaching does not need to read this meticulously — especially the parts bracketed with Ⓜ/~~Ⓜ~~. In fact, a *very* experienced reader may jump to the last two sections (p. 14) immediately.

Divisors of polynomial sequences: the simplest cases

In these notes we start with a given polynomial sequence P_m — which is the sequence of values of a polynomial P. (The group 0 of exercises on p. 22 covers our notations and the most elementary properties of such sequences. Unless the reader is fully fluent with such sequences, we strongly recommend going through these exercises now.) We look for numbers which divide one of P_m.

In other words:

$$\boxed{\text{For every number } n\text{, we ask: does it divide one of the numbers } P_m?}$$

The answer is a function of n with values **Yes** or **No**. (For pedants: above, "one" means "one or more".) We call the numbers with the answer **Yes** ***the divisors of a sequence***.

This may be restated as describing modular arithmetics in which a given polynomial equation[5] $P(x) = 0$ has a solution. Ⓜ However, our particular formulation allows us to introduce this problem to the kids quickly — all that is needed is a rudimentary knowledge of multiplication.[6] ~~Ⓜ~~ An impatient reader may want to skip the examples and jump to the section on p. 15.

Start with the cases of very small degree of P. The first two are completely trivial:

- If $P \equiv 0$, then for every number n the answer is **Yes**.
- If $P \not\equiv 0$ and $\deg P = 0$, then excluding finitely many numbers n, the answer is **No**.
- If $\deg P = 1$ and $P_m = am + b$, then for every n mutually prime with a the answer is **Yes**.[7]

[5] … or, in many applications, a system of polynomial equations.

[6] Ⓜ We introduce polynomial sequences by investigating the differences of differences of differences (etc.) and eventually finding a sequence of 0s.

[7] Ⓜ For example, for $n = 10$, if a is odd, then the last digit of P_m would go through all the digits (unless the last digit of a is 5). In particular, 0 appears as the last digit of P_m.

(We discuss the details of the last case in the group A of exercises on p. 23.)

One can "fix" the last statement so that it is more similar to the preceding ones:

If $\deg P = 1$, then excluding finitely many numbers p, every prime number p divides one of the values of P_m.

As this shows, even in the simplest cases, allowing a finite number of exceptions leads to significant simplifications of the statements. Moreover, restricting attention to prime divisors may lead to further similar simplifications. For example, one can cover all the cases as in:

If $\deg P \leq 1$, then excluding finitely many numbers p, the answer to "Is a prime number p a divisor of one of the values of P_m?" does not depend on p.

These two ways to achieve simplifications are going to influence our formulations of similar statements for higher degrees as well.

Example in $\deg = 2$: pizza numbers

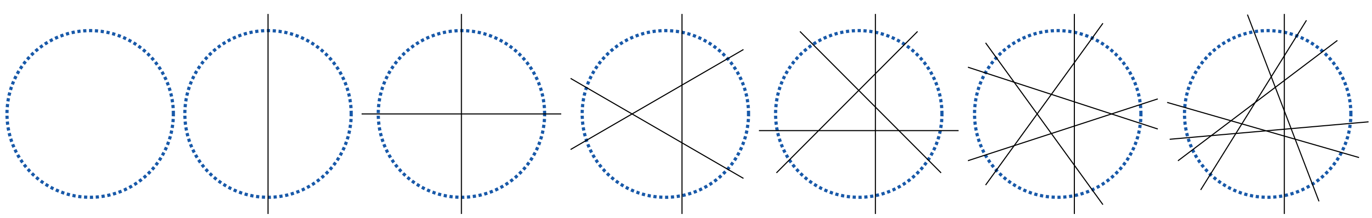
When I look at the history of mathematics, I see a succession of illogical jumps, improbable coincidences, jokes of nature.

Freeman Dyson, *Birds and Frogs*, Notices of the AMS (Feb 2009)

What we are going to discuss here is one of the most dramatic discoveries in arithmetic. The typical expositions try to play this dramactic aspect down; while we cannot have the Chorus singing "Something is going to happen! Just you wight, 'enry 'iggins, just you wight!", we want to start with a sequence P_m having a clear combinatorial significance, and check what are the divisors of the numbers P_m.[8]

Remark 1: A very adventurous reader may want to jump to the group C of exercises on p. 25. If one wants to do this group of exercises later, what we discuss here may work as hints for them.

Observe the largest number of pieces of pizza one can make with m straight cuts:



The answers join into the following table, and differences follow a simple pattern:

Cuts	0	1	2	3	4	5	6	7	8	9	10	11	12
Triangular number	0	1	3	6	10	15	21	28	36	45	55	66	78
Pieces	1	2	4	7	11	16	22	29	37	46	56	67	79

+1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12

Observing this pattern leads to an immediate description of pizza numbers P_m: they are 1 more than triangular numbers.

[8] This question may look like a joke from the epigraph to this section: *why* would one investigate divisors of the number of (so different!) pieces of pizza? The only excuse we can claim is that the answer *would* demonstrate the "improbable coincidences" from the same epigraph — by being completely unexpected, and eventually leading to a progression of interrelations between completely separate branches of math. And, as Dyson said in the same paper, "the deepest concepts in mathematics are those which link one world of ideas with another."

Indeed: We need to show that the next cut can increase the count of pieces by at most the green number (as above). The increase is the number of "old" pieces this cut goes through. Observe that the new cut is subdivided into the cuts made through these pieces. Moreover, these parts of the cut are separated by the points where the new cut meets the old cuts.

Hence to show that the pattern observed above continues forever, it is enough to show that the mth cut meets the old cuts in at most $m-1$ points. — However, there are only $m-1$ old cuts!

(In fact, one also needs a bound on the other side: that $m-1$ meeting points *is* possible. However, this is completely obvious after one notices that the answer for pizza is exactly the same as the answer for the whole plane. Indeed, one can shrink the cut lines until all the meeting points fit inside the pizza.)

The pattern above shows that $P_m - P_{m-1} = m-1$, which leads to the formula $P_m = m(m+1)/2+1$. Therefore P is a polynomial of degree 2.

Ask the same question as above: what are the possible divisors of (one of) the numbers P_m?[9] (Here we are interested in all divisors, not only the prime ones.) With the following table

Side	1	2	3	4	5	6	7	8	9	10	11
$\triangle$-number + 1	2	4	7	11	16	22	29	37	46	56	67
As products	1×2	1×4 2×2	1×7	1×11	1×16 2×8 4×4	1×22 2×11	1×29	1×37	1×46 2×23	1×56 2×28 4×14 7×8	1×67

one can see that 1, 2, 4, 7, 8, 11, 14, 16, 22, 23, 28, and 29 can be divisors of "pizza numbers".

Observation: the same table shows that no other number up to 30 can divide a pizza number! (We recommend solving Exercise C16 on p. 26 and the following exercises now. The group B of exercises on p. 24 may also be of help here.)

Indeed, consider pizza numbers mod n. If n is odd, then division by 2 in the above formula for pizza numbers makes sense mod n, hence pizza numbers mod n are periodic with period of length n; for even n, a similar argument shows that there is a period of length $2n$.

Moreover, if we continue pizza numbers to the left,[10] they form a palindromic sequence: $P_{-1-m} = P_m$. Hence if numbers $P_0, P_1, \dots, P_l$ are not divisible by n, then numbers $P_{-1}, P_{-2}, \dots, P_{-1-l}$ are also not divisible by n. If $2+2l$ is at least as long as the period of $P_m \bmod n$, we can also conclude that no number P_m would be divisible by n.

Conclusion: For $n = 2m+1$, it is enough to check that n does not divide $P_1, \dots, P_m$, and then n cannot divide any pizza numbers. Likewise for even $n = m+1$.

Looking in the list above, this means that if $n \leq 30$ is not in the list, and divides one of pizza numbers, then $n > 23$ for odd n, and $n > 12$ for even n. In particular, the list above is good up to $n = 17$.

In the odd case only 25 and 27 remain — and they cannot be divisors, since we *already know* that 3 and 5 cannot be divisors! In the even case we know that the answer about $n = 2l$ is **No** if it *is already known* that l cannot divide pizza numbers; one can see that this implies indeed that the list above is complete up to $n = 30$.

Conclusion: In the list

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 . . .

[9] Recall that mathematically, we want to describe the modular arithmetics for which the polynomial equation $P_x = 0$ has roots.

[10] This is one of the favorite subjects in our Math Circles. Note that it is trivial to continue the row of green numbers (in the first table of this section) to the left. After this, it is easy to continue the row of pizza numbers to the left so that the green numbers continue to work as differences.

the green numbers are divisors of pizza numbers, and red numbers are not divisors of pizza numbers.

This distribution of colors looks completely random. However, already Euler and Legendre knew how to find the pattern in this distribution of colors. (Moreover, Fermat might have known the answer too: he found patterns for several other polynomials of degree 2. In fact, he could *prove* that these patterns would continue forever for *all* similar sequences simpler[11] than this one.)

It turns out that a noticable proportion of people cannot see the pattern in the table below.[12]

Answer: to see the pattern, we need to highlight prime numbers, and rewrite the sequence of natural numbers using 7 columns (on the right). It is clear that bold numbers in certain columns are all green, and in the remaining columns they are all red.

Moreover, although the columns of 3, 5 and 6 are fully red, the columns of 1, 2 and 4 contain a mix of red and green. This means that the pattern, indeed, does not work for composite numbers. (For example, 50 and 64 are composites which are in the same column: the column of 1.)

Of course, every column on the right represents a residue modulo 7. Hence the pattern above shows that to find whether a prime number p can divide a pizza number, it is enough to know $p \bmod 7$. Yet another way to state it is that (slightly abusing notation):

1	**2**	**3**	4	**5**	6	**7**
8	9	10	**11**	12	**13**	14
15	16	**17**	18	**19**	20	21
22	**23**	24	25	26	27	28
29	30	**31**	32	33	34	35
36	**37**	38	39	40	**41**	42
43	44	45	46	**47**	48	49
50	51	52	**53**	54	55	56
57	58	**59**	60	**61**	62	63
64	65	66	**67**	68	69	70
71	72	**73**	74	75	76	77
78	**79**	80	81	82	**83**	84
85	86	87	88	**89**	90	91
92	93	94	95	96	**97**	98
99	100	**101**	102	**103**	104	105
106	**107**	108	**109**	110	111	112
113	114	115	116	117	118	119
120	121	122	123	124	125	126
127	128	129	130	**131**	132	133
134	135	136	**137**	138	**139**	140
141	142	143	144	145	146	147
148	**149**	150	**151**	152	153	154
155	156	**157**	158	159	160	161
162	**163**	164	165	166	**167**	168
169	170	171	172	**173**	174	175
176	177	178	**179**	180	**181**	182
183	184	185	186	187	188	189
190	**191**	192	**193**	194	195	196
197	198	**199**	200			

The pattern of colors is periodic when restricted to prime numbers.

What does it mean for a function of a prime number to be periodic?!
The pattern above suggests the answer:[13] a function $f(p)$ is periodic if there exists a periodic function $F(n)$ on $\mathbb{N}$ such that f is a restriction of F. The function F is in fact uniquely defined on n mutually prime with the length of its period. (This is due to existence of prime numbers in any arithmetic progression with mutually prime step and values.)

One can illustrate this pictorially. Observe the colored sequence above; we copy it below, and, in the next row, we write the sequence of colors with the period[14] ●●●●●●● (of length 7):

1 **2** **3** 4 **5** 6 **7** 8 9 10 **11** 12 **13** 14 15 16 **17** 18 **19** 20 21 22 **23** 24 25 26 27 28 **29** 30 **31** 32 33 . . .
1 **2** **3** 4 **5** 6 **7** 8 9 10 **11** 12 **13** 14 15 16 **17** 18 **19** 20 21 22 **23** 24 25 26 27 28 **29** 30 **31** 32 33 . . .

As you can see, these sequences match at prime numbers (marked bold) but not at 9, 15, 18, . . . ![15]

[11] Here the measure of simplicity is the number of necessary columns in the tables below.

[12] M Fortunately, for the kids the proportion is quite similar to one for adults; so recognizing this pattern is a reasonably challenging problem to give at a math circle.

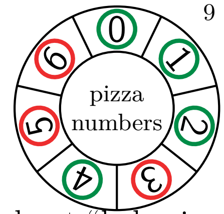
[13] Another way to see this is that “usually” N-periodicity of T_n means $T_n = T_{n+N}$ for any n. But this is equivalent to $T_n = T_{n'}$ if N divides $n' - n$.

The latter way makes perfect sense even if T_n is defined only for n in a subset of $\mathbb{N}$. Hence it makes sense to say that “the given function $f(p)$ of a prime number p is periodic”. (As usual, a *periodic* function is one which is N-periodic for some $N > 0$.)

Finally, for us “*a period*” is *the part* of a sequence which repeats forever. For example, N above is not a period, but a “length of a period”.

[14] Recall that for us, “a period” is a subsequence, as opposed to its length.

[15] This is a very general observation about how patterns involving conductors behave: given a sequence, we provide another sequence (defined by completely unrelated rules!) which matches the former sequence at prime numbers. However, in general there may be a few “exceptional primes” where the match breaks. Observe $\cancel{2}$ below, on p. 13.



Another way to restate this is to observe what happens if we *ignore all the non-bold (non-prime) numbers*. Since every column in the table above matches a particular position on the wheel of residues mod 7, we may color these position matching the color of bold numbers in the columns (on the right).

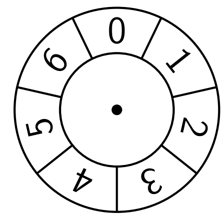
One can call the wheel on the right "the *conductor* wheel": to find out something about "behavior of pizza numbers mod p" ("is there a pizza number which is 0 mod p"), it is enough to know p mod 7 (provided p is prime). So we say that

The conductor for the problem of divisors of pizza numbers is 7.

Summing up the same way as on p. 9:

The answer to "Is a prime number p a divisor of one of pizza numbers?" depends only on $p \bmod 7$.

Wheels



The group A of exercises on p. 23 covers the usage of wheels ("clocks") to handle the behaviour of divisors of terms of arithmetic progressions. (It expands on our Observation on p. 7.)

Warning: the wheels drawn at the end of the preceding section (and in what follows), in discussions of sequences of degree 2, are of *very* different nature. The rest of this section is, essentially, a sneak summary of the architecture of such discussions. It is written very cursorily, and it may be wise to skip it at the first reading.

Recall that when dealing with arithmetic progressions, we take a particular number n (a "potential divisor of the sequence"); to answer the question: "is it a divisor of a number in our sequence?", we use the wheel of size n. The positions on this wheel are residues mod n, and we consider residues of the numbers in our sequence P_m. If one of these is 0 mod n, then the answer for the number n is **Yes**.

This may be summarized as the first row in the table below. On the other hand, for P of degree 2, we work with "the *conductor* wheels". There is one[16] such wheel per sequence P_m, its size is called the *conductor* c of the sequence (it is related[17] to the discriminant of P).

	Size	Take positions of:	Look at:	$\deg P$
n-wheel	n	Numbers P_m	Reaching the position 0 mod n	1
Conductor wheel	The conductor c	Potential divisors n	The color of a position	2

Every position on the conductor wheel is marked with a color, and there is a certain correlation between the *the answer* **Yes/No** for n (discussed in the section on p. 5) and *the color* at the position of n. In fact, there is 1-to-1 match color↔answer for prime numbers n.

In short, instead of inspecting in the linear case whether P_m on the n-wheel may hit the 0-position (for at least one m) — or $P_m \bmod n = 0$, in the case $\deg P = 2$ what we inspect is the color of $n \bmod c$ on the conductor wheel.

Essentially, due to 1-to-1 matching mentioned above this may work only if the conductor c is the length of the period of the function in:

If $\deg P \leq 2$, then the answer to "Is a prime number p a divisor of one of the values of P_m?" is a periodic function of p.

[16] Unless we are going to distinguish the "conductor" and the "level" so we get two wheels, as we do on p. 11. Compare with Footnote 22 on p. 11.

[17] While to shows the *existence* of the conductor is a very hard problem, there is a simple recipe *calculating* the conductor for a quadratic P. On the other hand, while analogues of conductors exist in higher degrees (later, we discuss mostly degree 3), the calculation of these conductors may be very involved.

The preceding section and several following sections provide examples clarifying this claim. ("Periodicity" is discussed in Footnote 13 on p. 8.)

Remark 2: Compare this statement in the case $\deg P \leq 1$ with the results of the section on p. 5. How come there is no need to exclude a finite number of exceptional values of p in the statement above? We handle this question in Exercise A4 on p. 24.

Remark 3: However, if we allow a finite number of exceptions, then the laws of the section on p. 5 show that for $\deg P \leq 1$ the periodic function may be taken constant (in other words: the conductor may be taken to be $c = 1$).

Likewise, if we allow a finite number of exceptions for $\deg = 2$, for many sequences the length of the period in the law above may be decreased.[18] However, for irreducible polynomials of $\deg = 2$, the "reduced" conductor is always greater than 1.

Conductor of another sequence of degree 2: "squares + 3"

It turns out that the pattern of colors we observed for divisors of pizza numbers is applicable to all polynomial sequences of degree 2. (Instead of reading through this section, the reader may want to solve the first two exercises in the group C of exercises on p. 25.)

Remark 4: In fact, some of polynomial sequences give easier answer than the others of the same degree. For example, if the polynomial has a factor of degree 1, then we get the same answer as for sequences of degree 1 (see p. 5).

Recall that for pizza numbers, after ignoring non-prime numbers, the red/green color pattern is fully controled by the residue of the (prime) number mod 7. Compare this with the coloring of positions on 7-wheel above.

The simplest similar answer is for the sequence $n^2 + 3$. Exercises mentioned above provide the handout with a table representing the numbers in this sequence as products (similar to the table on p. 7) up to $n = 60$; so one can see that the divisors of numbers $n^2 + 3$ with $n \leq 60$ are

1 2 3 4 6 7 12 13 14 19 21 26 28 31 37 38 39 42 43 49 52 57.

Moreover, using the same arguments as for pizza numbers, one can show that no other number up to 60 can divide numbers $n^2 + 3$. This gives the green/red coloring for divisors/non-divisors up to 60 (not all of which fits below):

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 . . .

One can see that to work with more data, it makes sense to omit the red numbers.

Recall that the preceding section suggests existence of a certain pattern:

- Select the prime numbers from the first list (the list of divisors).
- Choose *an approriate* size of a wheel (*the conductor*), and write numbers 1,. . . ,60 in that many columns.[19]
- Mark the prime numbers in these columns.
- Select *suitable* columns out of these tables.
- Then the prime numbers from the list above would coincide with prime numbers in the selected columns.

(Recall also that one may expect several mismatches — but there should be very few of these.)

Of course, the real challenge is to choose what stands for "*appropriate*" in this recipe!

[18] In other words: there is a sequence with a shorter period which *also* matches the answers **Yes** and **No** for primes p if we allow a finite number of exceptions. (Compare with Footnote 22 on p. 11.)

[19] Now each column matches one of the positions on the wheel.

The prime numbers in the "green" list above are:

1 **2** **3** 4 6 **7** 12 **13** 14 **19** 21 26 28 **31** **37** 38 39 42 **43** 49 52 57 ...

giving a rather short (so simple to work with) list of prime divisors 2 3 7 13 19 31 37 43 ... below 60. (Circling in green — instead of using green color — is easier to do by hand[20] on printouts of numbers arranged in columns.) What remains is to visualize this list using arrangements of natural numbers into several columns.

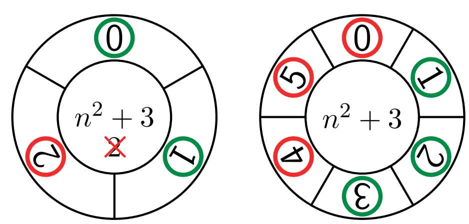
1	**2**	**3**
4	**5**	6
7	8	9
10	**11**	12
13	14	15
16	**17**	18
19	20	21
22	**23**	24
25	26	27
28	**29**	30
31	32	33
34	35	36
37	38	39
40	**41**	42
43	44	45
46	**47**	48
49	50	51
52	**53**	54
55	56	57
58	**59**	60

1	**2**	**3**	4
5	6	**7**	8
9	10	**11**	12
13	14	15	16
17	18	**19**	20
21	22	**23**	24
25	26	27	28
29	30	**31**	32
33	34	35	36
37	38	39	40
41	42	**43**	44
45	46	**47**	48
49	50	51	52
53	54	55	56
57	58	**59**	60

1	**2**	**3**	4	**5**
6	**7**	8	9	10
11	12	**13**	14	15
16	**17**	18	**19**	20
21	22	**23**	24	25
26	27	28	**29**	30
31	32	33	34	35
36	**37**	38	39	40
41	42	**43**	44	45
46	**47**	48	49	50
51	52	**53**	54	55
56	57	58	**59**	60

1	**2**	**3**	4	**5**	6
7	8	9	10	**11**	12
13	14	15	16	**17**	18
19	20	21	22	**23**	24
25	26	27	28	**29**	30
31	32	33	34	35	36
37	38	39	40	**41**	42
43	44	45	46	**47**	48
49	50	51	52	**53**	54
55	56	57	58	**59**	60

One can immediately see that only the table with 6 columns matches: the bold numbers in its columns 1, 2, 3 coincide with the list above!



In fact, there is another, smaller table which also matches — but only if we allow "a few" exceptions for the match. In the table with 3 columns the first and the last columns match all the prime numbers in the list above — with one exception ~~2~~.

This leads to the conductor wheels on the right.

Conclusion: based on the provided data, the conductor for divisors of number$^2 + 3$ is 6, and if we allow one exception, conductor 3 "would also work".[21] [22]

---

[20] M ... which allows conversion of what we do in this section into a rather simple exercise. We give the students the list of "green" primes, and the printouts of numbers arranged into columns (with nothing circled yet). The kids circle the numbers from the list and can find which of the arrangements lead to "observable patterns".

[21] We already saw (on p. 6) that it may be very convenient to allow a finite number of exceptions. Note that the lists we considered are just what is near the start of infinite lists. Comparing with these infinite lists, any finite number of exceptions is "relatively negligible".

[22] In some contexts the smaller number is called "conductor", and the larger one "*level*". Also, see p. 213 and Footnote 18 on p. 10.

Remark 5: This guess turns out to be correct. *The law of quadratic reciprocity* for n^2+3 shows that if we continue the tables with 3 or 6 columns, new circled numbers would appear only in the left column, and all the bold numbers in this column are going to be circled. We return to this in the section on p. 15.

Divisors of sequences of deg = 2: two more cases

(Instead of reading through this section, the reader may want to solve Exercise C3 on p. 25.)

Start with “squares + 1”. Same as in preceding section, one can use handouts from Exercise C3 on p. 25 presenting such numbers as products (up to 60^2+1). This gives the list of divisors up to 60

1 2 5 10 13 17 25 26 29 34 37 41 50 53 58 ...

The prime numbers in the list above are:

1 **2** **5** 10 **13** **17** 25 26 **29** 34 **37** **41** 50 **53** 58 ...

hence for “squares + 1” we should circle prime numbers in 2 5 13 17 29 37 41 53 .... Clearly, 3 columns do not work: the circled numbers are scattered among the first 2 columns — and these columns contain a lot of primes not in our list:

1	**2**	**3**
4	**5**	6
7	8	9
10	**11**	12
13	14	15
16	**17**	18
19	20	21
22	**23**	24
25	26	27
28	**29**	30
31	32	33
34	35	36
37	38	39
40	**41**	42
43	44	45
46	**47**	48
49	50	51
52	**53**	54
55	56	57
58	**59**	60

1	**2**	**3**	4
5	6	**7**	8
9	10	**11**	12
13	14	15	16
17	18	**19**	20
21	22	**23**	24
25	26	27	28
29	30	**31**	32
33	34	35	36
37	38	39	40
41	42	**43**	44
45	46	**47**	48
49	50	51	52
53	54	55	56
57	58	**59**	60

1	**2**	**3**	4	**5**
6	**7**	8	9	10
11	12	**13**	14	15
16	**17**	18	**19**	20
21	22	**23**	24	25
26	27	28	**29**	30
31	32	33	34	35
36	**37**	38	39	40
41	42	**43**	44	45
46	**47**	48	49	50
51	52	**53**	54	55
56	57	58	**59**	60

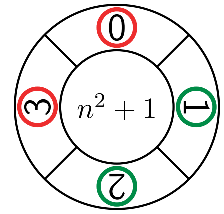
1	**2**	**3**	4	**5**	6
7	8	9	10	**11**	12
13	14	15	16	**17**	18
19	20	21	22	**23**	24
25	26	27	28	**29**	30
31	32	33	34	35	36
37	38	39	40	**41**	42
43	44	45	46	**47**	48
49	50	51	52	**53**	54
55	56	57	58	**59**	60

Likewise, the tables with 5 or 6 columns do not work — by the same reasons. However, the table with 4 columns looks absolutely different: if a column contains a circled number, then *all* the bold numbers in this column are circled.[23] (Observe that 2 is in the “exceptional” column — containing only 1 prime number — so for this column the description in the preceding sentence is void.)

[23] However, these observations are just guesses: they do not show that the observed patterns would continue forever.

Answer: the reasonable guess for the conductor is 4. (And, in fact, this is the correct answer: if we continue the table with 4 columns, the circled numbers would appear only in the left column, and all the bold numbers in the left column are going to be circled.)

Hence the conductor wheel looks like this:



Compare this with the answer from the preceding section (for "squares + 3"): the list of circled primes we obtained there

$$\cancel{2},\ 3,\ 7,\ 13,\ 19,\ 31,\ 37,\ 43,$$

matched the left and the right columns of the table with 3 columns — but to match, we needed to exclude ~~2~~. With "squares + 1", we do not need to exclude anything.

The next example is "squares − 2". (M Same handouts as in the beginning of the section give factorizations, up to $60^2 - 2$. ~~M~~) This gives the list of divisors up to 60 (with primes in bold):

1 **2** **7** 14 **17** **23** **31** 34 **41** 46 **47** 49 ...

which gives the list of primes to circle: 2 7 17 23 31 41 47 ....

With small number of columns, this leads to

1	**2**	**3**
4	**5**	6
7	8	9
10	**11**	12
13	14	15
16	**17**	18
19	20	21
22	**23**	24
25	26	27
28	**29**	30
31	32	33
34	35	36
37	38	39
40	**41**	42
43	44	45
46	**47**	48
49	50	51
52	**53**	54
55	56	57
58	**59**	60

1	**2**	**3**	4
5	6	**7**	8
9	10	**11**	12
13	14	15	16
17	18	**19**	20
21	22	**23**	24
25	26	27	28
29	30	**31**	32
33	34	35	36
37	38	39	40
41	42	**43**	44
45	46	**47**	48
49	50	51	52
53	54	55	56
57	58	**59**	60

1	**2**	**3**	4	**5**
6	**7**	8	9	10
11	12	**13**	14	15
16	**17**	18	**19**	20
21	22	**23**	24	25
26	27	28	**29**	30
31	32	33	34	35
36	**37**	38	39	40
41	42	**43**	44	45
46	**47**	48	49	50
51	52	**53**	54	55
56	57	58	**59**	60

1	**2**	**3**	4	**5**	6
7	8	9	10	**11**	12
13	14	15	16	**17**	18
19	20	21	22	**23**	24
25	26	27	28	**29**	30
31	32	33	34	35	36
37	38	39	40	**41**	42
43	44	45	46	**47**	48
49	50	51	52	**53**	54
55	56	57	58	**59**	60

All these tables do not look like what we want! With more columns, this looks like that:

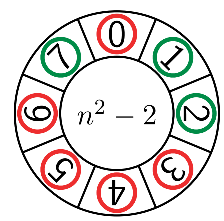
1	**2**	**3**	4	**5**	6	**7**
8	9	10	**11**	12	**13**	14
15	16	**17**	18	**19**	20	21
22	**23**	24	25	26	27	28
29	30	**31**	32	33	34	35
36	**37**	38	39	40	**41**	42
43	44	45	46	**47**	48	49
50	51	52	**53**	54	55	56
57	58	**59**	60	**61**	62	63

1	**2**	**3**	4	**5**	6	**7**	8
9	10	**11**	12	**13**	14	15	16
17	18	**19**	20	21	22	**23**	24
25	26	27	28	**29**	30	**31**	32
33	34	35	36	**37**	38	39	40
41	42	**43**	44	45	46	**47**	48
49	50	51	52	**53**	54	55	56
57	58	**59**	60	**61**	62	63	64

and

1	**2**	**3**	4	**5**	6	**7**	8	9
10	**11**	12	**13**	14	15	16	**17**	18
19	20	21	22	**23**	24	25	26	27
28	**29**	30	**31**	32	33	34	35	36
37	38	39	40	**41**	42	**43**	44	45
46	**47**	48	49	50	51	52	**53**	54
55	56	57	58	**59**	60	**61**	62	63

1	**2**	**3**	4	**5**	6	**7**	8	9	10
11	12	**13**	14	15	16	**17**	18	**19**	20
21	22	**23**	24	25	26	27	28	**29**	30
31	32	33	34	35	36	**37**	38	39	40
41	42	**43**	44	45	46	**47**	48	49	50
51	52	**53**	54	55	56	57	58	**59**	60



For 3, 4, 5, 6, 7, 9 and 10 columns, we see that a lot of columns contain both circled and non-circled prime numbers. This is not what we want to see.[24] On the other hand, with 8 columns we see exactly the same pattern as expected (and we do not even need "a small number of exceptions"): columns of 1, 2 and 7 contain *only* circled numbers, and there is no circled number in other columns. So, basing on the table above (with 8 columns), it looks like the wheel on the right controls whether a prime number can divide a number "squares − 2": if the position of the prime number on this wheel is circled green, it can; for red positions, it cannot.

Conclusion: the good guess for the conductor is 8. (Again, this is a correct answer: if we continue the table with 8 columns, new circled numbers would appear only in the columns of 1 and 7, and all the bold numbers in these column are going to be circled.)

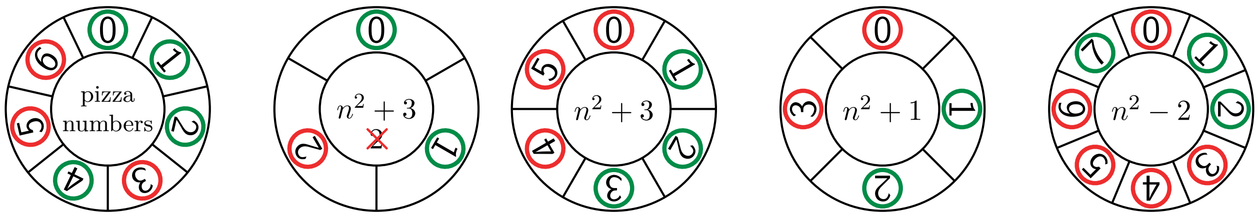
[24] However, these tables show the (relative) dearth of our data. Observe how the table with 5 columns contains no circled numbers in the column of 4. If this continues forever, this would be at least a *partial match* with the pattern we expect (the *full match* would be all columns having "all bold numbers circled", or "all bold numbers uncircled"; but already one column with this property is something "very interesting"[25]).

On the other hand, if we continue with numbers > 60, then soon we would find out that 79 divides $9^2 - 2 = 79$, so **79** would appear in the column in question. Likewise for other columns in which no (or few) circled numbers appears. — The only exception is the case of 8 columns — then the observed pattern would continue forever!

[25] In fact, for sequence of degree 3, such "partial matches" do actually appear (see Footnote 47 on p. 19). However, for degree 2, any non-trivial "match in one column" means that this is a "complete match" — here trivial matches are the columns which have at most one prime number.

Improved coloring

Combining together all the colorings discovered so far results in

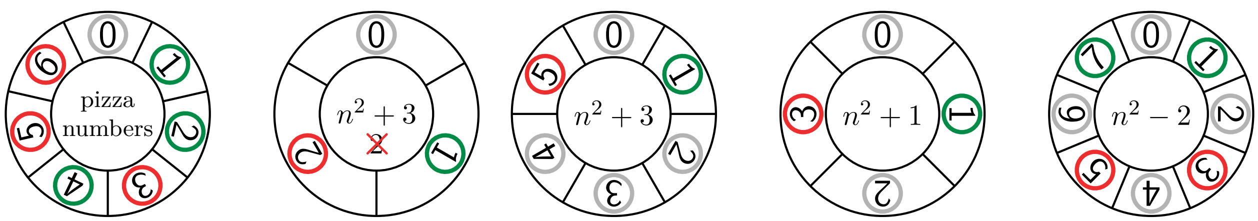


However, the naive way we obtained these pictures hides another extremely important property of these colorings. Recall: a particular position on a wheel matches a particular column in the corresponding table, and:

> The color of a position on a wheel reflects the color[26] of prime numbers in the matching column — with a few exceptional prime numbers allowed in a column.

Note that if we follow this rule literally, it is not clear how to color those columns which have only 1 prime number (and in examples above, we saw many such columns). Moreover, in the case "squares − 2" (the wheel on the right), there are columns mod 8 which contain no prime numbers at all — so in fact, we have no information about "the colors of these columns" whatsoever!

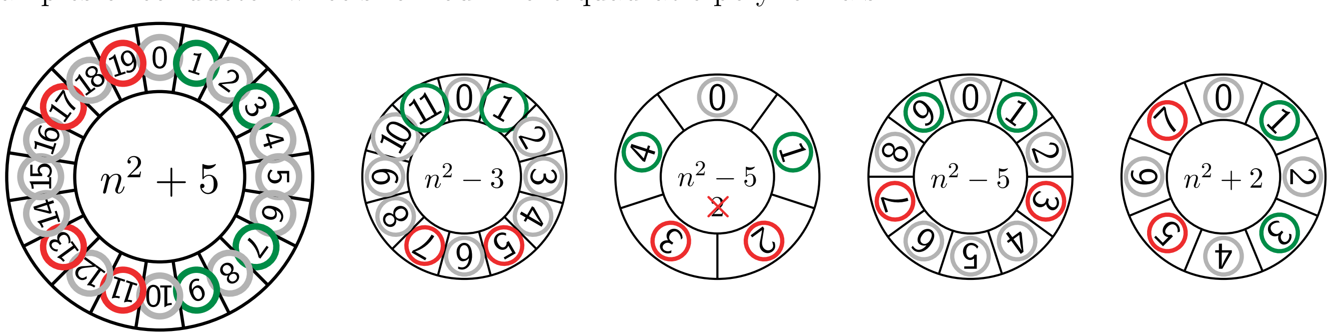
To be honest, we need to treat such columns in a special way. If we use gray color for the corresponding positions on the wheel, the pictures become very different:



These coloring are either preserved, or "made opposite" by a reflection in a vertical mirror! In particular, the coloring of the right wheel leads to a palindromic period (●)●●●●●●●, while the other colorings lead to anti-palindromic periods (●)●●●●●●, or (●)●●, or (●)●●●●●, or (●)●●●. Here we assume that −● = ●, and −● = ●, and put 0 in parentheses to make this symmetry of the wheels more visible in this "linear" rendition.

Euler's formulation: the "hidden symmetries"

To see what are the common features of our colorings of the conductoror wheels, and what are the differences, the more examples the better. First, the wheel for the pizza numbers coincides with the wheel for "squares + 7" (see Remark 6 on p. 16). Second, Exercises C4–C7 on p. 26 give five more examples of conductor wheels for four more quadratic polynomials:



[26] Well, recall that in some of the tables, instead of coloring prime numbers, we circled "the green ones", and left "the red ones" uncircled. M This allowed a conversion of these tables into problems for kids to solve.

Collect the info[27] on the size of the conductor wheel and (anti)symmetry w.r.t. vertical mirror:

Quadratic sequence	n^2-5	n^2-3	n^2-2	n^2+1	n^2+2	n^2+3	n^2+5	n^2+7
Size of the conductor wheel	5 (10)	12	8	4	8	3 (6)	20	7
(Anti)palindromic	Yes	Yes	Yes	Anti	Anti	Anti	Anti	Anti

This suggests **Conclusion:**

- the size of tahe conductor wheel is one of $|K|$, $|2K|$, $|4K|$, and
- the (anti)symmetry of the wheel for n^2+K is determined by the sign of K.

Since what is $|N|$-periodic or $|2N|$-periodic is also $|4N|$-periodic, these examples suggest that

(1) Whether a prime number p can divide numbers "squares $+\,N$" depends only on $p \bmod |4N|$.
(2) For $N<0$ the pattern of answers mod $|4N|$ is palindromic.
(3) For $N>0$ the pattern of answers mod $|4N|$ is anti-palindromic.

Here we use a special answer ("gray") for residues not mutually prime with $|4N|$.[28]

The first two of these three conditions[29] constitute what is now known as

Euler's formulation of Quadratic Reciprocity

invented by Swiss/Russian/Prussian mathematician Leonhard Euler. At his time, the proofs were known for $-5 \le N \le 4$ (some of these are trivial due to factorization of x^2+N). In fact, most of these known cases were established by Fermat (with proofs!) — almost a century before Euler. (Although Fermat stated his discoveries in a very different way.)

After Fermat, it took more than 150 years to prove the general case (done by Gauss — when he was 19)!

Remark 6: Similar questions about arbitrary polynomials of deg = 2 can be reduced[30] to questions about squares $+\,N$. For example, "completing the square" in pizza numbers $P_m := m(m+1)/2+1$ leads to $8P_m = (2m+1)^2+7$; hence the question about divisors of P_m can be rewritten as the question about divisors of l^2+7 for odd l. It is quite elementary that the latter questions has the same answer as "divisors of $Q_l := l^2+7$", and the Euler formulation implies that the colors for prime divisors of Q_l have a period of length 28.

This shows that among prime divisors of P_m only 2 can be an exception — due to the factor 8 above. However, since 2 is the only prime which is $p \equiv_{28} 2$, it cannot be an exception in this residue class!

Remark 7: However, we saw above that the *observed* period for the sequence (P_m) above has length 7. How to use Euler's prediction of a period of length 28 to show that the observed period of length 7 would continue forever?

Note that in the Euler formulation there is no need to allow exceptions. So if we know the color of a prime number p, one knows the color of its position $p \bmod 28$ on the 28-wheel (unless it is gray). This means that to find the period of length 28 (predicted by Euler's formulation), it is enough to find primes for every (non-gray) position on 28-wheel (call the smallest such prime *the check-prime*). There are 14 even positions (all gray); additionally, 7 and 21 are gray; so there are only 12 non-gray positions on 28-wheel. And after we know the period of length 28, there is a chance to show that length 7 will also work!

[27] Observe the missing $n^2 \pm 4$ and n^2-1. These omissions are discussed in Exercises C8 on p. 26 and C19 on p. 27.

[28] Since a gray column may contain at most one prime number, the first condition is trivial for such primes. Moreover, such prime $p \ne 2$ is a divisor of 0^2+N, so is of the "can divide" type. Likewise, $p=2$ divides either 0^2+N or 1^2+N.

[29] The third condition turns out to be an immediate corollary of the periodicity for $N=1$ and of top-multiplicativity (discussed below on p. 208). See Remark 120 on p. 210.

[30] See the section on p. 214 for details.

It turns out that the largest check-primes are 71 for the position 15 mod 28 and 83 for the position −1 mod 28; the check-primes for 10 remaining non-gray position are all below 60. Finding colors of these primes (e.g., from the table on p. 8), one can see that the period is ●●●●●●●●●●●●●● repeated twice. One can immediately see that the shown sequence is ●●●●●●● repeated twice, with every second color replaced by gray. — And this is exactly what is needed to see that the pattern of colors in the table on p. 8 would continue forever.

(We describe how to find the length of the shortest period in the general case in the section on p. 213.)

Remark 8: A surprising aspect of the discussions of Euler's formulation is that the people who already know about Quadratic Reciprocity may be at a disadvantage. The reason is that most of them know it in a very different formulation, one which was found about 50 years after Euler's by a French mathematician Legendre. On the surface, Legendre's formulation looks much more concrete and much more powerful. Until about 100 years ago, it was considered as "the only worthy" formulation. Most popular-math expositions of Quadratic Reciprocity discuss *only* Legendre's formulation.[31]

On the other hand, as the (amazing!) progress in number theory in 20th century has shown, it is Euler's formulation which has far-fetched generalizations.[32] Moreover, nowadays we know that either one of these formulations is an almost immediate corollary of the other![33]

Summing up: Euler's approach shows that studying divisors of numbers in a quadratic sequence uncovers "hidden symmetries" in the answers. Moreover, in what follows Legendre's formulation does not play any role.

For people who are already fluent with Legendre's formulation, to make it easier to switch gears to Euler's we discuss interconnections between these formulations in the Appendix on p. 207. (For the readers interested *only* in the Langlands program: since we use the topics discussed in this appendix in just a few very optional remarks, it may be safely skipped — unless you want to find more about Quadratic Reciprocity.)

Remark 9: For generic polynomial sequences P_n of degree 3 or more, the periodicity-in-prime-p described above is not going to be applicable. However, there are *particular* polynomials for which the divisibility properties are "almost the same"[34] as for quadratic polynomials.

Such polynomials are called *abelian*. There is a simple[36] criterion: P is abelian iff the solutions to $P(x) = 0$ are rational linear combinations[37] [38] of roots of 1. We return to this topic in Footnote 139 on p. 59.

---

[31] It looks like the majority of people who know quadratic reciprocity found it first in popular-math expositions. Contemporary "serious math" testbooks, and short overviews of Quadratic Reciprocity by prominent mathematicians would highlight Euler's formulation.

[32] The key difference is in Euler's formulation being way "more natural": it focuses on patterns in solutions to one particular problem about divisors of numbers in a quadratic sequence. On the other hand, Legendre's one highlights coincidences between answers to two different problems of this kind. (This is why "reciprocity" is a part of its name!) Like Euler's one, the most important generalizations target questions about divisors of values of a *particular* polynomial, — as opposed to questions about interrelations between divisors for different polynomials.

[33] Nowadays, the most useful application of Legendre's formulation is to prove the the Product Formula for Hilbert symbol — compare with the section on p. 144 and Footnote 280 on p. 100. However, this deduction is not immediate, so using Euler's formulation instead would not make the proof much harder.)

[34] The most important difference is that to encode the divisibility properties, one needs to colors the "conductor wheel" into more colors. These extra colors would not only describe whether the prime numbers p with this position on the wheel are divisors or not, but also "how frequent" are the numbers n with $p|P_n$. (For $\deg P = 3$, the frequency determines the color.)[35]

[35] **N.B. (???) Wrong!**

[36] Although non-constructive!

[37] For example, one proof of quadratic reciprocity writes $\sqrt{n}$ explicitly as a linear combinations of roots of degree $4n$ of 1 with coefficients ±1.

[38] **N.B. (???) Move to another place where we discuss abelian polynomials?**

Degree 3: a coarse-grained approach (Grades 3–4, Ilya 2018-05). The "hidden symmetries" are not as for degree 2

This part completes the digest of what we did at our Math Circles (this time, just in Grades 3–4). It is still OK to not read this meticulously (especially the parts between the signs Ⓜ ~~Ⓜ~~, written with Math Circles in mind!). The purpose of this part is only to *proclaim the existence* of "the Langlands pattern" — in the rest of the notes we are going to describe this pattern and the related issues.

From degree 2 to degree 3 (and the M-family)

In the investigation of divisors of polynomials of degree 2, one could restrict attention to polynomials $x^2 + N$ (see Remark 6 on p. 16). As we saw, there are two different important particular cases: for $N < 0$ the pattern is governed by a palindromic period, for $N > 0$ by anti-palindromic period. In a certain sense, "when N crosses the boundary $N = 0$", there is a "phase transition": "the hidden symmetries" of the answer make a drastic change.

Likewise, the cases of larger degree break into a few similar "regions" (two for degree 3), and "crossing a boundary" between these regions leads to a dramatic change in the type of "hidden symmetries". So if we want to restrict attention to a particular collection of, say, polynomials of degree 3, it is very important to ensure that this collection would have representatives from both regions. It turns out that this means that the collection should have both polynomials with 3 real roots, and (irreducible) polynomials with 1 real root.

This immediately rejects the "obvious first choice" of the family[39] "cubes$+N$". Ⓜ For our students, triangular numbers are as natural as squares, and, with 3D shapes, tetrahedral numbers are as natural as cubes. So in analogy to "triangular numbers + 1" they propose to use "tetrahedral numbers + 1. ~~Ⓜ~~ Unfortunately, this polynomial has a root when[40] side $= -3$ — and (Exercise on p. 27) such ("decomposable") polynomials sequences have every n as a divisor.[41]

As a workaround, we propose considering the sequences "tetrahedral numbers$+N$" with a suitable N, for example,[42] $N = 2$.

However, it turns out that the polynomials "tetrahedral numbers $+ N$" taken for integer values of N have either 1 real root (for $N \neq 0$), or are decomposable. Fortunately, considering a rational N makes perfect sense (prime factors of its denominator may be considered as "exceptions" allowed above[43]), so one can investigate "$M \cdot$ tetrahedral numbers $+ N$". Below, we consider cases $M = 1$, $N = 2$, as well as *the M-family* with $N = 1$.

When $M \in \mathbb{Z}$, the M-family has several elements with 1 real root, the rest has 3 real roots. As a bonus, among the latter, several are abelian (as in Remark 9 on p. 17). This gives a rich enough zoo of polynomials of degree 3, which is quite sufficient for our purposes.[44]

---

[39] This was one of the reasons for us to start with pizza numbers: since "cubes $+ N$" is not a good choice, we wanted to avoid "squares $+ N$" for as long as possible.

[40] Ⓜ When discussing triangular and tetrahedral numbers, we introduce continuation of these sequences "to the left". — This is easy to do due to our description via (repeated) differences, as in the table on p. 6 and the group 0 of exercises on p. 22. — So the students know that dealing with a tetrahedral number for "a negative side" makes perfect sense.

[41] While they are not directly related to the Langlands program, it turns out that using our approach with decomposable polynomial leads to very interesting effects. Compare with the section on p. 63.

[42] Ⓜ The kids also suggest another sequence: the "cake numbers". The difference with pizza numbers is that the cake is 3D, and we allow cuts to be non-vertical. This leads to "a tetrahedral number$-$a triangular number$+$it's side$+1$". Unfortunately, this is also decomposable (it vanishes when side $= -1$).

[43] For example, in Footnote 21 on p. 11.

[44] See also the section on p. 144.

Example: Divisors of "tetrahedral numbers + 2"

Proceeding as[45] for deg = 2, we fill the table of divisors:

Side	1	2	3	4	5	6	7	8	9	10	11
Tetrahedral number + 2	3	6	12	22	37	58	86	122	167	222	288
As products	1×3	1×6 2×3	1×12 2×6 3×4	1×22 2×11	1×37	1×58 2×29	1×86 2×43	1×122 2×61	1×167	1×222 2×111 3×74 6×37	1×288 2×144 3×96 4×72 6×48 8×36 9×32 12×24 16×18

This leads to possible divisors 1, 2, 3, 4, 6, 8, 9, 11, 12, ... Note that the numbers 5, 7, 10 do not appear in the row "As products". In fact, if we continue the table to the right, they would never appear (so we may color them red).[46] Extending the table far enough to the right, one may obtain the following color pattern:

1 **2** **3** 4 **5** 6 **7** 8 9 10 **11** 12 **13** 14 15 16 **17** 18 **19** 20 21 22 **23** 24 25 26 27 28 **29** 30 **31** 32 33 34 35 36 ...

It turns out that even after grouping into several columns, and looking only at prime numbers, the patterns of colors we observed for sequences of degree 2 won't work for this sequence.

For example: since **11** and **23** are of different colors, grouping into 12 columns would not help (unless 11 or 23 are exceptional — but similar mismatches also happen further to the right). From this it follows that grouping into 3, 4, or 6 columns cannot help either.

Likewise, the mismatch of **19** and **29** excludes 10 columns (hence 5), while mismatch of **17** and **31** excludes 14 (hence 7). The data above excludes also 16 (hence 8), 18 (hence 9), and 22 (hence 11). Extending the table, one would exclude more and more arrangements into columns.[47]

---

[45] Recall that mathematically, we want to describe the modular arithmetics for which the polynomial equation $P_x = 0$ has roots.

[46] For pizza numbers, we found (see our Observation on p. 7) how far in the table it is enough to check to be sure that the given number would never appear as a divisor listed in the table. It is easy to do the same for the sequence above.

[47] While we won't see the pattern "in some columns all primes are red, and in the remaining columns they are all green", in fact, with a suitable number of columns, a "partial pattern" *would* appear. The suitable number of columns is 971 (this is not a misprint![48] Compare with Remark 48 on p. 77).

With 971 columns, about half of them would contain only green primes. However, the remaining columns would not be "all red", even when one observes the primes only — every such column would contain a mix of red and green primes (the mix happens to be in "proportion" 2-to-1; compare with Remark 48 on p. 77).

Existence of such "all green" and "red/green mix" columns was first discovered even before Gauss; it was proven to hold in general about 100 years ago. However, until recently, the question

$\boxed{\text{Describe the pattern of colors \textit{inside} a mixed red/green column}}$

could be answered only in the particular cases covered by the Class Field Theory (compare with the section on p. 82). We discuss this in more details in Remark 49 on p. 77.

(Actually, *the particular polynomial considered above* has negative discriminant ($-3{,}884 = -2^2 \times 971$), so *it is* covered by the Class Field Theory. See Remark 18 on p. 36.)

[48] We investigate these effects in the group D of exercises on p. 27.

Since for about 20 years now we know the actual pattern of the colors for deg = 3, we may definitely exclude the patterns similar to those found above for sequences of degree 2. However, this requires using one of the most important (and impressive) tour-de-forces of math of 20th century![49]

Recent developments: the Langlands program

The sequence of colors above is a part of a big zoo of sequences. One can start with different polynomials of degree 3; one may also consider polynomials of higher degrees.

As usual, having a wider collection of examples may uncover a more beautiful landscape — and sometimes this makes the previously known examples easier to understand. In our settings, this happens with introduction of *polynomials in several variables*.

However, in this case instead of coloring a number according to whether it *can* divide a value of the polynomial, one should mark *how often* a given number is a divisor of the values. Compare with our "transliteration rules" on p. 60.

In fact, another extension of the pool of examples happens when one considers common zeros of several polynomials (with several unknowns).

These sequences of colors remained mysterious for a long time. Ⓜ A few weeks before considering this topic at Math Circles, we discussed discrete logarithms. Ⓜ One of the main messages was that mathematicians expect that this problem ("how discrete logarithms depend on the size of the wheel") *does not follow* any pattern. Until recent decades, there was no clue whether the color sequences like those above would all have a pattern (but possibly, a very complicated pattern), or sometimes the situation could be as with discrete logarithms.

Things changed about 50 years ago, when a Canadian mathematician Langlands started to ask his colleagues some "crazy" questions; a few years later, these questions crystallized into a chain of conjectures connecting

- questions about divisibility in polynomial sequences[50] (really hard; considered very important, but impenetrable before), and
- questions of mathematical analysis (hard, but much easier to handle).

These connections would show that all these problems about divisibility have a pattern in the answers — however, this pattern is extremely complicated even to describe (not mentioning proving this!). At the coarsest possible level, one can say that the (hidden) symmetries we saw in red/green coloring of prime numbers in the case of a polynomial of degree 2 — *periodicity* and *mirror (anti)symmetry*[51] — are replaced by yet "more hidden" symmetries.

Ⓜ To be able to expose the pattern of hidden symmetries, one needs to understand many different concepts:

- Wheels (= residue classes and their symmetries);
- symmetric tessellations (or "tilings");
- curved geometries,
- working with infinities,
- fractals,
- harmonies, harmonics and waves ("Harmonic Analysis"),
- heat propagation. Ⓜ

[49] In fact, in his review written in 1972 (before the importance of the Langlands program was fully realized), Wyman claims that it is possible to exclude these patterns using *only* the tools of the Class Field Theory. However, I do not recollect seeing this argument actually written down. (See the discussion on **MathOverflow**/11688, especially the answer by Keith Conrad.)

[50] . . . and tables=bi-sequences etc.

[51] In other words: (anti)palindromicity of the period.

Similar to the case of degree 2, there is a particular size of the wheel (the *conductor*) which is related to a particular polynomial sequence. However, it controls the sequence not directly, but by selecting a particular "size"[52] of a tessellation of a curved geometry (as mentioned above) in which we consider the heat propagation.[53 54] Tesselations of different "sizes" have different symmetries — and these symmetries become the "hidden symmetries" of the pattern of colors.

These conjectures (called the *Langlands program*) explain (among other things) how to find the conductor — however, the recipe is not straightforward. Before going half-way in writing these notes, I had no clue what the conductor for the sequence "tetrahedral numbers + 2" above (and similar sequences) may be![55]

One must keep in mind that initially Langlands has been working on a very specific circle of problems. Until his dreams crystallized, nobody expected these problems to be related to the questions of red/green coloring we consider here.

Following the parable about blind men and an elephant, Langlands have been investigating an ear of an elephant, while our questions concern the trunk of the elephant. What happened next is that, contrary to the parable, he could figure out the general appearance of the whole elephant using just the data from his research of the ear. From this, he unraveled how to access all the particular features of the elephant in a uniform way.

Meanwhile, during these 50 years, mathematicians managed to investigate "the trunk" by following the recipes of Langlands. Other mathematicians could prove that what Langlands visualized actually holds in "the particular case of the trunk". So today, we can discuss the trunk of an elephant in detail — which has not been dreamed of before Langlands.

After the Langlands program was thought up, it became one of the most important focus points of contemporary mathematics. *A lot* of mathematicians work on realizing this program. Moreover, about 20 years ago, one of the major way points of the program was achieved: the Langlands program was proved in the cases connected to 2D tessellations (as opposed to higher dimensions).

Such 2D tessellations are related to polynomial sequences up to degree[56] 3. In particular, this leads to a proof of Langlands' pattern for our sequence of colors for[57] "tetrahedral numbers + 2".

M Moreover, just a few weeks before we discussed that at our Math Circles, the achievements of Langlands were formally recognized as well: he won what is considered the most prestigious award for mathematicians: the Abel prize. This prize is in fact much more prestigious than the Nobel Prize. For example, every year 2 or 3 physicists are awarded the Nobel Prize — but typically, only one mathematician a year wins the Abel prize. M

[52] Note that in the "usual" geometry, given a tessellation, we can rescale it, and it remains a tessellation. However, curved geometries *allow no rescaling* — so every "type" of tessellation may exist in one size only.

[53] We provide examples of such tesselations in Chapter on p. 84. We discuss some idiosyncrasies of the heat propagation in curved geometry in Footnote 96 on p. 39.

[54] An alternative approach is to say that the conductor controls "the laws of fractality" of the graph of Fourier transform of the sequence in question. (However, if one calculates this Fourier transform "naively", one would get infinite values! We will start addressing this in Remark 20 on p. 37.)

[55] After finding the conductor, the Langlands program leads to a recipe describing certain integrals (see Remark 15 on p. 35). The values of these integrals are whole numbers matching the colors above: for example, the number may be 0, 1 or 3 (with 0 for red, 1 and 3 for green; compare with p. 60).

One can calculate these integrals approximately, then round to the closest integer. This gives a "practical" (meaning: computationally feasible) recipe to find colors of arbitrarily large prime numbers.

[56] They also cover a (small) subset of polynomials of degree 4 (see the section on p. 131), as well as polynomials of degree 5 if the discriminant is a perfect square.

[57] When discussing this in Math Circles, we cheated, and pretended that to treat this sequence one *needs* the Langlands program. In fact, *this particular sequence of degree 3* is covered by the Class Field Theory.

One must massage this sequence a bit so that one *needs* Langland's approach to see the pattern. For example, one may consider "$20 \times$ tetrahedral numbers $+ 1$." See Remark 18 on p. 36.

Exercises for the preceding chapters

Exercises 0: Polynomials

This section explains our notations first, then the exercises follow. On one hand, these exercises are not used in the main text directly. However, they explore the landscape of the fundamental properties of polynomials taking integer values.

In math, the symbol "=" is used in a multitude of different meanings. It may be used in the meaning "solve this", as in $x^2 + 3x - 7 = 0$. Or it may mean "define *this* in terms of *that*, as in "put $P(x) = x^2 + 1$". Or it may mean that two things "are the same", as in "now we can find the polynomial P, and the answer is: $P(x) = x^2 + 1$". Sometimes, it also appears as a part of larger notation, as in "$r = s \bmod 5$".

In these notes in potentially confusing situations we try to use different notations for different meanings of this symbol. In particular, $b \equiv c$ means that b and c are the same "as a whole"; for example, if they are polynomials, this means that their coefficients coincide (so the values coincide everywhere). When we define b in terms of c, we may write[58] $b := c$ or $c =: b$. Time to time we may use $b \stackrel{\text{def}}{=} c$ meaning that "to see that b equals c one does not need any tricks, just inspect the definitions". Furthermore, $r \equiv_5 s$ means that the residues of r and s modulo 5 (denoted $r \bmod 5$ and $s \bmod 5$) coincide.

Below, we consider a polynomial sequence with elements P_0, P_1, P_2, ... (As everywhere in these notes, this means that there is a polynomial $p(x)$ such that $P_0 = p(0)$, $P_1 = p(1)$, $P_2 = p(3)$, etc.) When we want to consider this sequence "as a whole", we use parentheses, as in (P_k), or just use one symbol P.

Exercise 01: Consider a sequence (P_k), $k \geq 0$, such that $P_{k+1} - P_k = k$. Show that P_k is a quadratic polynomial of k.

Hint: Write a formula for P_k in terms of k and P_0.

Exercise 02: Show that the sequence (P_k) from the preceding exercise can be "extended left", to $k < 0$, such that the equality $P_{k+1} - P_k = k$ holds for every $k \in \mathbb{Z}$.

Exercise 03: Show that in the preceding exercise $P_{-k-1} \equiv P_k$.

Below we denote by T_k the k␣th triangular number.

Exercise 04: Show that $T_{-k-1} \equiv T_k$.

Exercise 05: Given a polynomial sequence (P_k), define $Q_k := P_{k+1} - P_k$. Show that (Q_k) is a polynomial sequence of degree less than $\deg P$.

Exercise 06: Show that in the preceding exercise, if Q_k is odd (as a function of k), then $P_{-k-1} \equiv P_k$.

In the following exercises, we define a collection of polynomial sequences. We could denote them as a sequence (N_k), and a sequence (Q_k), and a sequence (R_k), etc. — but this quickly becomes unwindy. To simplify mentioning these sequences, we use "an extra index" on top left; so instead of (N_k) we write $({}^{(0)}P_k)$, instead of (Q_k) we write $({}^{(1)}P_k)$, instead of (R_k) we write $({}^{(2)}P_k)$, etc.

With these notations,

- define ${}^{(0)}P_k := 1$,
- then ${}^{(1)}P_k := {}^{(0)}P_k \cdot k/1$,
- then ${}^{(2)}P_k := {}^{(1)}P_k \cdot (k-1)/2$,
- then ${}^{(3)}P_k := {}^{(2)}P_k \cdot (k-2)/3$,
- then ${}^{(4)}P_k := {}^{(3)}P_k \cdot (k-3)/4$, etc.

---

[58] Sometimes people can use $b \stackrel{\text{def}}{=} c$ in this context — or maybe even $c \stackrel{\text{def}}{=} b$, but only when it is absolutely clear what is defined in terms of what. We avoid this; the notations we use are "directional", so less confusing.

(so ${}^{(5)}P_k \equiv k(k-1)(k-2)(k-3)(k-4)/5!$ etc.).

Exercise 07: Write down the first 8 elements of every sequence $({}^{(0)}P_k)$, $({}^{(1)}P_k)$, $({}^{(2)}P_k)$, etc. up to ${}^{(5)}$.

Exercise 08: Show that all the numbers ${}^{(n)}P_k$ are non-negative integers (if k, n are non-negative integers).

Exercise 09: Show that given any positive integer n, the sequence $({}^{(n)}P_{k+1} - {}^{(n)}P_k)$ coincides with $({}^{(n-1)}P_k)$.

Exercise 010: Show that for any quadratic polynomial S_k, there are three numbers α, β, γ such that $S_k \equiv \alpha \cdot {}^{(0)}P_k + \beta \cdot {}^{(1)}P_k + \gamma \cdot {}^{(2)}P_k$. Express numbers α, β, γ in terms of the numbers S_0, S_1, S_2, S_3, ... (you may use as many of values of S_k as you need).

Hint: If a quadratic polynomial $U(x)$ vanishes at $x = 0, 1, 2$ (so $U(0) = U(1) = U(2) = 0$), then $U \equiv 0$. (Indeed, if its degree is 2, it can have at most 2 roots, and if it is 1 then it has only 1 root.)

Hint: Guess α, β, γ, then put $Q := S - (\alpha \cdot {}^{(0)}P + \beta \cdot {}^{(1)}P + \gamma \cdot {}^{(2)}P)$.

Exercise 011: Start with any polynomial sequence written as a row of numbers, then proceed as below:

Side	0	1	2	3	4	5	6	7	8	9	...
Cube	0	1	8	27	64	125	216	343	512	729	...

+1 +7 +19 +37 +61 +91 +127 +169 +217 + ...

+6 +12 +18 +24 +30 +36 +42 +48 + ...

+6 +6 +6 +6 +6 +6 +6 + ...

+0 +0 +0 +0 +0 +0 + ...

write the differences of the numbers next to each other (as in the green line), then the differences of differences (as in the blue line), then the differences of differences of differences (as in the red line), etc. (Obviously, after one gets a whole line of 0s, there is no sense to continue!)

Denote the leftmost number in the row of the initial polynomial α, denote the leftmost number in the row of differences β, then continue with γ, etc. Write the initial sequence in terms of the numbers α, β, γ etc. and our sequences ${}^{(0)}P$, ${}^{(1)}P$, ${}^{(2)}P$ etc.

Exercise 012: Starting with the table in the preceding exercise, plug the corresponding numbers α, β, ... into the expression you obtained as the answer to this exercise. Simplify the result, and check that it matches the name of a row in this table.

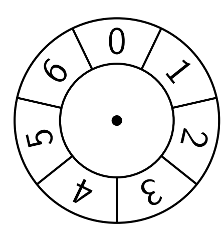
Exercise 013: Given any polynomial sequence such that all the numbers in this sequence are integers, show that it coincides with a polynomial whose coefficients are integer numbers divided by $d!$. Here d is the degree of the sequence.

Hint: Check that this holds for any sequence $({}^{(n)}P_k)$, then use the result of Exercise 011.

Exercise 014: Show that for any given polynomial sequence (Q_k), there is a polynomial sequence (P_k) such that $Q_k \equiv P_{k+1} - P_k$.

Hint: Exercises 09 and 011.

Exercises A: Wheels and modular properties of arithmetic progressions



One way to visualize the divisors of numbers in arithmetic progression is to use "wheels" (sometimes called "clocks"). To check whether, e.g., 7 is a divisor, organize residues mod 7 in a circle, as on the right ("*the wheel of size 7*"). An arithmetic progression with the step J corresponds to a sequence of jumps of equal length J between the positions on the wheel. (Here J may be larger than the size of the wheel. Then "such a jump" includes several full rotations around the wheel.)

Exercise A1: Say that a jump of length J is *universal* on a wheel of size S if one can travel between any two positions on the wheel by a suitable number of such jumps. (Note that a jump is universal iff one can go from the given position to one of its neighbor by several such jumps.)

Show that a jump of length J is universal on a wheel of size S iff a jump of length S is universal on a wheel of size J.

Hint: These satements may be written as identities of the form $\ldots - \ldots = \pm 1$.

Exercise A2: Show that a jump of length J is universal on a wheel of size S if J is mutually prime with S.

Hint: Induction based on: if $J \geq S$, then one can decrease J by replacing it by its remainder in division by S; otherwise, if $J > 0$ use the preceding exercise. (But the base may be non-obvious!)

Exercise A3: In the preceding exercise one can replace "if" by "if and only if".

The preceding two exercises essentially describe the Euclid's algorithm. It calculates the largest common divisor using the described inductive process.

Recall that "periodicity" of function defined on a subset of $\mathbb{N}$ is discussed in Footnote 13 on p. 8.

Exercise A4: Consider a function T_n of n defined on a subset S of $\mathbb{N}$. Suppose that S' is a subset of S and that the restriction of T to S' is N-periodic. Show that if $S \smallsetminus S'$ consists of numbers $n_1, \ldots\ n_k$ and the numbers in S' are not divisible by $n_1, \ldots, n_k$, then T is $n_1 \ldots n_k N$-periodic.

Exercises B: Modular properties of $\triangle$-numbers

Here (same as in the rest of these notes), "*the period*" of a periodic sequence (S_k) is a certain "range" of values which repeats indefinitely. (In particular, we need to use the term "the length of the period" for the *other* meaning of the word "period".) For example, the infinite word "Banananana… " has periods "an", "na", "nanana" etc.

Below we denote by T_k the k␣th triangular number.

Exercise B1: Denote by D_k the last digit of the number T_k. Find the minimal length of the period of the sequence (D_k).

Exercise B2: Show that D_k has a period which is a palindrome.

Below, the "*digital root*" of a number N is obtained by taking the sum S of digits of N, then the sum Σ of digit of S, then the sum of digits of Σ (repeating until the result does not change — which happens when the result is less than 10).

Exercise B3: Denote by R_k the "digital root" of the number T_k. Find the minimal length of the period of the sequence (R_k).

Exercise B4: Show that R_k has a period which is a palindrome.

Exercise B5: Show that for every integer $t > 0$ the sequence $(T_k \bmod 2t)$ cannot have a period of length $2t$.

Hint: For every t and every k the sum $(k+1)+(k+2)+\ldots+(k+2t)$ is not divisible by $2t$ (but it is divisible by t).

Exercise B6: Show that the length of the shortest period for the sequence $(T_k \bmod 2t-1)$ is $2t-1$, and for the sequence $(T_k \bmod 2t)$ it is $4t$.

Hint: reduce periodicity to a statement about divisibility of certain expressions; take into account that it should work for *every* k.

Below, "an interval" may be an interval in $\mathbb{R}$, or an interval in $\mathbb{Z}$ (depending on the exercise).

Exercise B7: Given a number $t > 0$, find the length of the shortest interval $\mathscr{I} \subset \mathbb{R}$ such that any even t-periodic[59] function $f(x)$ is positive for $x \in \mathscr{I}$ iff it is positive for every $x \in \mathbb{R}$.

Hint: If we know that $f(x) > 0$ for x in $[-t, 3t]$, does it follow that $f(x) > 0$ everywhere?

Exercise B8: Same question as in Exercise B7, but when $f(x)$ is defined only for integer x.

Exercise B9: Same question as in Exercise B8, but instead of "even" $f(x)$ we consider a function such that $f(-1-x) \equiv f(x)$. (In other words, the graph of $f(x)$ is symmetrical w.r.t. reflection not in the line $x = 0$, but in the line $x = -½$.)

[59] In other words, we require that $f(x+t) \equiv f(x)$.

Exercise B10: Show that for any function $f(x)$ defined on $x \in \mathscr{I}$ for the interval $\mathscr{I}$ of Exercises B7–B9, there is a way to extend it to a function of $x \in \mathbb{R}$ (or $x \in \mathbb{Z}$) such that f is t-periodic, and has the required symmetry ("evenness" or a "reflection in $x = -½$).

In presence of symmetries, such regions as the interval $\mathscr{I}$ above are called "*fundamental domains*" of this collection of symmetries. Note the similarity to what we discuss in the section on "toy fractality" on p. 43, on "$c = 1$ in Lobachevsky geometry" on p. 85, as well as the "polygones with colored lines as the sides" in the section "The case $c = 5$" on p. 87.

Exercise B11: Show that for any integer t the sequence $(T_k \bmod t)$ has a palindromic period.

Hint: The solution is different for even/odd t.

Exercises C: Quadratic Reciprocity modulo small numbers

The solutions of the following two exercises form the core of our discussion in the section on p. 10. We recommend the reader to solve them independently. (As an extra hint, one can use Remark 10 on p. 25.)

Exercise C1: For the last table of 3 attached tables in a handout[60] (the table for Number$^2 + 3$), circle the numbers below the table according to the instructions.

Exercise C2: Analyse the pattern of circled numbers from the preceding exercise using a printout with arrangements of numbers 1, ..., 60 into 3, 4, 5, ..., 10 columns (from the handouts for the main text[61]):

- In each arrangement, circle the same numbers as in the preceding exercise.
- In one (or more) of the arrangements, the circled numbers have a very strong tendency to appear in certain columns only.
- To see these patterns, one needs to pay attention to bold/non-bold numbers (same as we did in the section on p. 6).

Find all the arrangements for which such tendency appears. (There should be at most 1 exception[62] to the pattern you found!)

We say that the corresponding "number of columns" *works as a conductor* for this sequence. If there are no exceptions, we say that it *works as a level* (see Footnote 16 on p. 9). (Of course, what is interesting is to find *the conductor* and *the level*: the *smallest* numbers working as a conductor or as a level.)

The solution of the following exercise is discussed in the section on p. 12. We recommend the reader to solve it independently.

Exercise C3: For each of the first two tables of 3 attached tables[63] (the tables for Number$^2 + 1$ and Number$^2 - 2$) proceed as in two preceding exercises.

Remark 10: Summarize the main result from the first chapter: there is a number C such that the "green/red color" of a prime number p (i.e., whether p is a divisor of one of the numbers in the quadratic sequence) depends only on $p \bmod C$ (with only a finite number of exceptional p). (Moreover, one can avoid exceptions increasing C appropriately.)

In particular, if one arranges numbers into C columns and circles as above (in Exercise C2), the prime numbers in every column are going to be all circled ("green" in the notations of ???) or all uncircled/"red" (possibly with a finite number of exceptional p — if exceptions appear for the given C).

[60] The name of the handout is `handout-factor-squares.pdf`.
[61] The name of the handout is `handout-primes-in-columns.pdf`.
[62] Compare with the section on p. 213.
[63] The name of the handout is `handout-factor-squares.pdf`.

Remark 11: It may help to know that for quadratic sequences of the form $n^2 + \text{const}$ only $p = 2$ may be an exceptional prime.[64] Note that we saw that appearing for $n^2 + 3$ in the first arrangement on p. 11: the bold/prime number 2 is circled in green, while the rest of primes in this column are not — hence they "are red").

For more general sequences, the divisors of (the numerator of) the leading coefficient may also be exceptional.

The following exercises continue what we do on pp. 10–14 (as well as in exercises above); they investigate 4 more variants of quadratic sequences.

Exercise C4: Proceed as in Exercise C1 using 4 attached tables[65].

Exercise C5: Use the tables with circled numbers from Exercise C4. For each integer $n = -5, -3, 2, 5$ find whether one can use the arrangements of numbers into columns as on pp. 10–14 (or from Exercise C2) to expose the pattern of colors for sequences $(P_k) := (k^2 - n)$. (In other words, find whether the conductor C may take values below about 10.) Here we do not allow exceptional primes.

Exercise C6: The same with exceptional primes allowed.

Our tables with "the arrangements of numbers into columns" allow finding conductors C up to 10. However, if one pays attention only to every other row of the arrangement, one can also cover even C up to 20. So:

Exercise C7: Investigate the cases of two preceding problems where the conductors $C \leq 10$ could not be found. Show that these collection of "numbers to circle" match certain even conductors $C \leq 20$.

Exercise C8: Solve a similar problem about $P_k := k^2 + 4$ using the results for $k^2 + 1$ obtained in the text.
Hint: $p = 2m - 1$ divides $k^2 + 4$ iff p divides $K^2 + 1$ for a certain K. (For an extra hint see Remark 12.)

Exercise C9: For every n between -5 and 5 find whether the polynomial $P_k := k^2 - n$ allows $C \leq 10$ (separately for the case "without exceptional primes", and "with them").

Exercise C10: For every n between -5 and 5 for which you could find C (with or without exceptions) in the preceding exercise, fill a column in a table with 3 rows: a row for n, a row for the minimal possible C with exceptions, and a row the minimal possible C without exceptions.

Exercise C11: Find the pattern for the table, and try to fill the missing cells. Check whether your guesses work. (May be hard. . .)

The exercises below may be solved either heuristically using numerical experiments (provided one can write a simple program), or by inspecting the results from the Appendix starting on p. 207.

Exercise C12: Find four integers n with the smallest possible $|n|$ such that one cannot use the arrangements into ≤ 8 columns to find the pattern of colors for sequences $(P_k) := (k^2 - n)$. (In other words, the size of the conductor wheel for each of these sequences should be larger than 8.) Here we allow exceptional primes.

Exercise C13: The same questions (with five integers n) when we do not allow exceptional primes.

Exercise C14: Find seven integers n with the smallest possible $|n| > 100$ such that $C \leq 8$ for $(P_k) := (k^2 - n)$. Here we allow exceptional primes.

Exercise C15: Same question (with six numbers) if we do not allow exceptional primes.

Remark 12 **(Extra hint for Exercise C8 on p. 26):** Show that a prime $p = 2m - 1$ divides $k^2 + 4$ iff p divides $K^2 + 1$ for a certain K. Here one can take $K = mk$ with $m \in \mathbb{Z}$ or $k = 2K$.

The following exercises explore the fine points of what we did in the observation on p. 7. One can use the results of the group B of exercises on p. 24.

Exercise C16: In the context of Exercise C4, explain why there is no marked numbers in the bottom third of the tables.

[64] We discuss the details in the section on p. 213.

[65] The name of the handout is `handout-factor-more-squares.pdf`.

Exercise C17: Explain why marked numbers appear so rare in the middle third of the table.

Exercise C18: In fact, using only the factorizations in these tables, one can find all numbers (up to a certain number N) which may appear as divisors of numbers in each quadratic sequence. Find N.

Hint: There is a simplified variant of this exercise in Remark 13 on p. 27.

REMARK 13 **(Variant of Exercise C18):** It might be easier to split this exercise into two. (Here we assume that the table is extended on the left with a column for 0.)

(1) Given the list 1, ..., 60 at the end of the problems in the handouts, using the whole table in the beginning of the problem is an overkill. In fact, Exercise C16 shows that one can use a much shorter table, and still *verify* that any uncircled number cannot be a divisor.

How much can one shorten the tables for polynomials we consider (of the form $P_k \equiv k^2 - n$ for a fixed $n \in \mathbb{Z}$)? Try to find the smallest feasible answer.

(2) Given the table as in the problem going from 0 to 60, one can make the list 1, ..., 60 at the end of the problem longer, and still verify that any uncircled number is a divisor (for any polynomial of the form $P_k \equiv k^2 - n$ for a fixed $n \in \mathbb{Z}$). How much longer? Try to find the largest feasible answer.

In exercises below, we assume that P takes integer values in integer points.

Exercise C19: Show that if a polynomial $P(x)$ has an integer root x (we do not prohibit $x < 0$) then any integer $d > 0$ is a divisors of one of the numbers $P(n)$ with integer $n > 0$.

Hint: Increasing d this may be reduced to the case of P having integer coefficients.

Exercise C20: Show that if a polynomial P has a rational root x then all prime numbers (with a finite number of exceptions) are divisors of one of the numbers $P(n)$ with integer $n > 0$.

Exercises D: "Qubic Reciprocity" cannot be exactly the same as quadratic

When dealing with Quadratic Reciprocity, we start with a quadratic polynomial sequence P_k, and try to find a "colored conductor wheel": a number C (a "level", or a "conductor", see Footnote 16 on p. 9) such that for a particular position on the wheel (which is "a residue mod C") all the prime numbers p with this residue "have the same color w.r.t. (P_k)". In other words, given every residue $r \bmod C$, we can assign to it a "color" **Yes** or **No** such that all primes $p \equiv r \bmod C$ are either all divisors of certain numbers in the sequence (P_k) (with k which may depend on p; this is the **Yes**-case), or no number P_k is divisible by any such p (the **No**-case). (This condition is, of course, vacuous if there is no such prime number $p \equiv r \bmod C$, or only one such number — which happens if $(r, C) \neq 1$.)

Recall that "a complete" generalization of this simple pattern (which we call "the hidden symmetry of translations by $C\mathbb{Z}$") to polynomials of higher degree requires the (much more complicated) approaches of the Langlands program. (Exposing these approaches is the main aim of our notes.) However, if we ignore the Langlands program, a *partial* "direct generalization" of these patterns is possible.[66]

In this "partial approach", we do not require the statement above to work for *all* residues $r \bmod C$, but only for some of them. We investigate this effect in the exercises below.

For the exercises below, the number C may be larger that one may investigate "by hand" using only the "arrangements of whole numbers into columns" from the handout `handout-primes-in-columns.pdf`. So we assume that the reader can write a program checking the conditions of the exercises below — so we do not provide the table of divisors of the numbers P_k, and do not promise that the provided in the handouts "arrangements into a few columns" are enough to cover these exercises. (Example subroutines needed for such a program are provided below.)

---

[66] When the dust settles down, it turns out that such "partial results" have very little relation to the "actual" hidden symmetry which may be exposed for divisors of numbers in the sequence (P_k). This is why we pay so little attention to this effect in the main body of the notes. (We discuss it in Remarks 48, 49 on pp.77–77.)

Moreover, what is easy is to check whether "*many* prime p␣s behave as expected" but not whether "all prime p␣s behave as expected". So for the purpose of the exercises below, it is permitted to replace "all primes in a column have a property $\wp$" by "the first 100 primes in a column have a property $\wp$". (Below, we call this number 100 "*the poor man's infinity*".)

Exercise D1: Consider $P_k := k^3 - k - 1$. Find the smallest integer C such that there is a residue $r \bmod C$ satisfying the condition[67] $(r, C) = 1$ and

- Either: for all primes $p \equiv r \bmod C$, there is an integer k such that $p | P_k$.
- Or: $p \nmid P_k$ for all primes $p \equiv r \bmod C$ and all numbers P_k in the sequence.

Exercise D2: Double-check the result by increasing "the poor man's infinity" to 10,000.

Exercise D3: Same question for $P_k = k^3 - k + 1$.

Hint: there is no need to rerun the calculations!

Exercise D4: In conditions of Exercise D1, count how many residues $r \bmod C$ satisfy $(r, C) = 1$. (This is called "*Euler's φ-function*".)

Exercise D5: Which part[68] of the count of residues r from Exercise D4 satisfies the conditions on r from Exercise D1?

Exercise D6: Would one get false positives (in other words, the "fake=extra" pairs (r, C) with a smaller C) if in Exercise D1 one uses 10 instead of 100 for "the poor man's infinity"?

(We investigate these effects in the next group of exercises E.)

Exercise D7: The same questions as above, but for $P_k := k^3 + k - 1$.

(We mention this sequence in Footnote 445 on p. 145.)

Exercise D8: The same questions as above, but for[69] P_k being the k␣th tetrahedral number plus 2. (Expect a very long calculation!)

Hint: To speed up the final step of calculations, one should make "the poor man's infinity" as small as possible (but so that it still avoids false positives; of course, one needs to recheck such "preliminary" answers similarly to how we did it in Exercise D2).

Exercises E: Aside: search's cutoff and the degree of certainty

The following exercises are not fully tuned/debugged yet.[70]

In the preceding exercises, we completely ignored the question "How to *choose* the poor man's infinity?". The experiments described in these exercises show that *a lot* of prime numbers "demonstrate the needed behaviour". How can this be translated into "the degree of certainty" that this behaviour would continue for *all* prime numbers?

On one hand, strictly speaking this question is completely irrelevant to what we do in the main body of our text. However, the spirit of our exposition is that a lot of "hidden symmetries" could

[67] This condition ensures that there are infinitely many prime numbers $p \equiv r \bmod C$. Compare with Dirichlet's theorem on arithmetic progressions.

[68] As Remark 48 on p. 77 suggests, one should expect this number to be something like ⅙, or ⅓, or ⅔, or ½, or ⅚ of "residues mutually prime with C". This is more or less the claim of Chebotarev's density theorem.

However, explaining the connection of this exercise with the Chebotaryov's density theorem requires results on Galois extensions of residues mod p. Since in these notes we only mention these results really cursorily, there is no hope to draw on the actual connection. (On the other hand, let us note that for the fractions above, the common denominator is 6 — which is the order of the "Galois group of P". Since elements of this groups are permutations of the roots of P, its order divides $(\deg P)!$ — which is $3! = 6$ for our P.)

[69] We discuss this polynomial in the mentioned above remarks.

[70] **N.B. (???) Test?**

have been (or have been!) discovered by examining the results of "mathematical experiments". Such experiments lead to questions very similar to our: "would the observed behavior continue?".

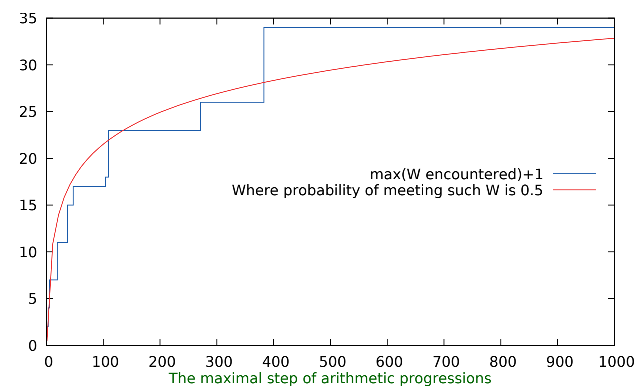
In a lot of cases there is no way (and no *heuristical* ways) to assign a "degree of certainty" to an experimentally observed pattern *to continue* for all values of parameters. (Of course, this excludes the case when one can *prove* this!) However, sometimes there is

- A lot of similar problems where the pattern *does break* at some moment;
- A heuristical probabilistic model describing the distribution of "the threshold where the pattern breaks".

Hence if the thresholds found by the "failing" experiments match the heuristical model, one can estimate "the heuristical probability" that the results of our — *apparently* successful — experiment are only a fluke (so the *observed* pattern is going to break later, beyond the range of our experiment).

This "heuristical probability" can be interpreted as "a degree of uncertainty" of our experiment. (This is close to the Bayesian approach to experiments.) Here we inspect a toy example: the experiments of the preceding section.

In the exercises above, one checks arithmetic progressions for "the color" of the primes in these progressions. Before we find a "good" arithmetic progression, we need to check many progressions which would "eventually fail"; in other words, we would eventually find a "prime of 'another' color'" inside this progression. In other words: at this moment, we inspected the first several primes in the progression, and found a few (denote this number W) primes of the same color, but the $W+1$st one is of different color.



How large may be W we would actually encounter? The answer to this question, obviously, affects our choice of "the poor man's infinity": it should be larger than any such number W. The plot on the right shows how this number grows when we allow progressions with larger and larger step in Exercise D8 on p. 28 (but, of course, we must exclude the step *found* in this exercise). One can see that these data can be approximated "reasonably well" by a red curve drawn according to a certain rule. Below, we examine how to guess a suitable rule for this curve. (Since 100 is much larger than the values on the blue and red graphs, *and* since the "sock model" below gives an exponential decay in probability w.r.t. this value, these plots substanciate our choice of 100 for "the poor man's infinity".)

Exercise E1: Calculate how many arithmetic progressions one should check to cover all $C \leq C_0$; here $C_0 = 400$.

Hint: Exercise D4 on p. 28

.

In fact, the answer is very similar to the answer at the end of Remark 41 on p. 73: we need to check about 61% of the pairs (r, S) with $0 < r < S < S_0$. See also the remarks at the end of the section on the average growth rate of φ on WikiPedia (and the group F of exercises on p. 30).

We continue discussing this after Exercise E3.

As mentioned in the Remarks considered above,[71] approximately 1/3 of primes p are going to be "red" (meaning that $p \nmid P_k$ for every k). How can this help us to estimate the maximum value of W after considering about 50,000 arithmetic progressions (as in Exercise E1)?

After we choose a polynomial, the color of every prime is uniquely determined by our rules. However, these rules are very unwindy to work with; very surprisingly, it turns out that *for a large*

[71] Compare also with Footnote 68 on p. 28.

category of questions there is no *radical* change of answers if we suppose instead that colors are assigned *randomly* (with expected frequencies)!

It is “reasonable to expect” that if ⅓ of socks are red (and the rest are green), and we take socks from a drawer at random, then “on average” we will pull 2 green socks before getting a red one.[72] However, the intuition tells us very little about how large this “run” of green socks may turn out if we try many times! We start with examining this question.

Exercise E2: Show that the probability of the first W socks being all green is $(2/3)^W$.

Exercise E3: Show that when we repeat “taking socks until you get a red one” 50,000 times, it is more probable than not that for every one of these sequences the initial run of green socks is less than 29 socks long.

To believe that these results are related to questions of color of primes in arithmetic progressions, we need to make the following two leaps of faith:

- When we consider 50,000 arithmetic progressions from Exercise E1, the color of prime numbers near the beginning behaves as if the colors were random, with probabilities as above.
- The colors in different arithmetic progressions behave as if they were independent.

(Here we interpret the word “random” as meaning what the Probability Theory calls “independent identically distributed”. The word “behave” means *only* the behaviour of the maximal number W we can see.)

Such statements are extremely hard to verify.[73] Nevertheless, it turns out that surprisingly often the corollaries of such “heuristical assumptions” are compatible with numeric experiments (e.g., as the plot above shows)!

For example, one can compare the value 29 obtained above with what is actually needed to avoid “false positives” in Exercise D8 on p. 28. The largest value W encountered for $C \leq 400$ is 33, which is larger than 29 — but it is not much larger (as promised above)! Moreover, the largest value W encountered for $C \leq 380$ is 26, which is smaller than 29.

Conclusion: a jump in the maximal observed value of W happens near the value of C_0 where the “sock model” gives a “significant” probability of encountering a larger value of W. (Here “significant” is of order of magnitude ½.)

Exercise E4: Using the results of preceding exercises, write a formula suitable for the red plot mentioned above.

Exercises F: Aside: Counting mutually prime numbers

Below we continue the discussion started immediately after Exercise E1 on p. 29. As in the preceding section, these exercises just clarify a phenomenon encountered in the exercises above. Nevertheless, the number 61% which appears below is mentioned once in the main body of the text, in Remark 41 on p. 73. Ⓜ These exercises examine routine “elementary calculus” questions. However, they are written so that the readers who do not know calculus have a chance to solve them.

Note that to remove all unsuitable pairs (r, C) in Exercise E1 on p. 29, one needs to remove those for which $2|r$ and $2|C$ (this is about ¼ of all pairs); then remove those for which $3|r$ and $3|C$ (this is about ⅑ of all pairs); then remove those for which $5|r$ and $5|C$ (this is about $1/25$ of all pairs); etc.

[72] Theoretically, with our description of the number W, we also need to consider the case of pulling out W red socks before pulling a green one. However, it is intuitively clear that this chance is *way* smaller than pulling out W non-red socks before pulling a red one (for $W \gg 1$). So we ignore this possibility below.

[73] … especially when we know that this is going to break for a certain larger value of C, when suitable $r \bmod C$ would give all-green sequences.

Exercise F1: Show that if the limit C_0 of Exercise E1 on p. 29 is large enough, then after doing 3 steps of the preceding paragraph, about $\alpha_5 := (1 - 1/2^2)(1 - 1/3^2)(1 - 1/5^2)$ of all pairs remain. In other words, if we start with S pairs, and R pairs remain, then $R/S \approx \alpha_5$, and the precision improves arbitrarily high when C_0 grows to ∞.

Hint: If one replaces 5 by 4, this is not going to be correct!

Exercise F2: Show that if numbers $0 < a_n < ½$ are summable (in other words, there is a number M such that $a_1 + a_2 + \dots a_n < M$ for any n), then $(1 - a_1)(1 - a_2)\dots(1 - a_n) > e^{-2M}$ for any n.

Hint: Use derivatives to show that $-2x < \log(1 - x) < 0$ for $0 < x < ½$.

Exercise F3: Show that $a_k + a_{k+1} + a_{k+2} + \dots + a_m < 1/(k - 1)$ for every $m \geq k > 1$ if $a_n := 1/n^2$.

Exercise F4: Show that the sequence $(a_n) := (1/n^2)$ is summable.

Exercise F5: Define numbers α_p as in Exercise F1: $\alpha_p := (1 - 1/2^2)(1 - 1/3^2)(1 - 1/5^2)\dots(1 - 1/p^2)$; here p is a prime number, and the factors $1 - 1/l^2$ are taken for all prime numbers $l \leq p$. Show that the numbers α_p have a certain limit α.

Exercise F6: Using prime numbers below 1,000, show that $0.605 < \alpha < 0.61$.

Hint: Use programs to calculate α_{997}; use the estimate from Exercise F2 on p. 31 to understand what happens with α_p/α_{997} for primes $p \geq 1{,}009$.

It is not very easy to show that the limit α from Exercise F5 coincides with the number we discuss after Exercise E1 on p. 29. However, Euler found a heuristic using the Taylor series for $\sin x$ to suggest that $\alpha = 6/\pi^2 \approx 0.6079271$. Furthermore, now we know how to make his arguments into real proofs. (Compare with this section on WikiPedia.) The original Euler's argument is in Polya's book "How to solve it".[74]

Example code to use in exercises

For a quadratic P (such as "a triangular number + 1") one can check what happens for $r \bmod C$ with $r = 4$, $C = 7$ in `GP/PARI` by

```
check_p(p,P)=for(k=1,p,if(0==subst(P,'x,k)%p,return(1)));0;
my(r=4,C=7,p);for(l=0,100,p=r+C*l;if(isprime(p),print([p,check_p(p,x*(x-1)/2+1)])))
```

To answer other exercises, one may use the following subroutines and use the example code as templates:

```
check_r(r,C,P,L=100)=my(p,c,cc=[0,0]);for(l=0,oo,p=r+C*l;if(isprime(p),\
  cc[1+check_p(p,P)]++;c++;if(cc[1]&&cc[2],return([cc,0]));if(c>=L,return([cc[2]!=0,1]))));
check_C(C,P,L=100)=my(l=List);for(r=1,C-1,if(gcd(r,C)==1,listput(~l,[r,check_rr(r,C,P,L)])));l;
check_C_count(C,P,L=100)=my(l=check_CC(C,P,L),c=0);for(j=1,#l,if(l[j][2][2],c++));c;
my(r);for(C=383,383,r=check_C_count(C,x*(x^2-1)/6+2,33);if(r,print([C,r]);return);if(C%100,,print(C)))
check_C_count_elevate(C,P,L,Lmax=100)=my(r,l=L);for(j=1,oo,r=check_C_count(C,P,l);if(r&&l<Lmax,l++,return([l,r])));
check_elevating(fr,to,LOG=1,Lmax=100)=my(L=1,l,r);for(C=fr,to,l=L;[L,r]=check_C_count_elevate(C,x*(x^2-1)/6+2,L,Lmax);\
  if(r,print([C,r]);return);if(L>l,print([C,l,L]));if(!LOG||C%100,,print(C))); \\ a few minutes to calculate up to 400
```

This `gnuplot` code produces the plot above (remove `set output` and `term` to output to the screen):

```
##### perl -wlne "s/[\[\]]//g; @F=split /,\s+/; print qq($prev\n$F[0] $F[1]) if $prev; $prev = qq($F[0] $F[2])" W-values > W-crv
set key right center
set xlabel "The maximal step of arithmetic progressions" offset 0,0.5 tc "dark-green"
set output "Wmax.pdf"
set term pdfcairo
plot "W-crv" w l title "max(W encountered)+1" lc "#1859a9", \
  log(x*(x+1)*6/pi**2)/log(3/2.) title "Where probability of meeting such W is 0.5" lc "#ed2d2e"
```

These commands use this input file `W-values` generated with the preceding `GP/PARI` code:

[74] **N.B. (???) A problem I forgot (with separate constants for each polynomial).**

```
1, 1, 1
2, 1, 2
3, 2, 4
5, 4, 7
19, 7, 11
37, 11, 15
47, 15, 17
104, 17, 18
109, 18, 23
271, 23, 26
383, 26, 34
1000, 34, 34
```

The simplest Langlands' patterns in more detail: the "hidden symmetries" in $\deg = 3$ are fractal

Bread crumbs: A very coarse outline of the Langlands' pattern

We did not discuss what follows at Math Circles.

On p. 20 we gave very vague hints about what one should be fluent with to be able to understand the Langlands' patterns (the "hidden symmetries") for our sequence of colors encoding divisors of "tetrahedral numbers + 2" (on p. 19). These patterns also fit other similar sequences of colors constructed, for example, from divisors of "$20 \times$ tetrahedral numbers + 1" (although the finer details for this example would be very different; we postpone them until Remark 18 on p. 36). Here we want to leave a tiny bit more bread crumbs on this path.

This section is just a very coarse outline. Later we are going to clarify the details.

While we tried to keep this outline as accessible as possible, there is a limit to this. Your mileage may vary. **All discussions below are heuristical only; it would take 100s of pages to give rigorous arguments.**

Exposing the pattern goes in 3 steps.

- First one needs to apply several "transliterations" to the colors. They are very straightforward, though the technical details are quite involved. To cut the long story short: in the outcome, we replace colors with "suitable" whole numbers.

  It is simplest to describe what happens to "bold" colors (colors of prime numbers): for sequences of degree 3, we replace red by -1; green becomes either 0 ("non-interesting green") or 2 ("interesting green"). (Which of the greens are "interesting" will be discussed later.[75]) Moreover, a few prime numbers[76] may need a special treatment.

  In fact, this is the step where we forget about colors of non-prime numbers: for example, the whole number assigned to pq does not depend on the color of pq, but only on the whole numbers assigned to p and to q.[77] We discuss this in more detail in the section on p. 59.
- Denote the resulting sequence of numbers by N_n. The second step is to take the Fourier transform of this sequence.[78] This is, automatically, a periodic function $F(t) = \sum_n N_n \cos nt$.[79]

---

[75] So, in fact, it is not "pure transliteration": we need a bit more information than our colors! However, the extra information is contained in what we already know: the color sequence corresponding to a certain polynomial of degree 2. For our example, it is "square numbers + 971" (this is not a misprint!). See Remark 40 on p. 72 for details.

[76] Divisors of the discriminant, of the denominators of coefficients, and of the numerator of the leading coefficient.

[77] For sequences of degree 2, already this first step exposes the pattern (so we do not perform the other two steps): the sequence N_n is periodic. In fact, we already saw this result (in disguise): it is the second row of colors on p. 8.

To unmask the disguise, note that in this case, the numbers N_n given by transliteration rules are either -1 or 1. For example, red or green primes p are replaced by $N_p = -1$ or $N_p = 1$ correspondingly. For composite n, one uses the rule $N_{ab} = N_a N_b$. (For more general polynomials, this works only if k, l are mutually prime.) Since N_n takes only two possible values, one can change N_n "back to" red/green colors. This makes it into a "double transliteration": "colors" $\to$ Numbers N_n $\to$ "colors". It turns out that it replaces our row of colors by a periodic row of colors (see p. 8). On prime n, the colors are automatically unchanged.

(Here we ignore "the exceptional primes" of the preceding footnote. They may lead to a mismatch between these two rows of colors in a few bold places.)

The obtained sequence N_n is called the "Legendre symbol". (Compare with p. 208.)

[78] Recall that these notes need only very cursory knowledge of Fourier series — see Footnote 2 on p. 1.

[79] A lot of things become simpler if we consider $F_{\mathbb{C}}(t) := \sum_n N_n e^{int}$ instead, so $F = \operatorname{Re} F_{\mathbb{C}}$. However, since until the discussion on p. 84 we are concerned mostly with plotting, it is much easier to ignore the imaginary part of $F_{\mathbb{C}}(t)$.

- At last, we can state how the Langlands program describes the pattern of numbers N_n. This goes through *fractal properties* of the function $F(t)$:[80]

$$\boxed{\text{The graph of } F(t) \text{ is an exact fractal.}}$$

Note that the word "fractal" is used in math with two different meanings:
  - A shape where every small part may be obtained from the whole by certain transformations, called "the fractality laws" (we call such a shape an "exact fractal").
  - A shape of fractional Hausdorff dimension.

  It is the first meaning which we need above, and "the fractality laws" play the role of "hidden symmetries".[82] Here is an example of such a fractal behavior of a graph from an about-15-years-old paper:

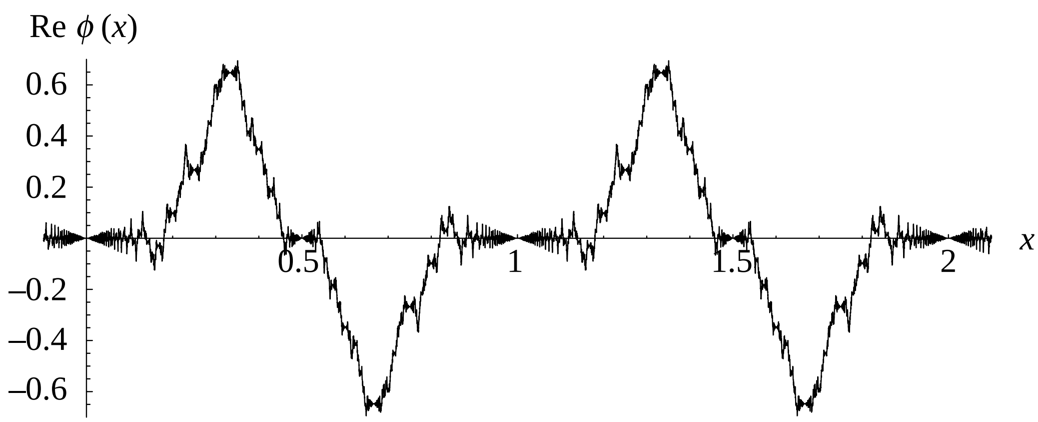


Figure 1. The real part of the antiderivative $\phi(x)$ of the automorphic distribution corresponding to the Mass form for $\mathrm{SL}(2,\mathbb{Z})$ with $\lambda \approx 27.56\, i$.

Note the pattern in the graph near $x = 0$. This pattern is in fact repeated near every point of this graph. The copy may be centered at any rational point $x = R/S$ — though the larger S is, the smaller is the copy (zooming into this graph can uncover many such copies corresponding to small S). Moreover, every "oscillation" of this pattern is, in fact, a particular "fractal transform" of the period of the graph (on this period x changes between 0 and 1).

Remark 14: Let us clarify in which sense the "fractal properties" above may be thought of as "a pattern in the sequence of numbers N_n" (or, transliterating back, as a "pattern of the sequence of colors"). If we know just "a very coarse overview" of the graph of $F(t)$, the fractality laws translate this information to "the coarse overview" of every small piece of this graph; combining these together, one gets "a much finer overview" of the graph. Repeating the process, one gets more and more details about $F(t)$. In a certain sense, the fractality laws "fill in" the information about the fine details of the graph which was missing in the overview.

So it should not be surprising that given the fractality laws and sufficiently many details of "the coarse overview", one can reconstruct *the whole graph* of $F(t)$. Since the "coarse overview" of a periodic function is given by its first few Fourier coefficients, it is natural to expect that

$$\boxed{\text{The fractality laws and the first few numbers } N_n \text{ determine all the numbers } N_n.}$$

[80] This may look very indirect as far as we are interested in numbers N_n — or red/green colors. However, first, this is expected to be "as good as it gets": probably, there is no pattern which is "more direct" than this. Second, currently mathematicians gradually learn how to extract "more useful" information about N_n out of such fractal properties.[81]

[81] This started with the circle method of Hardy–Littlewood.

[82] The corresponding law is the red-framed formula in Footnote 102 on p. 42.

And this is what actually happens![83] Moreover, this is exactly what one expects from "a sequence having a pattern": knowing "the type of the pattern"[84] and a few first terms, "the pattern" would allow us to reconstruct the rest of the sequence.

For example, for the graph above, it looks like all the "bumps" on the graph are fractality-law images of the "principal oscillation" on the graph. Then knowing the period (1) and amplitude (≈ 0.7) of the "principal oscillation" would allow one to find heights (and positions) of all the "bumps" on the graph, in effect reconstructing the whole graph.

We discuss how the regions where we "fill in the details" are positioned with respect to each other in the section on p. 87.

Remark 15: Note that to find whether a prime number p may be a divisor of numbers in our sequence of degree 3, it is enough to calculate the whole number N_p. On the other hand, if we know $F(t)$ then N_p is just a certain integral (the "inverse Fourier transform") involving $F(t)$ and p.

This shows that the questions of divisibility are inherently related to the questions of calculus.[85]

Remark 16: In discussions on the future (and history) of science, the prevailing mood is to claim that science becomes more and more fractured, so that even specialists in relatively similar areas cannot understand each other. Nevertheless, many leading mathematicians champion the exactly opposite point of view.

Yes, if one observes what happens on the bleeding edge of science *now*, one would see that people may focus on quite narrow questions. However, there is nothing new in this — this is the natural way the human mind works. Moreover, such narrow interests might be just "tactical" in nature, and such a close focus can be temporary only. (This is the *synchronous view* on science.)

On the other hand, the *diachronous view* would show a completely different perspective. Instead of looking at what people thought about what was "the bleeding-edge research" *at that particular moment of time*, this point of view focuses on a particular theme, and observes how it was perceived at *different* moments of time, from the time it was "bleeding-edge" till today. It turns out that as time goes we understand more and more the interrelations of these themes. What may have looked "very specific and narrow" when it was discovered, later would turn out to be included in wider and wider vistas. New points of view appear all the time; they interconnect things which were previously thought to be *completely* dissimilar.

This confluence of mathematical theories leads to the idea of "Unity of mathematics".[86] Remark 15 on p. 35 provides one of the most striking examples of such a unity.

Remark 17: While "Unity of mathematics" is a very captivating phenomenon, it may also lead to hard-to-surmount difficulties. *This* is what happens with the Langlands Program!

It brings together a dazzling amount of very different branches of contemporary mathematics. Even if one could make an intelligible sketch of every one of these themes, the sheer count of the involved topics would overwhelm all but the most persisting readers.

To cope with this, we go over the same ideas in several passes, trying to increase the amount of details gradually. Additionally, inside every pass we attempt to use strokes as bold as possible, cloaking all the "fine print" into footnotes, and interconnecting[87] the passes by cross-references.

[83] After explanations above, it should not be too surprising. *What is* surprising is that all this "filling in of details" does not lead to contradictions. In other words, the existence of *any* non-0 function satisfying the fractality laws is an amazing miracle!

[84] For example, in the case of the pattern of periodicity, the "type" is the length of the period. If we know that many first numbers N_n in a periodic sequence, the rest may be reconstructed by periodicity.

[85] Moreover, the famous circle method of Hardy–Littlewood is based on a very similar observation. Compare with Footnote 80 on p. 34.

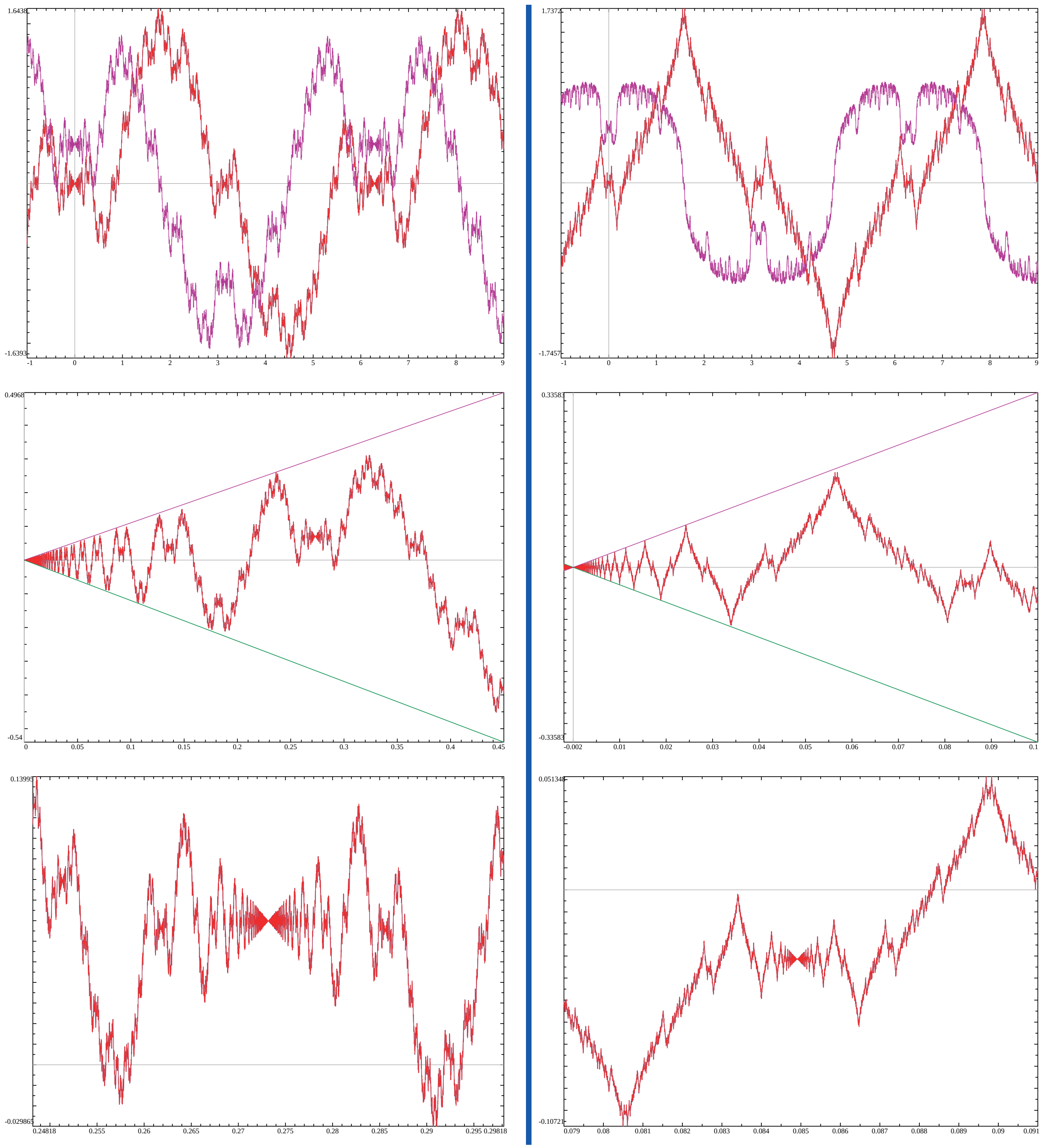
[86] … although for most mathematicians, maturing to this idea takes much longer than it took I. M. Gelfand in an epigraph to these notes!

[87] … as sparsely as possible, to avoid making these notes into Borges' Ts'ui Pên's *The Garden of Forking Paths*.

The appetizers for what follows

We continue laying the bread crumbs on the way to the Langlands Program. This is still just a very coarse outline of hidden symmetries (in degree 3)!

Remark 18: As an appetizer for the following discussion, here are the "real-life examples" of two types of behaviour of plots of functions related to our sequences of colors for polynomials of degree 3:



For each of two columns above, we picked a polynomial of the corresponding type for which the patterns of fractality are easiest to recognize.[88] Each plot in the top row shows two graphs: about 1½ periods for the real and the imaginary part of the function $F_{\mathbb{C}}^{(-1)}(t)$ (see Footnote 79 on p. 33). The

[88] Mathematically, this means that the conductor is as small as possible.

second row zooms into the red graph of the graph above it near the origin; the third row zooms yet more into the plot above it near its most interesting point.

One can see that the "shape of oscillation" in the second row matches the period in the first row — but on the left it matches the violet shape, while on the right it matches the red shape. However, for both columns, the "shape of oscillation" in the bottom row matches the *red shape* of the top row.

This difference between these two columns suggests that one may need to consider two different flavors of fractality — and this is what actually happens. By historical reasons, in math these flavors are called by unrelated names: "modular form" fractality, and "Maass form" fractality. (Due to harder-to-explain mathematical arguments of the Langlands Program, nowadays they are also called "the odd case" — on the left, — and "the even case".)[89]

The "odd" case was understood a few decades before Langlands — but before the Langlands Program it was just a mathematical curiosity. The investigations of the "even" case succeeded only very recently.[90] We examine another approach to these two cases in Remark 24 on p. 39. See also Remark 36 on p. 59, and the section on p. 82.

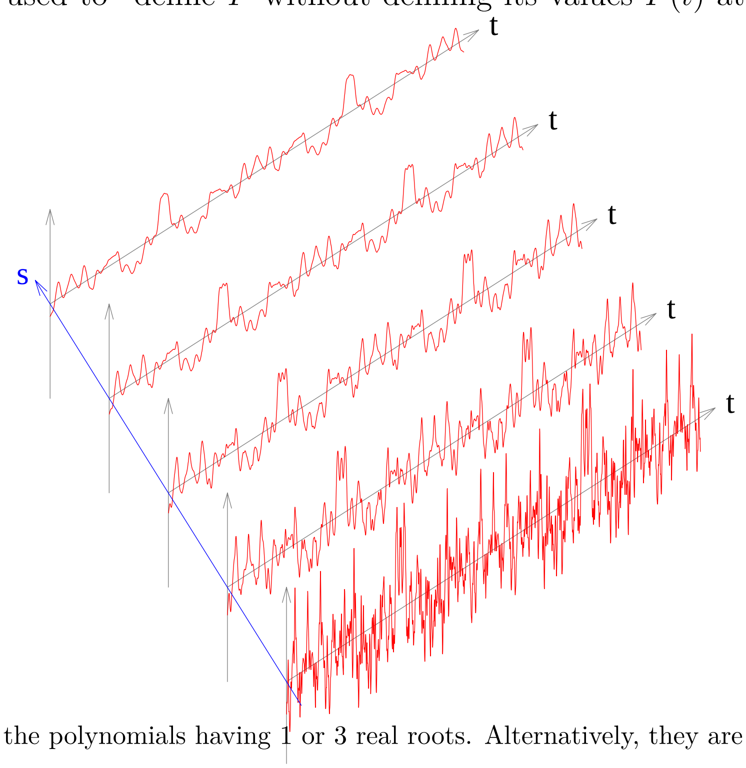
Remark 19: In the outline above, we needed to cheat to circumvent certain delicate points. Note that the graph above, on p. 34, plots not the function $F(t)$, but its antiderivative $F^{(-1)}(t)$. (Same for plots of Remark 18 on p. 36.)

The reason for this is that, in a naive sense, the function $F(t)$ has no value anywhere: the Fourier series defining $F(t)$ diverges for every value of t. In particular, the graph of $F(t)$ itself does not make a lot of sense. However, the antiderivative of $F(t)$ has "a much milder" Fourier series; and it has an honestly defined graph. (Note how this is similar to the relation of "white noise" and "Brownian motion" — see Remark 32 on p. 48.)

Essentially, the phrase "the fractal properties of the graph of $F(t)$" should be understood as a metaphor. To proceed any further, one needs to assign a precise meaning to this metaphor. There are two approaches to "infinities" which are used to "define F without defining its values $F(t)$ at particular values of t".

Remark 20: One approach provides ways to work with these infinities directly. This has immediate advantages of "visually obvious" fractality (see the plots above — and below). It also helps to internalize why the fractality laws allow a few initial values of N_n to define the rest of values of N_n, as we discussed in Remark 14 on p. 34. (See the section on p. 48 for details.)

The plots of functions shown above (and those below!) are results of application of this approach.



Remark 21: The other approach "regularizes" the infinities away altogether. Here "regularization" means a particular way to

[89] In elementary terms, these cases correspond to the polynomials having 1 or 3 real roots. Alternatively, they are the cases of a negative or positive discriminant.

[90] This is just my reconstruction — I could not find any appropriate reference.

It looks like during the last couple of decades, there is a widespread understanding that "this follows directly" from what is already proven about the Langlands Program. — However, apparently, nobody wrote this statement down explicitly.

morph a function which makes it "more smooth". In fact, this morphing process can be applied repeatedly (as done above). So one can "regularize" with different "strength"; the "strength" parameter s shows how many steps of "morphing" were used. Moreover, interpolation is possible, so the parameter s may be fractional as well.

Start with the function $F(t)$ and apply regularization with strength s; this leads to a function of two variables $f(t,s)$ (as on the plots above). For more details, see the section on p. 84.

(In fact, these pictures[91] illustrate a repeated application to $F(t)$ of a certain type of low-pass filtering with lower and lower cut-off frequency $1/s$. Compare with the discussion in Remark 32 on p. 48.)

Remark 22: At first, the fact that we need to work with a function of 2 variables may be seen as an inconvenience. On the other hand, with 2 variables one gets many more possibilities in interpreting *what these variables mean.* In particular, while all geometries with 1 degree of freedom are essentially the same, with 2 degrees of freedom a new opportunity appears: some of these geometries are "curved" (somewhat similar to how the geometry of the surface of Earth is "curved").

It turns out that

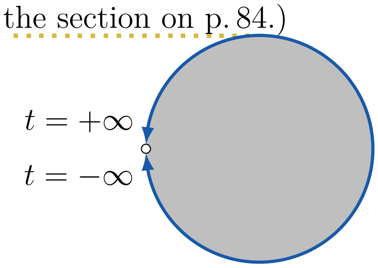
- If one chooses a "suitable" way to regularize, and
- if one chooses a "suitable" curved 2-dimensional geometry,

then the transformations of the fractality law for $F(t)$ become just "rotations" (or "shifts") in this curved plane of parameters (t,s). One gets the following translation rules:

$$\boxed{\text{Fractality laws for } F(t)} \longleftrightarrow \boxed{\text{Rotational/Translational symmetries for } f(t,s).}$$

These rules show that "the fractality laws" *are* symmetries, explaining how they can play the role of "hidden symmetries".

Moreover, the "rotations" (or "shifts") in question happen to be symmetries of a tessellation (or tiling) of this *"Lobachevsky" geometry*. (We return to this topic later, in the section on p. 84.)



Remark 23: In the second approach, the domain of definition of the function $F(t)$ becomes "the absolute", or the "horizon" of the curved geometry.[92] A point t of the absolute encodes "the azimuth" φ of the direction going to this point (the encoding is similar to the rule $t = \tan\frac{\varphi}{2}$ in the usual geometry which sends $(-\pi,\pi)$ to $(-\infty,\infty)$).[93] In Remark 22 on p. 38, we worked with a point of Lobachevsky geometry writing it as

[91] For technical reasons, these plots are based not on our function $F(t)$, but on a function $\Phi(t)$ with *random* Fourier coefficients of approximately the same magnitude as for $F(t)$. However, since "the degree of smoothness" of a graph depends on how quickly the Fourier coefficients decrease, "the roughness" of these graphs is very similar to the graphs for $f(t,s)$. (However, because of randomness, $\Phi(t)$ allows no fractality laws.)

To unclutter the picture, we avoid small values of s: they would result in very high spikes; these spikes would ruin the plots. Above, s changes in $[0.015\ldots0.095]$, while t changes in $[0\ldots14]$.

In fact, the scales of variables s and t are closely interconnected (see Remark 22). Us using a very different scale for s means that we scaled s up about 200 times. So the plots illustrate what happens in *a very narrow strip* near the line $\{s=0\}$.

[92] A point of the absolute is "a point at the 'infinity' of the geometry"; different points of the absolute correspond to different *azimuths*: "directions to look at" (this assumes that we look at something "very far" away).

This notion works equally well in non-curved (Euclidean) and in Lobachevsky geometries. While each observer living in this geometry would have their own coordinate system for "azimuths", what is crucial for existence of the absolute is that if two cowboys ride "to infinity", and their azimuths become closer and closer for one observer, the same would happen with any other observer. So a particular value of azimuth for one observer "matches" a certain value of azimuth for another observer.

This identifies "the absolute" with something observant-independent. We do not want to reuse the word "horizon" in this context since we need it below in the non-curved situation.

[93] People familiar with the *Stereographic Projection* may recognize the significance of this formula.

(t, s) with $s > 0$. This is the half-plane model of this geometry; however, there is another, equally useful model in a disk,[94] where the absolute is the circle which is the boundary of this disk. One point of this circle matches $t = \infty$, the rest is identified with the t-axis.

The pairs of numbers (t, s), $s > 0$, used above are coordinates on a half-plane. However, they may be also thought of as curvilinear coordinates in this disk; very vaguely speaking, s corresponds to how far away from the boundary is the point. In particular, points with $s = 0$ are on the absolute, matching the setup of Remark 22 on p. 38.[95] Moreover,

Regularization $F(t) \to f(t, s)$ is the interpolation of $F(t)$ from the boundary to the inner part of the Lobachevsky disk.

Assume that $F(t)$ describes "the temperature on the absolute". In other words, $F(t)$ is the temperature "far away in the direction encoded by t". Keep this temperature on the boundary steady, and let the temperature inside the Lobachevsky plane "settle down", eventually reaching a steady state. What may be the distribution of temperature in this state of stable equilibrium? The answer to this question turns out to be exactly our choice of $f(t, s)$.

In this language, $f(t, s)$ is "the steady-state-heat-propagation interpolation" of $F(t)$ from the boundary of the unit disk into the whole disk. Moreover, $F(t)$ may be interpreted as the "boundary trace" of $f(t, s)$. Hence, when the description above is applicable, one gets *an "intertwining" compatibility rule*:

If the function $F(t)$ on the boundary has a symmetry, then its interpolation $f(t, s)$ has a "similar" symmetry.

and vice versa.

Moreover, it turns out that "fractality laws" for $F(t)$ may be considered as such symmetries. This shows that the fractality laws are indeed "hidden symmetries" we have been looking for:

If $F(t)$ is an exact fractal, then $f(t, s)$ is highly symmetrical.

(And vice versa.) *This* is the reason for the rules from Remark 22 on p. 38.

Remark 24: In the preceding remark, we hid a very important effect: it turns out that the ordinary process of heat propagation in our familiar non-curved geometry has *two* analogues in the case of curved geometry. Some of the features of steady-state temperature distributions in our "flat" geometry are inherited by one analogue, while some other features are inherited by the other.[96]

These two different analogues of the heat transfer process lead to *two different choices* of the interpolation $f(t, s)$ of $F(t)$ into the disk.

Compare this with two flavors of "fractality laws" mentioned in Remark 18 on p. 36. It so happens that one of them is compatible (in the sense of preceding section) with one type of heat transfer, while the other one is compatible with the other type. This way, modular/Maass forms corresponds to different kinds of heat propagation in a curved geometry.[97] We illustrate this in the section on p. 55.

---

[94] There is no best way to visualize this curved geometry. Sometimes the half-plane model ((t, s) with $s > 0$) used in Remark 22 on p. 38 is more convenient; sometimes the disk model.

[95] We illustrate these coordinates on p. 85.

[96] This is, eventually, related to so-called "non-*amenability*": the area of the circle in this curved geometry grows exponentially with its radius. Therefore, even if you heated a part of radius 999 of a disk of radius 1,000, when this heat propagates to the whole disk, the temperature would drop several times.

Essentially, all our intuition breaks in this case. Mathematically, this corresponds to appearence of a "spectral gap" for the heat propagation operator. One analogue of heat propagation "ignores" this gap, the other analogue introduces a new term cancelling this gap.

[97] In fact, this is how these forms were first discovered: not on the absolute, but on the Lobachevsky plane.

Remark 25: We must stress out that what people *recognize* as exact fractals are the fractals "optimized for beauty". When repeated due to fractality laws, the features of such shapes can remain sufficiently large to be immediately recognizable. *This* makes these shapes attractive enough to be put on a wall.

Unfortunately, most (or all?) examples of fractals in these notes are not "beautiful" in the above sense. One *needs* to zoom in to recognize repetition of features. In fact, the smallest needed zoom ratio is the conductor — and there are no polynomials of degree 3 with a small conductor!

However, even if not "beautiful enough to be put on a wall", exact fractals remain exact fractals. While the pictures below require zooming in to see the self-similarities, mathematically, they are on equal footing with "beautiful fractals".

Remark 26: For example, the plot above, on p. 34, *is* optimized for beauty: it has conductor 1. To achieve this, the authors used a certain "tuning parameter" λ (mentioned in the caption to the plot).[98] In our context λ must be 0. In fact, they took the smallest $|\lambda|$ allowing conductor 1.

In more detail

The exposition of the previous two sections was intentionally made very sketchy, to avoid drowning the reader in excessive details. I expect that for many readers, already the level of details in the sketches above may be an overkill — and then here is a good place to stop reading.

On the other hand, the rest of this report is written for people left unsatisfied by *the vagueness* of the preceding exposition. From this point on, the notes are going to become way more technical.

Anyway, to make the level of difficulty raise as slow as possible, we start with topics which allow a "more visual" presentation, and would postpone "dry algebraic" themes for as long as possible.

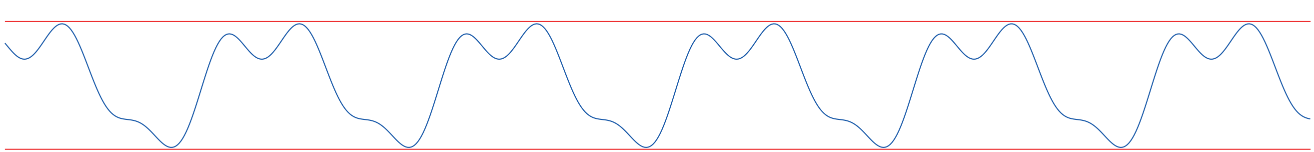
Unfortunately, the usual way the Langlands program is stated is extremely technical and very far removed from the simplified point of view discussed above. Translation to down-to-earth terms is error-prone if one is not a specialist; on the other hand, there are very few published attempts to do this — and all the attempts I know cover just the cases of negative discriminant (such as "tetrahedral numbers + 2"), which were, in fact, understood well before Langlands. (Compare with Remark 18 on p. 36.)

The (pseudo-)exposition we did in class (and do in these notes) is based on scratches of information extracted from "the attempts mentioned above" combined with what I could distill from the original papers. As I said, this is an error-prone process; apply salt as needed — one grain may be not enough.

Fractality laws: the simplified example

The first thing we want to describe more precisely is the "fractal transformations". Recall that these transformations map the whole graph of the function to its small parts. In fact, we want to start with a "toy example": it does not match "the actual transformation" exactly, but is its very close cousin.

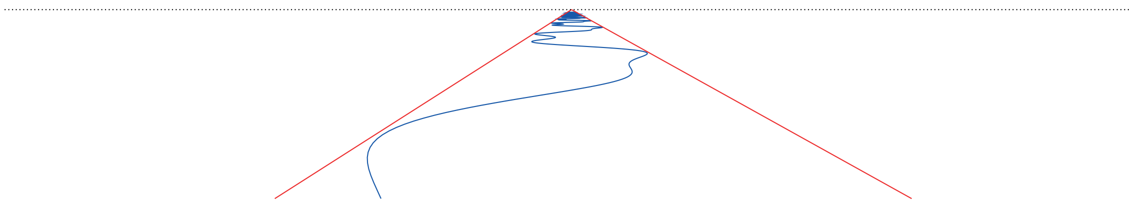
Take a graph of a periodic function $g(t)$:



squeezed between two horizontal lines. The graph continues forever to the left and to the right; image it drawn on a horizontal floor, and look at this graph from above. When our gaze follows the graph

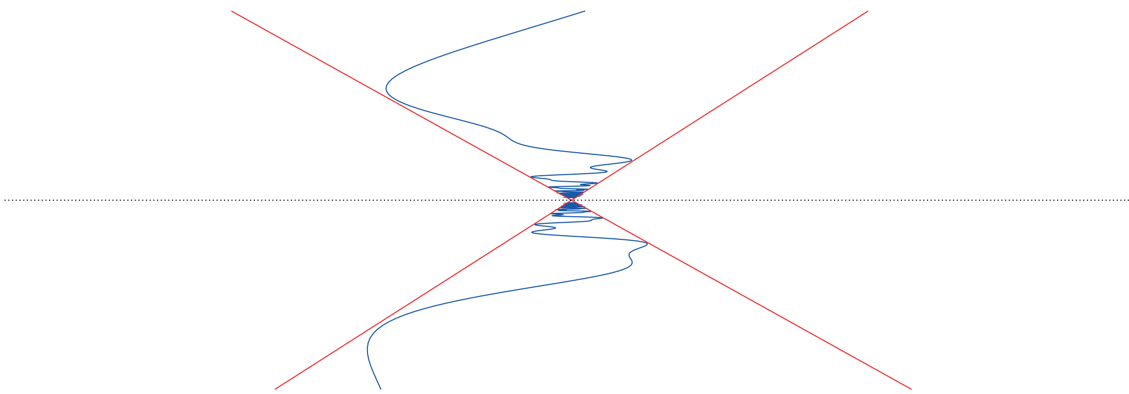
[98] This number is related to the eigenvalue for this eigenfunction of the heat transfer operator.

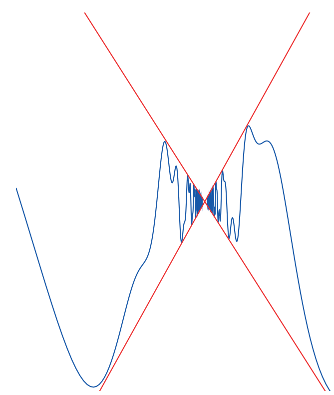
to the horizon, near the horizon we see the picture like this:



Here two red lines "converge" near the horizon like rails of a straight railroad.

As it is customary done in "Projective Geometry", above the horizon we put the reflected picture of what is "behind us":





Note that rotating this, we can make it into a graph of a function (on the right). And *this* is the transformation we had in mind.

As it is easy to see, given any periodic function $g(t)$, the graph on the right is the graph of the function $tg(-1/t)$. (This assumes that the intersection of the red lines is the origin.)[99] Call this *the toy transformation* of the graph of $g(t)$.

With this transformation defined, we may state the required "toy fractal property" of the graph of the function $F(t)$: (after appropriate rescaling and horizontal shift) every small piece of the graph of $F(t)$ coincides with the "toy-transformed" graph of F (the graph of $tF(-1/t)$).[100]

More precisely: In fact, even more is true. Shift the graph of $g(t)$ so that a point P of the graph moves to the origin. Suppose that there is a periodic function $g_P(t)$ such that the shifted graph is the "toy-transformed" graph of $g_P(t)$. We say that near P, the graph of $g(t)$ is *horizon-similar* to $g_P(t)$.[101]

The periodicity of $g_P(t)$ is already an extremely strong condition on the graph of g. For the function $g(t) = F(t)$, it holds for any P whose t-coordinate is $t = 2\pi R/S$ with whole numbers R, S. Furthermore, the exact-fractality property can be restated as this amplification: "for many" such points P, the function $g_P(t)$ is "a shifted and rescaled" function $g(t)$ itself. In other words, $g_P(t) = A_P g(B_P t + C_P)$. Such points P appear arbitrarily close to

[99] Moving our "observation point", one can also get functions $tG(-1/t)$ with $G(t) = Ag(Bt + C)$.

[100] What we said above is a simplification; in fact, instead of applying this law to the graph of $F(t)$, it should be applied to the graph of $1/F(t)$.

This may be restated as follows: one should apply not the "toy transformation" $tF(-1/t)$, but the "actual transformation" $F(-1/t)/t$. (*This* restatement is applicable even though $1/F$ does not make sense for "white-noise-like" generalized functions F we consider in our notes. Compare with Footnote 111 on p. 47.)

[101] In other words, $g(t + 2\pi R/S) = tg_{2\pi R/S}(1/t)$.

any given point of the graph. Which particular points of the form $t = 2\pi^R/S$ "work this way" is determined by the conductor; call them "*horizon-self-similar points*".[102]

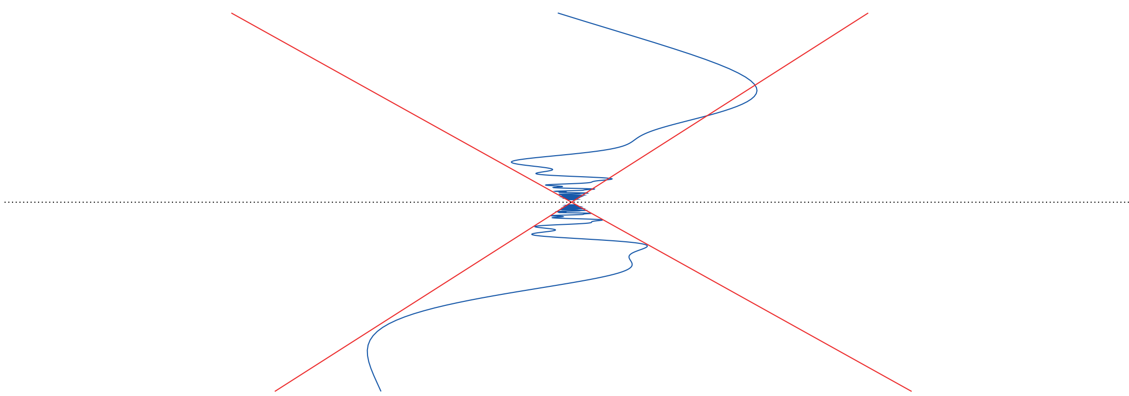
Remark 27: Above, what we did "above the horizon" looks very logical — provided one knows projective geometry. Indeed, when we look in some direction, our gaze "hits" everything on the half-line starting at our pupil, and going in the direction we look at. Now, half-lines are not very natural geometric objects; a projective geometer would try to replace them with whole lines.

After such a replacement, we imagine that we "can see" not only along the "forward" half-line, but also along "backward" one. How would it play out in practice?

When one looks above horizon, there is nothing along the "forward" half-line, but along the "backward" half-line one "can see" the objects hit by the "backwards continuation of our gaze" — which are below the horizon! So the objects on the ground *behind us* "would appear" above horizon in front of us. (This is the central symmetry with the fixed point in our pupil.) This is exactly how we plotted the illustration above. In turn, this led us to the fractality law $tF(-1/t)$ stated above.

For many years, this law was known to "provide" the pattern in sequences of colors considered above, at least for *some* of polynomials of degree 3 (those of negative discriminant, see Remark 18 on p. 36). On the other hand, a lot of polynomials were not covered by this kind of fractality.

Eventually, due to the Langlands program, it was understood that to cover these "remaining" cases, we need to change what we do "above the horizon". There is *another way* to attach the top part of the picture above: reflect it flipping left and right:



This way, the "reflected" "toy" fractality law sends the graph of $F(t)$ to the graph of $|t|F(-1/t)$, and the "reflected" "actual" fractality law sends it to $F(-1/t)/|t|$.

These absolute values are very unnatural, almost sores in the eye — but this is what turned out to actually work (in the cases of positive discriminant; see Remark 18 on p. 36). The contrast between having t and $|t|$ leads to the difference of the graphs in Remark 18 on p. 36: the right one needs $|t|$.

[102] Using the formula from Footnote 100, horizon-similarity "to $G(t)$" at $t = 0$ means:

$$\boxed{F(t) = \varepsilon \cdot G(-{}^1/_{\gamma t})/t. \qquad \text{Alternatively: } tF(t) = \varepsilon \cdot G(-{}^1/_{\gamma t}).}$$

With "self-similarity" $G = F$. Likewise, horizon-self-similarity at $t = 2\pi^R/S$ can be written as

$$\boxed{tF(t + 2\pi^R/S) = \varepsilon \cdot F(\zeta - {}^1/_{\gamma t}).}$$

for certain constants ε, ζ, and γ. (Note that the relation between the arguments of F on the right and on the left coincides with what is described in the section on p. 78.)

Below, we illustrate the notions of "horizon-similar" and "horizon-self-similar" with many plots.[103]

We quantify the notion of "many such points P" in the sections on p. 79, p. 90, p. 96. Moreover, the possible values of R and S — and the corresponding ζ and γ — are described in Footnote 201 on p. 78.

[103] However, since what we plot is $F^{(-1)}(t)$ we need the transformation law for the antiderivative of F. For visual comparison, it turns out to be very similar to the *toy* law! (See the section on p. 42.)

The zoo of fractality laws

Let us collect together the fractality laws we use in these notes:

The **“actual”** fractality law

$F(t) \rightsquigarrow F(-1/t)/t$
or $tF(t) \rightsquigarrow F(-1/t)$

From Footnotes 100, 102.

$F = 1/g$

Antiderivative
$\Phi = F^{(-1)}$

The **“toy”** fractality law

$g(t) \rightsquigarrow tg(-1/t)$

Illustrated by the pictures above.

The **“honest law for antiderivative”**

$\Phi(t) \rightsquigarrow t\Phi(-1/t) +$ **extra term**

Introduced in the section on p. 49

Visually very similar (see the plot on p. 81)

(The **extra term** comes from integration by parts. See the calculation on p. 81 for details.)

Moreover, every one of these laws comes in two flavors: one is as above, the other has $|t|$ instead of t as a factor or a denominator.

It is the “actual” (or the “honest”) fractality laws which are “the hidden symmetries” of the Langlands Program. In these notes, we play with the “toy” fractality law *only* for instructive purposes, because

- It is so simple to deal with.
- It has a very strong visual similarity to the “honest” fractality law.
- The dashed connection above ($F = 1/g$) permitted us to quickly *introduce* the “actual” fractality law (in Footnote 100 on p. 41).[104]

Recall (see the section on p. 33) that we are interested in particular (generalized) functions $F(t)$: the Fourier transforms of “arithmetic” sequences N_n. The main message of these notes is that these functions satisfy the “actual” fractality law. However, the graphs of functions $F(t)$ turn out to be “unplottable” (see the section on p. 48), and the best choice we have is to plot their antiderivatives; in this context the blue arrow above leads to the “honest” fractality laws.

Finally, the “honest” law leads to pictures practically indistinguishable from those of the “toy” law — hence the features of such “fractal plots” are easy to recognize. The only important difference is that the “extra term” can move these features up or down on the graph. (See the section on p. 81 for details.)

Example: the toy fractality law as a symmetry

Now we want to demonstrate how the “toy” transformation discussed above works as a part of a fractality law. We want to simplify the situation above yet more so that we may discuss a handy example. With this in mind, replace the property stated on p. 41 before Remark 27 by a much weaker property:

The origin $P = (0, 0)$ of the graph is “horizon-similar” to the function itself.

(So, first, we require horizon-self-similarity near one point P only. Second, we do not need to shift the graph.)

[104] Recall that the particular function $F(t)$ we study *cannot* be written as $1/g(t)$. Hence the dashed connection above is again “didactic only”. (Compare with Footnote 111 on p. 47.)

In other words:

$$\boxed{\text{The graph is a rescaled toy transformation of itself.}}$$

Can this happen with a periodic function? Since the typical gut reaction to this question was: "this is not possible", we start with an example of such a graph.

The idea of our construction is very simple:

$$\boxed{\textit{Force}\text{ the graph to be preserved by toy transform, }\textit{and force}\text{ periodicity.}}$$

Forcing preservation by toy transform is easy: keep the given definition of the function far from 0, and define it near 0 by the formula for the toy transform. Likewise forcing periodicity is easy: one can extend any function on $[-\pi, \pi]$ periodically.[105] We are going to apply these two steps alternatingly, and see what happens.

So we start with a smooth function $g_0(t)$, then define $g_1(t)$ as $tg_0(-1/t)$ on $[-\pi/2, \pi/2]$, and extend periodically so that the shift $g_1(t + {}^{\pi}\!/\!_2)$ of the resulting function $g_1(t)$ is even. Then we get $g_2(t)$ likewise, etc.[106] Every next function would have "a thicker pool" of non-smooth points than the previous one.

Very quickly (for plotting purposes, it reaches the limit already about $n = 8$) the process above leads to a sequence of functions flipping between 4 states. Essentially, $g_{n+2}(t)$ almost coincides with $-g_n(t)$ when $n \gg 0$. In other words, putting $G(t) := g_n(t) - \mathrm{i}g_{n+1}(t)$ with $n \gg 0$ gives a function

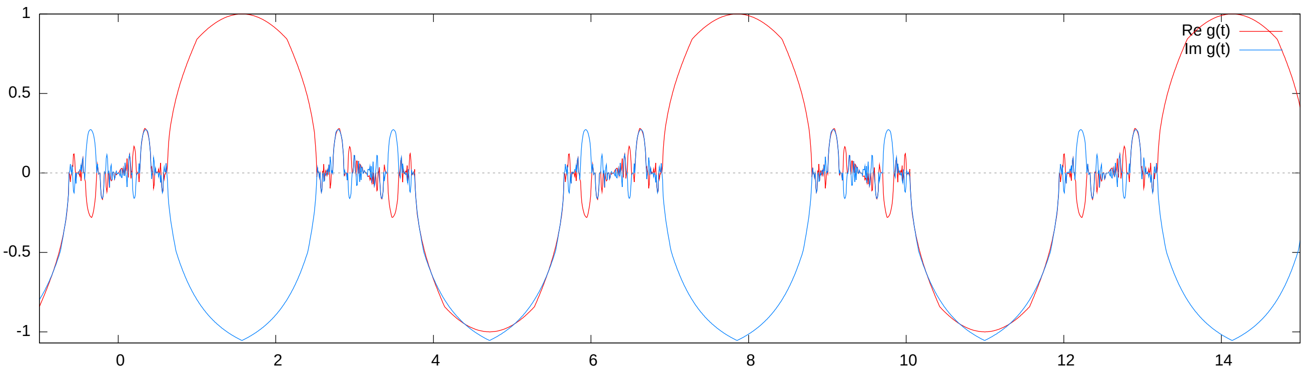
105 For continuity, it is better to start with $[-{}^{\pi}\!/\!_2, {}^{\pi}\!/\!_2]$, then extend it to $[-{}^{\pi}\!/\!_2, {}^{3\pi}\!/\!_2]$ so that $f(t + {}^{\pi}\!/\!_2)$ is even. *Then* one can extend from $[-{}^{\pi}\!/\!_2, {}^{3\pi}\!/\!_2]$ by 2π-periodicity.

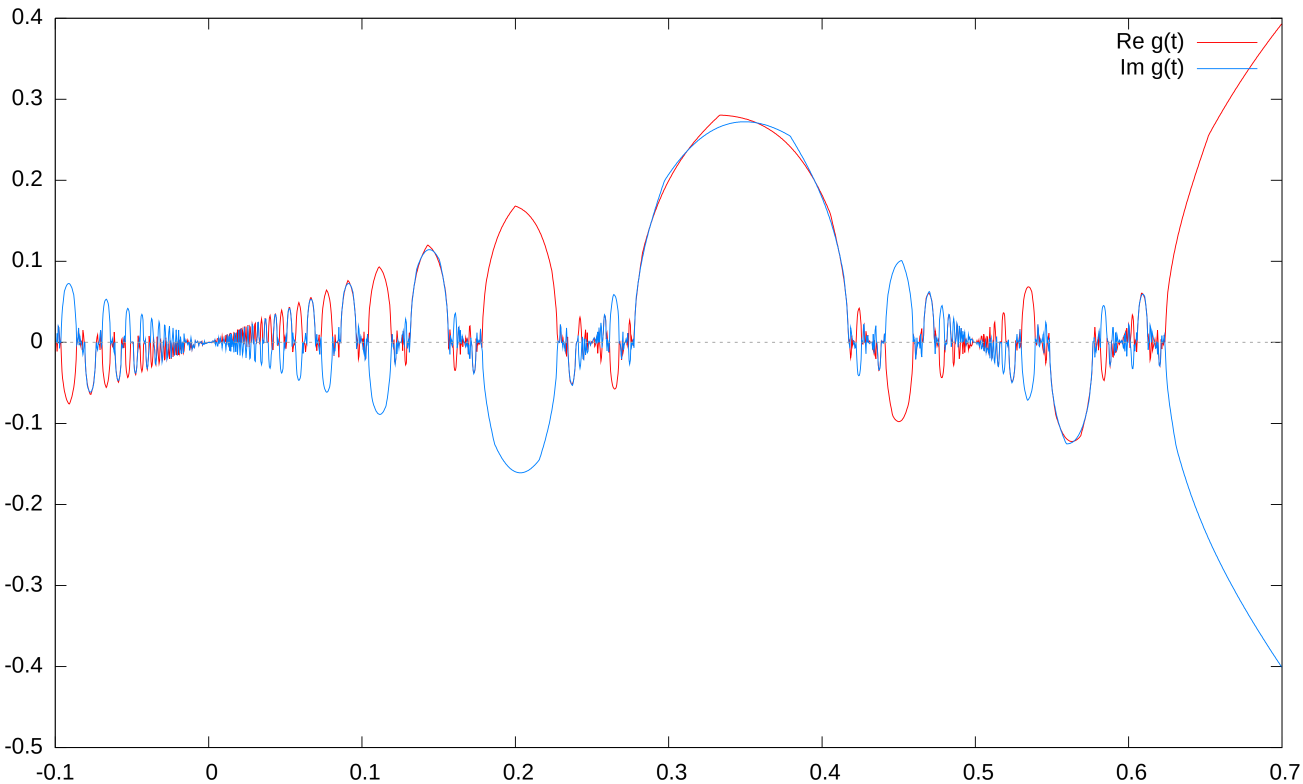
This is what we do below. However, to improve the visibility of the pattern, we rescale the t-axis; essentially, we use $g_{k+1}(t) := tg_k(-C/t)/\sqrt{C}$ with $C = \pi/2$. (This particular choice has no significance except for $t = 0.5$ being non-smooth.)

(Note that this creates discontinuity of derivative at $t = {}^{\pi}\!/\!_2$. With a bit more ingenuity one could extend avoiding this discontinuity. Instead, we are going to just ignore this defect.)

106 However, to improve the visibility of the pattern, we rescale the t-axis; essentially, we use $g_{k+1}(t) := tg_k(-C/t)/\sqrt{C}$ with $C = \pi/2$. (This particular choice has no significance except for $t = 0.5$ being non-smooth.)

such that $tG(-1/t)$ is $\mathrm{i}G(t)$; in other words, it is $G(t)$ rescaled by the imaginary unit i. Observe:





The first graph plots a bit more than 2 periods of this function. The second shows a small part of its period.[107]

The Cantor set of non-smooth points on the example plot

Here we continue inspecting what happens if a periodic function $G(t)$ is symmetrical w.r.t. the toy fractality law.

Automatically, the graph of $G(t)$ near the origin looks "at least as bad" as the graph on p. 41 used in the definition of the toy transform. In fact, it *must* be much worse! That graph was a "toy transformation" of a smooth function $g(t)$ — and this transformation had a "very non-smooth point"

[107] It should not be "very surprising" that we obtained a complex-valued function. Recall that above we promised that $F_{\mathbb{C}}(t)$ is easier to deal with than $F(t)$. Indeed, $F_{\mathbb{C}}(t)$ has "better" fractal properties than $F(t)$ — and it takes complex values.

The simplification comes from the fact that we allow our fractal transform to rescale the function — and there are "more ways to rescale" a complex number than a real number. For example, one can multiply it by i. (Algebraically, appearence of i is unevitable since the toy transform chained with itself sends $G(t)$ to $-G(t)$.)

with oscillating behavior near the origin.[108] However, as the example graph of the preceding section shows, there is a process "proliferating" already known non-smooth points.

Proceed marking the already known non-smooth points in color:

- The origin is a "bad" ("non-smooth") point for $G(t)$:

- By periodicity, G has many such "non-smooth" points going to infinity:

- Since the graph of $G(t)$ near the origin is a "toy transformation" of such a *non-smooth* graph of $G(t)$, these non-smooth points "near the horizon" (so when $t \approx \infty$) are transformed to non-smooth points accumulating near the origin:

- Now take the periodicity into account again: this red "family" of non-smooth points near the origin must be repeated near every blue point:

- Use the toy transform *again*. The red points near the origin were "toy transforms" of the blue points. However, now every blue point is surrounded by "a red family". So every red point near the origin must be surrounded by a (tiny!) "toy transform" of the red family near the corresponding blue point; draw this in green. Here we zoom about 10 times near the origin:

  Together, the red and green points accumulating at the origin form a "super-family". (The origin is surrounded by red points, and every red point is surrounded by green points.)
- By periodicity, there is a repetition of this super-family near every blue point.
- Time to use the toy transform again! Since every green point near the origin is a toy transform of a red point, *and* now we know that every red point is surrounded by points of a super-family, near every green point there is a toy transform image of this super-family. This forms "a super-duper-family". (The origin is surrounded by red points, every red point is surrounded by green points, and every green point is surrounded by its own family.)

  Etc.

Conclusion: every non-smooth point of the graph of $G(t)$ is surrounded by a whole "pool" of non-smooth points. Taken together, these points form an exact fractal. Call it *the Cantor hyper-family*.[109]

Warning: do not confuse the exact fractality of this set with exact fractality of the graph of $F(t)$. *This* fractal is formed by the *arguments* t of the function $G(t)$ where it has singularities (so it is a fractal in dimension 1). (The latter function is still too uncomplicated for its graph to have the required fractality property!)

Fortunately for our construction of the graphs of the functions $g_n(t)$ in the preceding section, while new steps add "more and more points of oscillation", it turns out that every next step "thickens" the pool in smaller and smaller increments. So, as far as visualization is concerned, this leads to a very quickly converging process.

Remark 28: Due to the nature of toy transform, the constructed functions $g_n(t)$ vanish at their non-smooth points. Hence the non-smooth points on the graphs above are where the graphs meets

---

108 This is yet more pronounced when the "toy transformation" $tg(-1/t)$ is replaced by the "actual transformation" $g(-1/t)/t$.

109 The closure of this hyper-family is a Cantor set (a closed totally disconnected subset of $\mathbb{R}$ of full cardinality; it is homeomorphic to $\{0,1\}^{\mathbb{N}}$).

For those who know continued fractions, this set is quite similar to the set of numbers such that the coefficients a_n of their continuous fractions are all larger than c. Here c depends on how much we shrink the transformed graph of $G(t)$ to match the graph of $G(t)$ (and we allow negative numbers as coefficients).

In examples related to the Langlands Program, c depends on the conductor.

the t-axis. For our plots of $F^{(-1)}(t)$, which are symmetric w.r.t. the "honest" fractal transform, the "extra term" (see p. 42) can move these points off the t-axis.

Remark 29: Since the non-smooth points of the graph form a fractal, "for most[110] of the points t" the function $G(t)$ is smooth and non-0. In particular, $1/G(t)$ makes sense "for most of the points t".

Moreover, since $G(t)$ satisfies the rule above with the "toy transformation", $H(t) := 1/G(t)$ satisfies the similar rule with the "actual transformation" $H(1/t)/t$ instead. This gives an example of a function satisfying the "actual fractal transformation" law for one point P: the origin.

We do not plot the graph of $H(t)$: if we want its interesting parts to fit the page, most of them are going to be too small. However, it is not hard to imagine how this graph looks like.[111]

Remark 30: One can see that near any point from the Cantor hyper-family the graph above looks like a toy transform of itself. And indeed, this is what necessarily happens. (In other words: chaining any number of operations of the toy transform and shifts of the arguments would not give any new transformation comparing to just "shift argument, then toy-transform, then shift argument again". We discuss more of this on p. 78.)

Summarizing: if we know that a graph of a periodic function allows a fractality law which works at $t = 0$ (in other words, the function is not changed by a "toy transformation" *at one point 0*), then there is a huge collection of *other* points t for which the fractality law holds. These points are horizon-self-similar (see p. 41 before Remark 27).

These points (together with their accumulation points) break the real line into intervals; in every one of these intervals the mentioned above fractality laws do not restrict the behaviour of $\operatorname{Re} g$ whatsoever. (Recall that above we, essentially, defined the function $\operatorname{Re} G$ in such an interval almost *arbitrarily*.) Two plots above show an example when the function changes smoothly on such an interval (with a few corner points).

Moreover, our fractal transforms *interchange* these intervals; combining these transforms, one can send any such interval to any other. Additionally, there is a fractal transform which "inverts" a given interval (and multiplies the function by i). In particular, if we know the graph of $\operatorname{Re} g$ in one of the intervals, it determines g on the whole real line.

For more details, see Remark 51 on p. 78.

Remark 31: The fractality laws of the preceding remark work at particular points t (the horizon-self-similar points), and these points avoid certain intervals. This allows us to define the function $\operatorname{Re} g$ arbitrarily on one of these intervals. This means that these fractality laws still leave infinitely many degrees of freedom for the choice of function g.

Compare this with the promised fractality laws for the function F: the horizon-self-similar points appear in every interval.[112] Moreover, the fractality laws determine F up to a finite number of degrees of freedom (compare with the discussion near Footnotes 83 and 84 on p. 35).

In fact, the contrast between these situations reflects what was happening in number theory for half a century before Langlands. In 1918 Erich Hecke has shown that our function $F(t)$ is horizon-self-similar at 0 (hence in all points from the "Cantor hyper-family" on the graphs above).[113] Until Langlands, mathematicians wouldn't suspect that there must be many more points of self-similarity,

---

[110] . . . meaning: outside of a "meagre set of measure 0".

[111] In fact, this trick with replacing $F(t)$ by $1/F(t)$ may be a complete red herring. Here, we could use it only because $G(t)$ was behaving nice at a lot of points — and this won't happen for functions satisfying the fractality law at every point $2\pi^R/s$.

The functions $F(t)$ considered below are "too singular" — I do not know any mathematical approach which would make sense of the expression $1/F(t)$. One is forced to proceed as in Footnotes 100, 102 on p. 41.

[112] In a certain very precise sense a positive fraction of the set of numbers $2\pi^R/s$ are horizon-self-similar. Compare with Footnote 205 on p. 79.

[113] In fact, he found another — equivalent — formulation. (In Footnote 128 on p. 54 we have a few more details.)

and that these laws would severely restrict how our red/green coloring (or numbers N_k; see p. 33) may look like.[114]

All the fractal transformations together: infinities and regularizations

Return back to the situation when "horizon-self-similar points"[115] appear everywhere. Now *every* small piece of the graph contains a smaller piece which "looks the same" as "what happens with the graph near horizon". Comparing with two graphs above, the function should be at least as pathological as that — but the behavior of the graph above near the origin should now happen near every point of the graph. With "actual" fractality law we get a pole instead of each zero on the graph — and this means that such functions are not *possible to graph* at all!

How can it happen that a function is impossible to graph? Above, we described $F(t)$ as the Fourier transform of the sequence N_n. On the other hand, numbers N_n are whole numbers; one can immediately see that at any real point t, the series $\sum_n N_n \mathrm{e}^{\mathrm{i}nt}$ diverges! In other words: we *defined* the function $F(t)$ using a summation which does not makes sense anywhere!

Did we cheat? In fact, no! Mathematicians established a solid foundation for working with similarly divergent series (in a certain sense, "to work with infinities") already in mid-20th century.

For example, one can write $F(t) = -H''(t)$, with $H(t)$ being the Fourier transform of the sequence N_n/n^2. *This* sequence decreases quickly enough for its Fourier transform to make perfect sense; so $H(t)$ is a well-defined continuous function. While not every continuous function has a derivative which makes sense as a "usual function", every continuous function may be thought of as "a generalized function"[116], and any generalized function has a derivative which is also a generalized function. **Conclusion:** $F(t)$ makes perfect sense as a generalized function.

We can describe this generalized function as a second derivative of a continuous function. In other words, the second antiderivative of $F(t)$ is continuous. This gives us a way to work with $F(t)$ via "its *regularization*" $H(t)$ (since it carries all the info about $F(t)$!); this is what we meant in Remark 20 on p. 37.

In fact, already the first antiderivative of $F(t)$ is plottable. In what follows we work with this antiderivative $F^{(-1)}(t)$ as a "regularization" of $F(t)$.

Remark 32: The reason why this generalized function "is impossible to plot" is that it has "too much energy" in high-frequency harmonics; the situation is quite similar to the theory of "white noise".[117] When we filter out high frequencies from white noise (low-pass filtering), we get "a usual function" with well-behaving graph. However, adding higher and higher frequencies (i.e., raising the cut-off frequency) adds more and more "bumps" on this graph, and the amplitude of these bumps grows larger and larger. When we draw the graphs of results of low-pass filtering with growing cut-off frequencies together, the lengths of these graphs increase, so every next graph "requires much more ink than the previous graph". The "un-inked white space" left on these graphs "shrinks" when we raise the cut-off frequency.[118]

Conclusion: the graph of unfiltered white noise "would fill the whole plane". The same would happen with the graph of $F(t)$.

---

[114] Indeed, since there is just a finite number of degrees of freedom, knowing a color of a few prime numbers plus the fractality laws should determine the colors of the rest of prime numbers.

[115] These are defined on p. 41 before Remark 27.

[116] In other words, "a function which may have no value at any particular point, but 'weighted averages' of these values still make perfect sense".

[117] Any particular white noise function is also "only a generalized function". It is a derivative of the corresponding Brownian motion — which is a continuous function with "no derivative in the 'usual' sense".

[118] Compare this with the plots in Remark 21 on p. 37. While the filters there are not the "usual" low-pass filters (they are much stronger on high frequencies), these filters also have a characteristic frequency which goes down as s grows.

Remark 33: If the graph of $F(t)$ does not make sense, what is the description that "it is an exact fractal" good for? Indeed, this should be understood "as a metaphor only".

On the other hand, the property like "$F(t)$ is the same as $F(-1/t)/t$ up to rescaling" makes perfect sense for generalized functions as well.[119] So our description of the fractal behavior of the graph is a metaphor for the "transformation properties" of the function $F(t)$.

Fractality law for antiderivative

In the previous section, we established that

- The function $F(t)$ satisfies the "actual" fractality law — but we cannot plot $F(t)$.
- The antiderivative $F^{(-1)}(t)$ may be plotted.

Fortunately, the antiderivative $F^{(-1)}(t)$ also satisfies a certain "fractality law".

However,

- When written down as a formula, this law looks way more complicated than the "toy" and "actual" fractality laws considered above. For example, it includes integration.
- On the other hand, in these notes we *use* fractality laws only "visually": essentially, we observe graphs, and recognize "features" related to a fractality law.

It turns out that *for the purpose of visual comparison*,

the fractality law for the graph $F^{(-1)}(t)$ is indistinguishable from the toy fractality law.

In fact, this claim has one exception. Essentially, there is "an extra term" in the fractality law, and this term "moves the features of the graph up and down a bit" — comparing to the toy law.[120]

For example, compare the graph on p. 45 with the graph on p. 34. With purely-toy fractality law, all the non-smooth points are on the t-axis — while in the "Maass" plot the similar features appear at different heights.

From this moment on, all our plots are graphs of antiderivatives of functions satisfying the "actual" fractality laws.

Such graphs closely resemble a graph of a function satisfying "the toy law", except for vertical shifts.

So to recognize the type of fractality dictated by the Langlands program, we inspect the graph of $F^{(-1)}(t)$ looking for features related to *the toy* fractality law — but we allow these features to appear at different heights. (We return to this theme and show some plots in the section on p. 81.)

Hidden symmetries in degree 3: the first "real life" case

Return back to the function $F(t)$ which was constructed based on our sequence of red/green colors related to "tetrahedral numbers + 2" (on p. 19). Recall that (see p. 33) we "transliterate" a sequence of colors to a sequence of numbers N_n, and the function $F(t)$ is the Fourier transform of this sequence. We claimed that the graph of this function follows the fractality laws described in the last three sections (at least in a "metaphoric sense").

Hopefully, the preceding section gives an idea which kinds of nastiness one may expect from this graph:

- This function is impossible to plot directly.
- Its antiderivative is plottable.
- This plot satisfies a fractality law very similar to the "toy fractality law".
- However, in contrast to the "toy fractality law", the "matching pieces" may be at different heights.

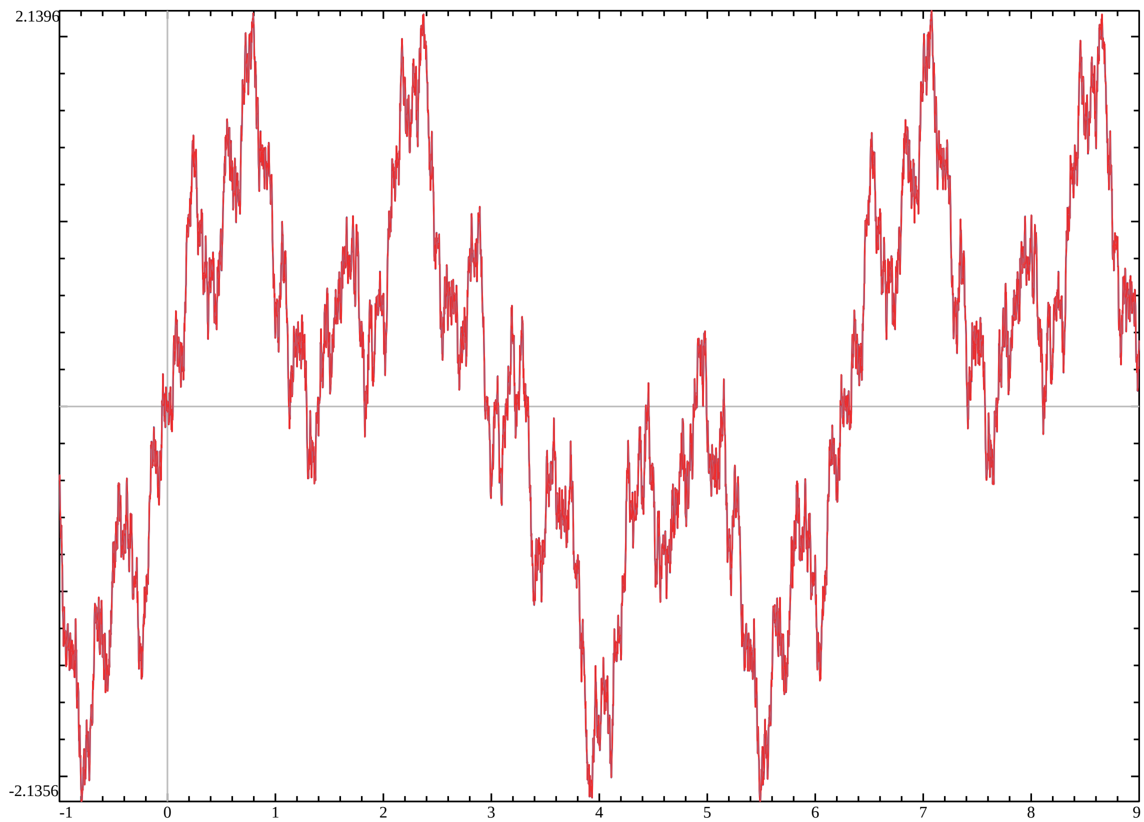
---

[119] At least if one understands it as $tF(\mathrm{const}\cdot t) = \mathrm{const}\cdot F(-1/t)$, as in Footnote 102 on p. 42.

[120] We discuss this extra term below, on p. 81.

Essentially, these expectations are fully satisfied by the graph on p. 34. For example, near the origin this graph looks very similar to a "toy transform" of itself".

However, the actual graph[121] of the antiderivative $F^{(-1)}(t)$ (about $1\frac{1}{2}$ periods)

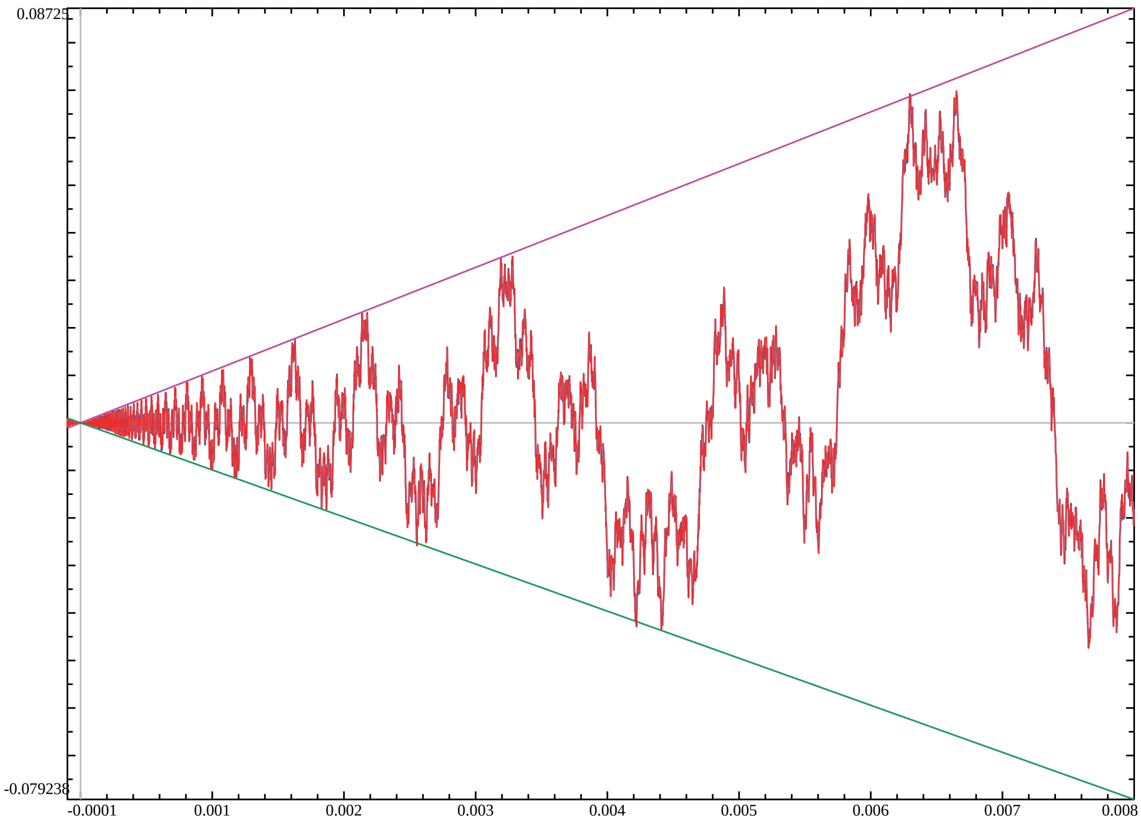


does not look this way — there is no piece similar to the toy transform of this periodic graph! What is the reason for this?

Answer: what is spoiling the fun in the graph above is the conductor! For the graph on p. 34, the conductor was 1. For the graph above, the conductor is 971 — and the larger is the conductor, the smaller are the parts where "the patterns of toy transformation" are clearly visible.

[121] The specs of blue (hardly) visible on this graph are due to this being two graphs of top of each other: blue for 500,000 terms of Fourier series, red for 1,000,000 terms. So where blue is visible, this means that 500,000 terms were not enough to get the required precision of calculation.

So, to see this pattern near the origin (to observe the “hidden symmetries”), we need to zoom into the graph with a very strong magnification (about 971 times):



Now the pattern is clearly visible.[122] Moreover, it can be seen that while this looks like a toy transform of a periodic function, this is a toy transform of a function *different* from $F^{(-1)}(t)$ (for example, the parts below the t-axis look very different from the parts above).[123]

To see the part of the graph which is recognizable as the toy transform of $F^{(-1)}(t)$ itself (we called such points “horizon-self-similar”, see p. 41), we need to zoom *again* scaling 971 times near, for example, $2\pi/971$. Unfortunately, the computational facilities accessible to me right now are not enough for doing this plot: without further speedups, it would take several weeks to plot this! (We revisit graphs of this function in the section on p. 79. For a heuristic estimate of zoom factors needed to expose the extent of fractality see Remark 62 on p. 100.)

[122] Note how the graph gets separated from the (violet and purple) straight lines when we get closer to the origin. This is due to numerical errors. There are two contributions: first, the finite number of terms of Fourier series we take (16,000,000 for the red graph). Second, as we get closer to the origin, the plot gets fewer and fewer samples on one “period” of oscillation, missing the maximal/minimal values more and more (in the graph, we used 7,500 samples — with density increasing near 0).

Zooming in, one can see a completely flattened region near 0. The experiments show that it is a result of the first contribution (as above) — but I cannot invent any simple argument explaining this! (Saddle-points calculation clarifies that this has the same nature as the Riemann–Siegel summation formula for ζ-function.)

[123] Additionally, recall that $F(t)$ is even (by definition), hence $F^{(-1)}(t)$ is odd. If we can write $F^{(-1)}(t) = t\Phi(-1/t)$, then Φ must be even — so it cannot be $F^{(-1)}(t)$ rescaled! (Indeed, looking at the graph of $F^{(-1)}(t)$, no shift would make this function even.)

However, it turns out that $\Phi(t)$ *is* $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ rescaled. We return to this theme in Footnote 128 on p. 54 and in the section on p. 81.

A simpler-to-plot example: $M = 6$

As the preceding section shows, the plots related to the polynomial "tetrahedral numbers + 2" turn out to be very hard to draw. However, eventually, to get closer to the situation which could not be dealt with without Landlands program,[124] we would need to consider different sequences anyway. For example, for a fixed number M, one can consider the sequence[125] "$M \times$ tetrahedral numbers + 1".

The difficulty encountered in the previous section is related to the fact that discriminants[126] of polynomials of degree 3 tend to be quite large in magnitude (hence the conductors are also expected to be large). For the example of the preceding section, the discriminant is -4×971. In fact, the smallest value for the magnitude of discriminant is 23 (for discriminant -23).

Fortunately for us, this smallest value *is reached* on one of the example sequences we just defined, for $M = 6$. Moreover, zooming twice into the graph, each time scaling 23 times is quite within the grasp of the software I have. Finally, this discriminant is negative, so one does not *need* the Langlands program to see that "the toy transformation" is going to be applicable to the graph.[127]

So let's redo what we did above, starting with the polynomial "$6 \times$ tetrahedral numbers + 1".

- Assign colors to numbers according to whether they can be divisors of "$6\times$tetrahedral numbers$+$ 1".
- Transliterate colors to numbers N_n (for details, see the section on p. 59).
- Take the Fourier transform $F(t)$ of the sequence N_n.
- Plot the antiderivative $F^{(-1)}(t)$

[124] See the section on p. 82.

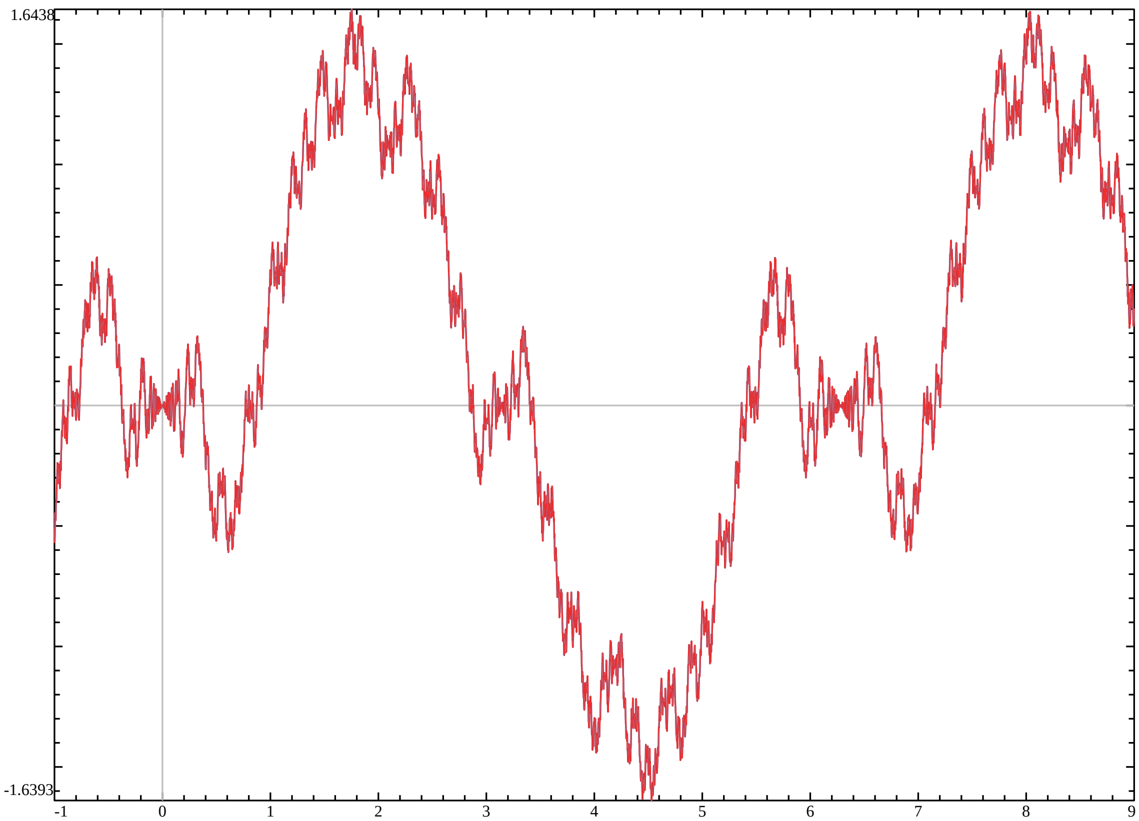
[125] Note that doubling this sequence to become "$2M \times$ tetrahedral numbers + 2" leads to the same prime divisors, with a possible exception of 2. However, such an exclusion is "negligible", since when matching the patterns of colors, we allow a few primes to be exceptional anyway (compare with 2 above, on p. 13).

This shows that the sequence "tetrahedral numbers + 2" is, for all practical purposes, also covered by this scheme, since it is "$2M \times$ tetrahedral numbers + 2" with $M = \frac{1}{2}$.

[126] Discriminant of a polynomial is a very fundamental "invariant" of the polynomial. It correlates with the minimal distance between roots of the polynomial. It also governs many features of the modular reductions of the polynomial. (This number is a polynomial expression of the coefficients of coefficients of the polynomial.)

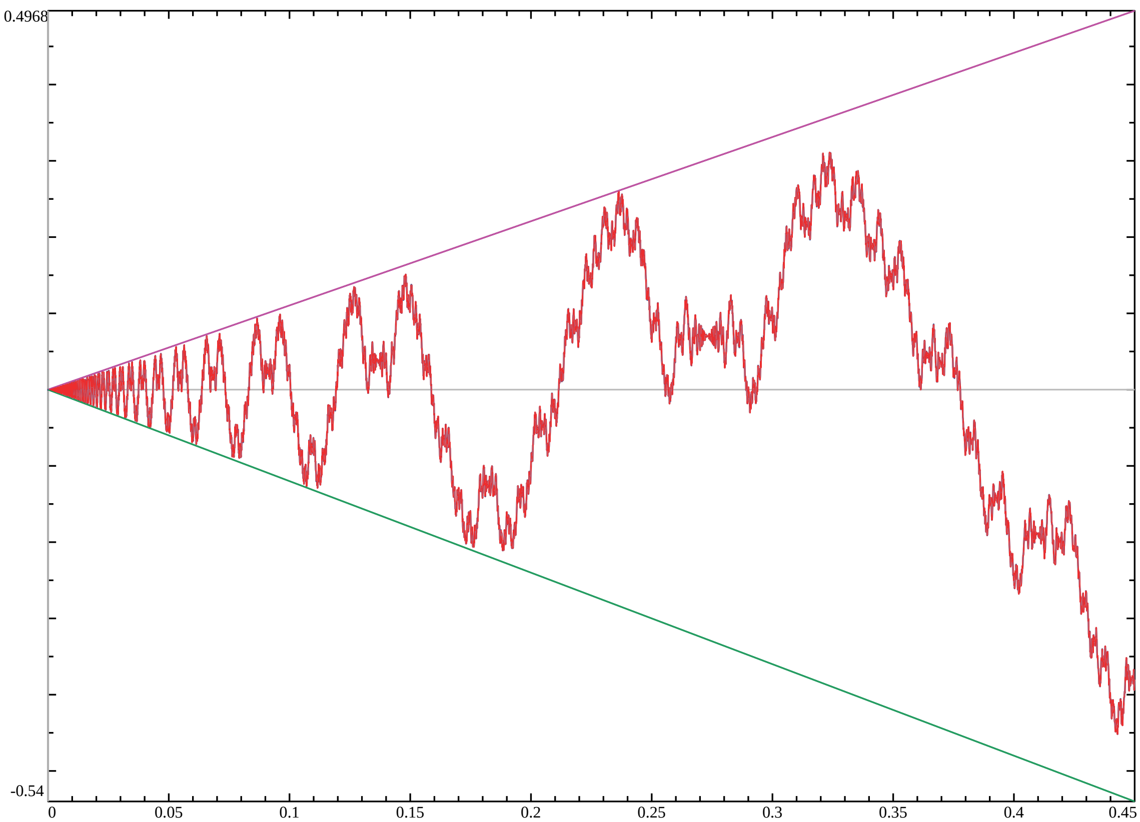
[127] Again, compare with the section on p. 82.

Here is the result (about 1½ periods)



This time, one can *guess* that the region near 0 may resemble the toy transform of a periodic function. Still, with this graph, it takes a leap of faith to trust that it actually happens.

Now zoom in (about 23 times) near the origin:



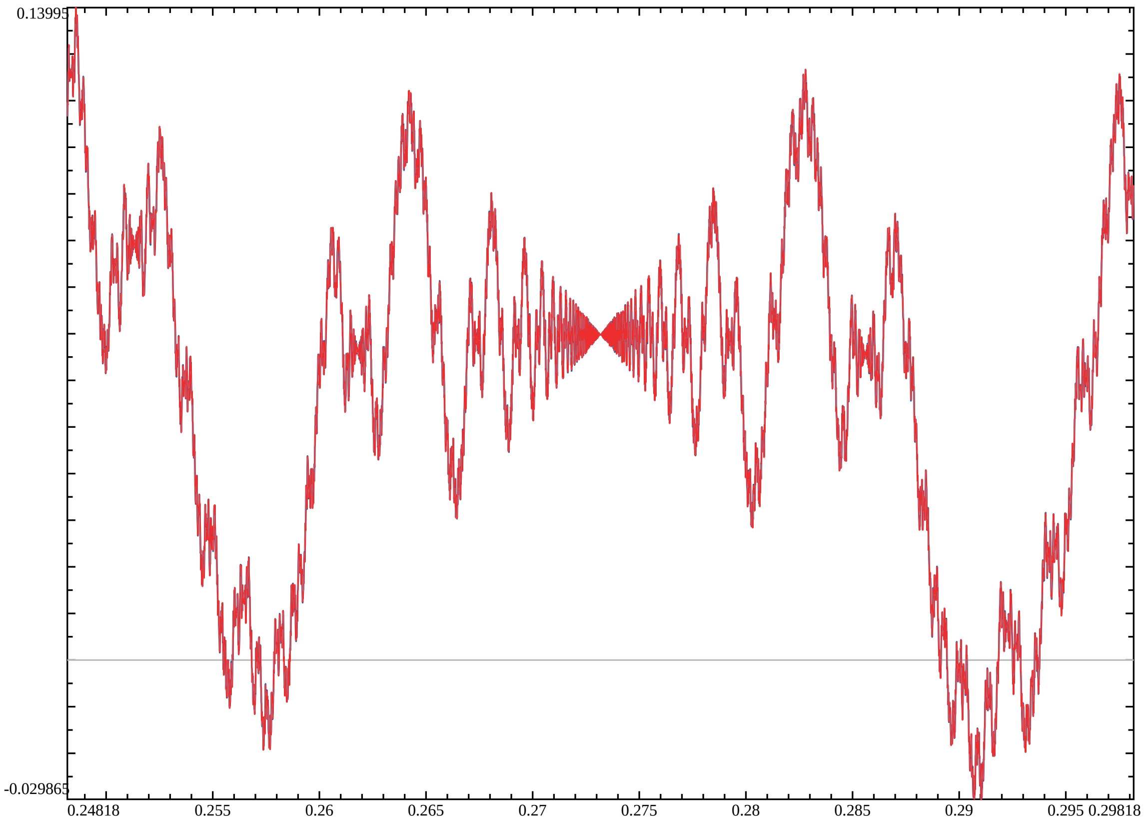
The pattern "a toy transform of a periodic function" is again clearly visible. In notations introduced on p. 41 before Remark 27, this periodic function is $(F^{(-1)})_0(t)$. Moreover, the same as in the previous section, comparison of two preceding graphs shows $(F^{(-1)})_0(t)$ *is different* from $F^{(-1)}(t)$. Again, the parts below the t-axis look very different from the parts above.[128]

Next, zoom *again* with scale 23 times near, for example, the point with $t = 2\pi/23 \approx 0.27318$. This point is clearly visible on the graph above; around it is the largest region away from 0 which

[128] In fact, this is one of the situations where $F_{\mathbb{C}}(t)$ is easier to deal with than $F(t)$. One indication of this is that $F_0(t) = \operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$. (Compare with the violet graph of $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ in the top-left plot of Remark 18 on p. 36.)

It turns out that the point $t = 0$ is very special from historical point of view. Its horizon-similarity can be explained by the functional equation for the Dedekind ζ-function which was discovered more than 100 years ago — half a century before the Langlands program. See the section on p. 82 for details.

resembles "a toy transform of $F^{(-1)}(t)$ *itself*":



Finally, this part of the graph indeed looks very similar to the toy transform of the whole graph — as expected! Indeed, every "oscillation" of the graph is similar in shape to the period of the whole graph. (So here we encounter the first "real" example of "horizon-self-similar"[129] point — defined on p. 41, and the first explicit "hidden symmetry".)

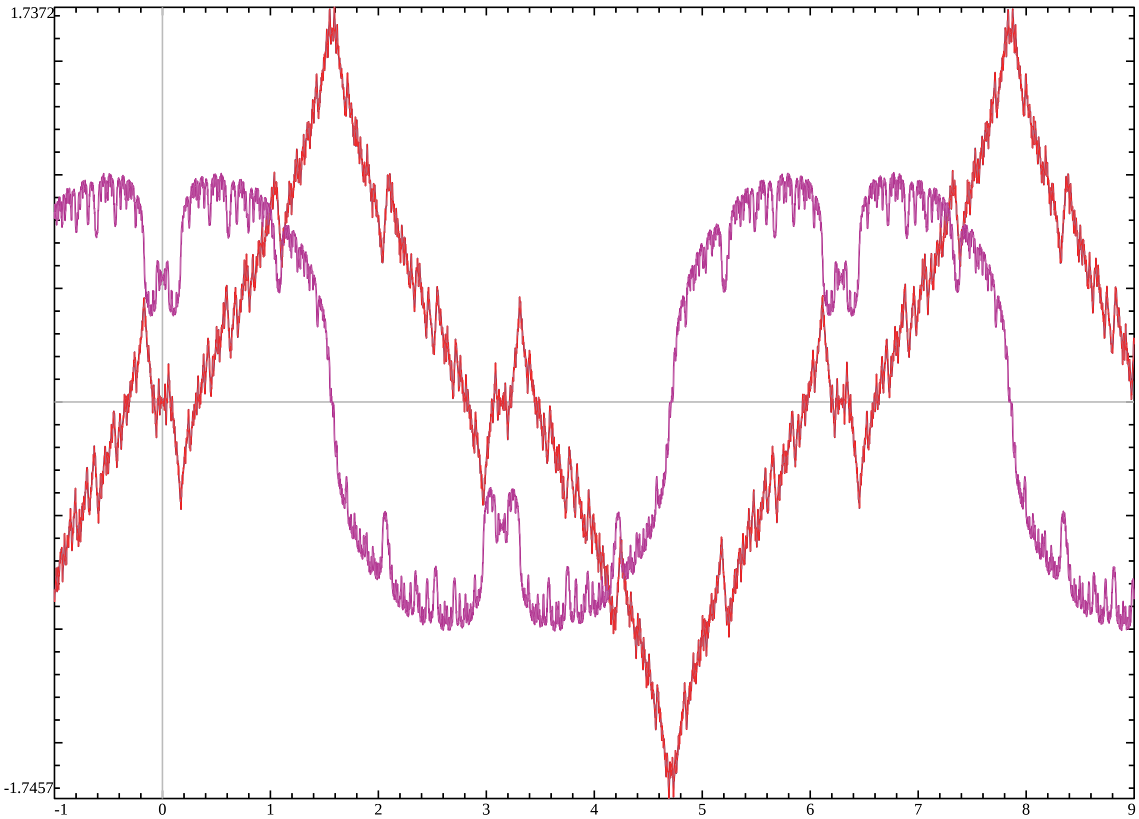
Maass fractality laws

Above, all our graphs were for the "odd case" (or "modular forms"), when the fractality laws for the function $F(t)$ could have been described by the Class Field Theory (see p. 82). This happens for polynomials "$M \times$ tetrahedral numbers $+ 1$" with a whole number $M \leq 15$. At last, here we consider what happens in "the other" case.

Unfortunately, the smallest conductor in "the other" case is $c = 2^2 \times 37 = 148$ (for $M = 24$, when the discriminant is $2^4 \times 37$). In general, this would require zooming in 148^2 times for our method of plotting. This may be too large for the software we use (would take days to calculate).

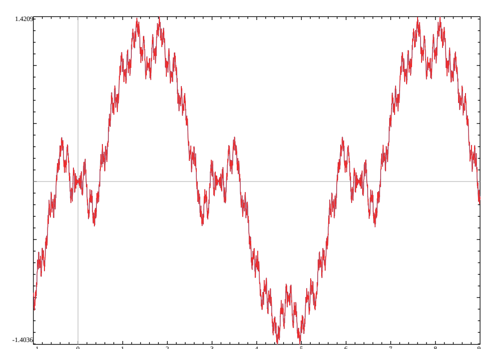
[129] With a correction that here we get not a "toy" transform, but the "honest" fractality law (see p. 81).

Fortunately, "an extra coincidence" happens, leading to extra symmetries in this case, which make zooming feasible:



Above, $M = 24$; the plot of the corresponding function $F^{(-1)}(t)$ is in red, and the corresponding imaginary part $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ is in violet. Note the mirror symmetry of the red graph w.r.t. the line $t = \pi/2$.

This extra symmetry (which may be suspected from the factor 2^2 in the conductor $c = 2^2 \times 37 = 148$)[130] makes our zooming factors behave as if this number was about 4 times smaller. This makes plotting feasible.[131]



Aside: It is interesting to note that the zooming factors needed for the simplest polynomials in the cases of positive and negative discriminant (37 and 23, with $M = 24$ and $M = 6$ correspondingly) are of the same order of magnitude — although the smallest conductors in these cases (148 and -23) are very different in magnitude.[132] (On the other hand, this coincidence may be a red herring: a similar symmetry may decrease the needed zoom factor in the "odd" case as well. On the right, we show what happens when $M = 12$: the conductor is $-2^2 \cdot 11$; it is larger than 23 in magnitude, but the zooming factor of 11 is enough. So it is not 37 vs. 23, but 37 vs. 11.[133])

[130] This mirror symmetry is due to the transliteration rules for the prime 2 following the last case of Step (d) on p. 60. Because of this, $N_{2^k} = 0$ for any k, which implies $N_{2k} = 0$ for any k by the rule of Step (e) on p. 60.

[131] For the graphs below, we used N_n for n up to 1, or 4, or 16 millions.

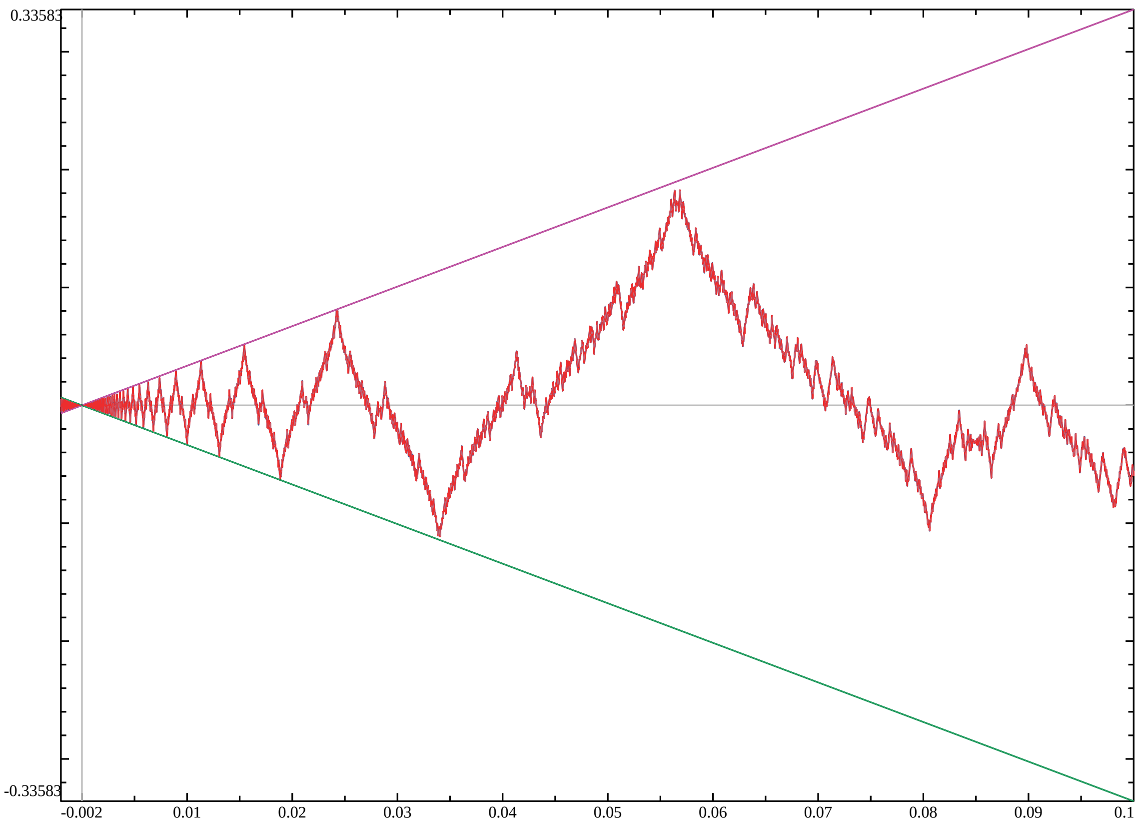
[132] Recall that the case $M = 6$ is the cubic polynomials with negative discriminant, which are covered by the Class Field Theory (see p. 82) — so do not *need* the Langlands program.

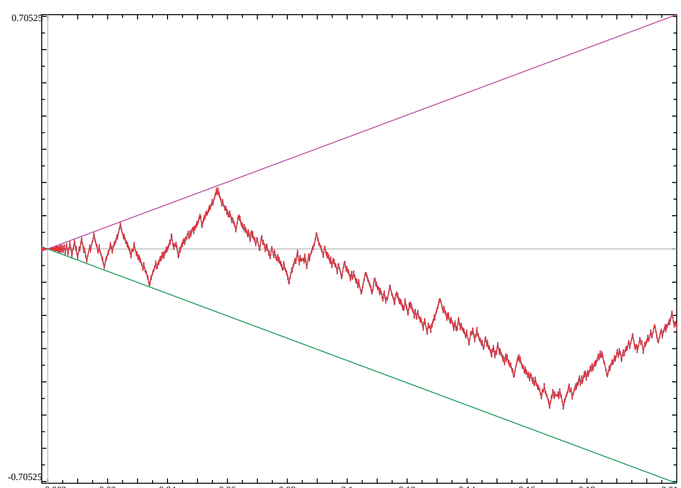
[133] On the other hand, Arnold's *Principle of Fragility of Good Things*[134] focuses on behaviour of roots of $x^3 + px + q = 0$ for small p, q; just "a minority" of these have 3 real roots. This scarcity may explain why conductors which are "large enough" to "work" for general polynomials may be "not large enough" for polynomials with 3 real roots.

[134] ... referenced in WikiPedia article on "Anna Karenina principle".

In this section, we focus on the case $M = 24$.

Recall what we saw in the left column of Remark 18 on p. 36 (discussed in the section on the case $M = 6$ on p. 51): in the "odd case" what happens near 0 on the red graph is visually indistinguishable from the toy transform of the violet graph. Now the situation is, in a certain sense, much easier (this is the right column of Remark 18 on p. 36): near $t = 0$ the graph is visually indistinguishable from the toy transform of *itself*:



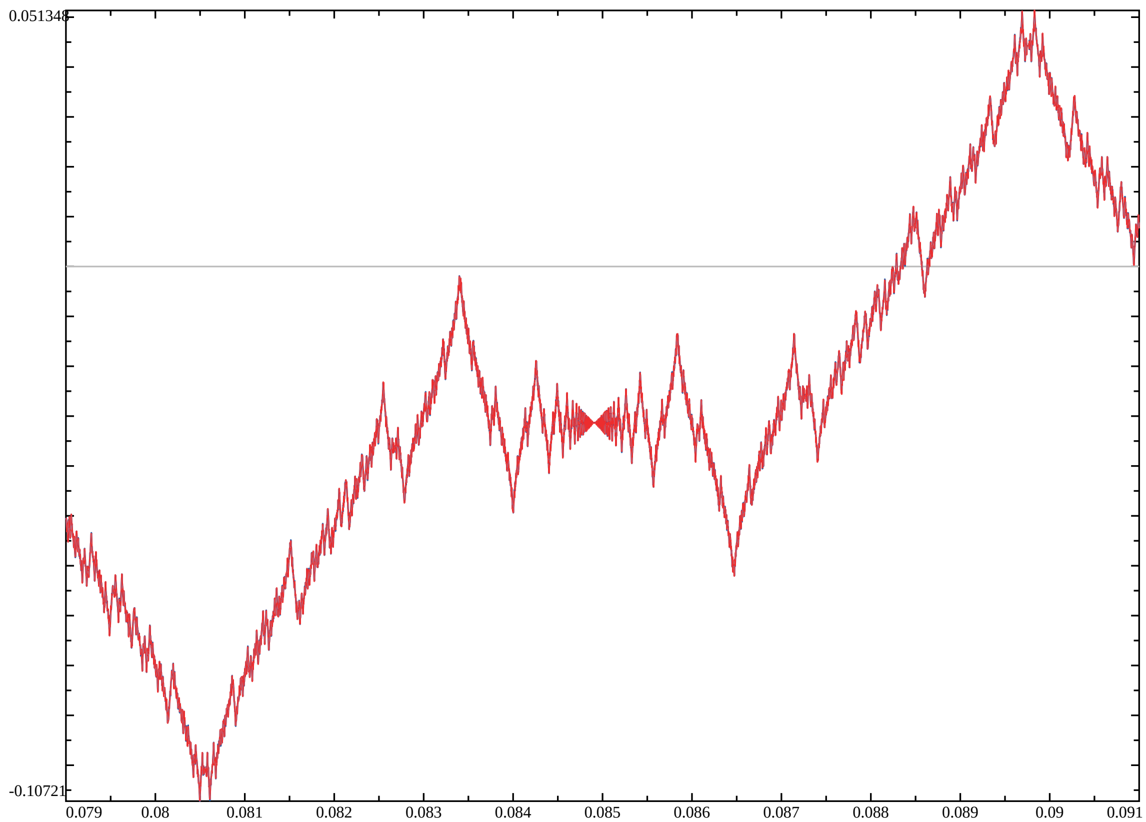


Remark 34: Note that the rightmost maximum, near the violet asymptote, is the transform of the maximum of the first red graph near $-\frac{3\pi}{2}$. In particular, we could have used 3 times smaller magnification so that the extended graph would also include the transform of the minimum at $-\frac{\pi}{2}$ (on the right). Unfortunately, the difference between the honest fractal transform and the toy transform becomes very large in such an extended domain[135]. In particular, this minimum is far away from the green asymptote of our graph — and this makes the extended graph too confusing (compare with Remark 53 on p. 82).

Near the right edge of the large graph above (close to $t = 0.085$ and near the t-axis) one can see what looks like a tiny "copy" of this whole picture. This much more magnified view of what happens

[135] Compare with the section on p. 49, where we called this difference "the extra term".

near the point $t = \frac{\pi}{37} \approx 0.0849079$ confirms this: the graph behaves similarly to $t \approx 0$:



Conclusion: in this case considering the complex-valued function $F_{\mathbb{C}}(t)$ gives no benefits — the whole theory becomes completely real! The shapes of oscillations in all these red graphs match each other.

Remark 35: Recall what we did in Remark 21 on p. 37: in the "odd" ("modular forms") case, we would extend the function $F(t)$ to a function $f(z)$ on the upper-half plane (with $z = t + \mathrm{i}s$); this changes $F_{\mathbb{C}}(t) := \sum_n N_n \mathrm{e}^{\mathrm{i}nt}$ to $f(t,s) := \sum_n N_n \mathrm{e}^{\mathrm{i}nt - ns}$. In other words, we replaced N_n by $N_n R(ns)$ with $R(s) = \mathrm{e}^{-s}$ being the "regularizing factor". (See also the section on p. 84.)

Such a replacement turned out to be *compatible* with fractal transforms (in the sense of Remark 23 on p. 38). In the "even" case we use $|t|$ instead of t as a factor in our fractal transform — so for compatibility, we need *a different* regularization. The Langlands theory predicts which regularization is needed, leading to the case of "algebraic Maass forms".

The answer: instead of $R(s) := \mathrm{e}^{-s}$ used for the "modular forms" regularization, one should write $R(s) := \sqrt{|s|} K_0(|s|)$ with K_0 the Bessel function.[136] In particular, instead of looking at $f(t + \mathrm{i}s) := \sum_n N_n \mathrm{e}^{\mathrm{i}nt} \mathrm{e}^{-ns}$, one should write $f(t + \mathrm{i}s) := \sum_n N_n \sqrt{|s|} \mathrm{e}^{\mathrm{i}nt} K_0(|ns|)$. Additionally, this summation involves negative indices n as well; in particular, one needs a way to extend N_n to negative values of n (for the plots above, we use $N_{-n} := N_n$).

This time, $f(z)$ is not complex-analytic (but it is still real-analytic). The formula above ensures that $F(t)$ is "the trace" of $f(z)$ on the absolute:[137] the main term in the asymptotic of $f(z)$ when

---

[136] The precise form of K_0 is not important for our purposes.

[137] (Ignore this footnote unless you are used to different notations!) *This* is the reason for us taking a shortcut in the formula above: if one follows our recipes *literally*, then one would need to write $f(t + \mathrm{i}s) := \sum_n N_n^{\mathrm{lit}} \sqrt{|ns|} \mathrm{e}^{\mathrm{i}nt} K_0(|ns|)$. So to get the "intertwining" property with $F(t)$, an extra substitution would be needed: $N_n^{\mathrm{lit}} = N_n / \sqrt{|n|}$.

Keeping track of both N_n^{lit} and N_n turns out to be a quite messy approach; compare with Remark 12 of Kevin Buzzard's *Explicit Maass forms*.

$s = \operatorname{Im} z \to 0$ is $F(t)\sqrt{s}\log s$. Moreover, taking "the trace" is compatible with Lobachevsky-moves ("intertwinging"), which means that the fractal trasforms of $F(t)$ are the traces of Lobachevsky-moves of $f(z)$.

This means that with this "'intertwingingly' compatible" regularization, $f(z)$ should behave in the same way with respect to the tessellation as in the "odd" case: knowing $f(z)$ on one piece, one can find it everywhere: a Lobachevsky-rotation or Lobachevsky-translation which sends one piece to the other preserves $f(z)$.

The properties of $K_0(S)$ show that there is another condition on $f(z)$ replacing the complex-analyticity; it is one of a curved-geometry analogues of the condition of being harmonic in flat geometry (compare with Remark 24 on p. 39). Functions satisfying this condition are called *algebraic Maass forms*.

Remark 36: Historically, the "odd case" was easier to deal with since it could be treated using the techniques of the Class Field Theory, developed about 90 years ago.[138] In fact, the first conjectures about particular examples of the "odd case" started to appear yet before Gauss; the first proofs for the cases of these examples were discovered by Gauss.[139]

I cannot find precise references for who completed the "even case" (for polynomials of degree 3) and when. I *expect* that this case should be completely understood now, judging basing on the (essentially) second-hand information about which parts of the Langlands Program are already completed.

Remark 37: The last graph is very special among the graphs of these notes. It is the only graph which *requires* the Langlands program to explain it. For details, see the section on p. 82.

The transliteration rules

Here we explain how to construct the sequence N_n from a polynomial of degree 3.

Recall our process; essentially, we do this (compare with the section on p. 33):

- Start with a particular polynomial sequence of integers (of degree 3);
- Collect all the possible divisors of the numbers in this sequence;
- Color all the whole numbers: green for possible divisors (as above), red for the rest;
- Transliterate this sequence of colors into a sequence of numbers N_n;
- Take Fourier Transform of the obtained sequence;
- Inspect the fractal properties of this function.

What is left unexplained is the transliteration process. As we said, it is quite straightforward (with the exception of how to treat prime divisors of the discriminant).

Below, we first go through the steps of transliteration, listing only the rules one should follow to perform these steps. Here we treat these steps as a "black box" recipe; in the next chapter, we try to demystify these rules — but only as far as it is possible: no matter how trivial the step may look, all of them have extremely deep connections to very profound themes of contemporary math.[140]

---

[138] Recall again: this theory was a triumph for mathematics of the first half of 20th century. However, nowadays it settled down to be a run-of-the-mill feature of mathematical landscape.

[139] In addition to "odd" and "even" cases discussed above, there is also "a special" case of the *abelian*, or *cyclic* polynomials of degree 3. In this special case (also covered by the Class Field Theory), the colors follow the *exactly* the same pattern as in the case of degree 2: prime numbers are colored according to their position on the **conductor**-wheel. (See also the section on p. 82 and Remark 78 on p. 130.)

This happens when the discriminant is a complete square. For the polynomials of the type we consider here, $M \times$ Tetrahedral Numbers $+ 1$, for integer M this happens for $M = (k + 3^5/k)/2$ with integer k, which means $M = 18, 42, 122$. (Likewise for rational M.)

[140] The manipulations below may look purely algebraic in nature. However, one of the major achievements of math of 20th century was to expose very deep connections of such steps to setups of geometry. Unfortunately, the format of these notes does not allow us to dwell on this connection.[141]

Essentially, our "chain of demystifications" goes up to the section on p. 120.

[141] **N.B. (???) Give a reference to a book?**

(a) In the sequences above (see example for degree 2 on p. 10 and one for degree 3 on p. 19), we used two colors: red and green. Recall how to color a particular number n: we use the residues mod n, and write down residues of several elements of the sequence. We know that these residues should be periodic, and know the length of the period; after we have that many residues, we can check whether 0 mod n appears in this period. If it appears, we mark the number n green, otherwise red.

For degrees higher than 2, and for several variables (as in the beginning of the section on p. 20), one should replace these colors by more detailed information: instead of marking whether the residue 0 mod n occurs in the sequence or not, mark how many times it occurs in the shortest period. While nothing special happens to red (it is transliterated to 0), green "attains several tints": it may be replaced by different numbers.

Call this count $\widetilde{N}_n^{\mathrm{res}}$ (for the residues mod n). For a polynomial of degree 3, for almost all prime numbers p the value $\widetilde{N}_p^{\mathrm{res}}$ is 0, 1 or 3.[142]

(b) We obtained a sequence $(\widetilde{N}_n^{\mathrm{res}})$ of numbers which are 0, 1, or 3 at all prime positions (with a few exceptions). To get a fractal behavior for $F(t)$, we need to "distill" this sequence a bit.

Recipe: Put $N_p := \widetilde{N}_p^{\mathrm{res}} - 1$ for a prime number p (with exceptions from Footnote 142; we cover them in Step (d)).

(c) Next, we need to define N_q for $q = p^k$ with a prime number p. **Recipe:**[143] for every prime p, choose one of the following sequences:
- $-1, 0, 1, -1, 0, 1, \ldots$ (3-periodic);
- $0, 1, 0, 1, 0, 1, \ldots$ (2-periodic);
- $2, 3, 4, 5, 6, 7, \ldots$ (a linear function),

so that its first number matches the known value for N_p. Assign these values to N_{p^k}.

(d) For an "exceptional" prime number p (of Footnote 142), one cannot find N_{p^k} given $\widetilde{N}_p^{\mathrm{res}}$ only, even for $k = 1$. The procedure is quite involved; it suffices to say that for the sequence N_{p^k} one should choose one of the sequences above, or one of:
- $1, 1, 1, 1, 1, 1, 1, \ldots$ (1-periodic);
- $0, 0, 0, 0, 0, 0, 0, \ldots$ (1-periodic).

While there is a recipe explaining which of 5 variants to choose,[144] it is easier to note that since only finitely many primes p are involved, this ambiguity leads to only finitely many choices of the sequence N_n. Exactly one of these choices would lead to the desired fractal behavior of $F(t)$.[145]

(e) For composite indices of the form $p^k q^r$ with different primes p and q, put $N_{p^k q^r} := N_{p^k} N_{q^r}$. Likewise for indices with more than 2 distinct prime factors.

Example: For "tetrahedral numbers + 2", the discriminant is -4×971, so *small* prime numbers greater than 3 are covered by the rule (c). Inspect the sequence of colors on p. 19. This shows that 11 is green, and 7 is red. So $N_7 = -1$; moreover, checking the table on p. 19 shows that 11 divides only one number of our sequence for sides 1,…,11 — which is the shortest period of our sequence mod 11. Hence $\widetilde{N}_{11}^{\mathrm{res}} = 1$, and $N_{11} = 0$. Picking up a matching sequence above, $N_{11^2} = 1$ (the second number

[142] The exceptions are the prime divisors of the discriminant, where the value may be 2 as well. Moreover, one should include the divisors of the leading coefficient (and of denominators of coefficients, if present), and $p = 2$ or $p = 3$, when the period is longer than p.

[143] See Remark 43 on p. 74 for more details.

[144] See the section on p. 115.

[145] Compare to the answer of 2010-08-14 in the discussion *Zeta Functions: Dedekind Versus Hasse-Weil* in **n-Cat Café** discussing how the errors at "exceptional" primes would break the horizon-self-similarity at $t = 0$ (which is due to Hecke's functional equation — see the section on p. 82 for details).

in the sequence 0, 1, 0, 1, 0, 1, ...), and $N_{7^4} = -1$ (the 4th number in the sequence -1, 0, 1, -1, 0, 1, ...). Finally since say, $290{,}521 = 7^4 \times 11^2$, we conclude that $N_{290{,}521} = -1$.[146]

Keep in mind that any error made during transliteration would ruin the function $F(t)$ — it won't have the desired fractal behavior. To obtain the graphs used in this report, we followed these steps precisely (treating divisors of discriminant by hand — which turned out to be very error-prone[147]).

Remark 38: As we explained, for degree 3 and residues mod a prime number, the count $\widetilde{N}_n^{\mathrm{res}}$ may be 0, 1 or 3 (with exceptions as in Footnote 142). The count 3 appears less often than the others; in the part of the colored sequence shown above (on p. 19), it appears only for prime 3. The first few other occurrences are for the primes 37, 61, 83, ...[148]

Remark 39: Replacing the sequence of colors by the sequence of counts $\widetilde{N}_n^{\mathrm{res}}$ (as in Step (a) on p. 60) was not needed for sequences of degree 2: then for the residues mod a prime number p the count is 0 or 2 (except for finitely many p␣s — and since above we allowed a few exceptions in the pattern of colors anyway, these would not matter). So two colors were enough to encode all the information in these counts for prime n (and eventually, we ignored the colors for non-prime n anyway!).

Essentially, this finishes our first goal (started on p. 33): to give the simplest possible self-contained rough outline of how to get a fractally-symmetrical function starting with a polynomial of degree 3. This example exposes both sides of the Langlands program: on the arithmetic side we have a problem about divisors of numbers in a polynomial sequence; the other side is related to fractal symmetries of $F(t)$ (or Lobachevsky-symmetries of $f(t, s)$).

In the rest of these notes, we unravel a few clarifications and finer points related to the steps of this outline.

146 This illustrates that in general, whole numbers $|N_n|$ grow very slowly.

147 Compare with Footnote 145.

In fact, a few months ago an α-release of `GP/PARI` mathematician's calculator (version `2.10.1`) changed this: it has tools to automate these tasks.

148 As we will see in Remark 48 on p. 77, close to ¼ of green primes are going to have the count 3, the rest — the count 0. However, for relatively small prime numbers, the proportion is going to be measurably smaller than ¼ (compare with the section on p. 156.

Appendix: More patterns, and additional pictorial examples of symmetries

If all you have is a hammer, everything looks like a nail.
Abraham Maslow, *The Psychology of Science*, 1966

The preceding chapter sets up the minimal possible context illustrating how (and when) the “hidden symmetries” of the Langlands program may be expressed as fractality of certain explicitly written Fourier series. Here we provide more bread crumbs to connect this setup with more customary accounts of the topics related to the Langlands program. We also expose a few beautiful effects which we kept hidden in the rough outline of the preceding chapter.

It turns out that when our hammer is the statement “‘hidden symmetries’ mean the fractality of $F(t)$”, a lot of themes related to the Langlands program happen to work very well as nails!

Plots for degree 2

A very natural question to ask is: what happens if we make a plot following the same recipe as before, but starting with a polynomial of degree 2 instead of degree 3? It turns out that while one does not need to *change* the sequences listed in Steps (c)) and (d) on p. 60, one still needs to *extend* the list of such sequences for deg = 2: for some primes one of the “old sequences” should be used, and for the remaining primes two new types of “cases” appear. While the first “typical case” recipe below was relevant for degree 3 too (as “an exceptional case”), the other two are new.

Now the modified recipes are: for every odd prime p which is not “exceptional”, choose one of the following sequences:

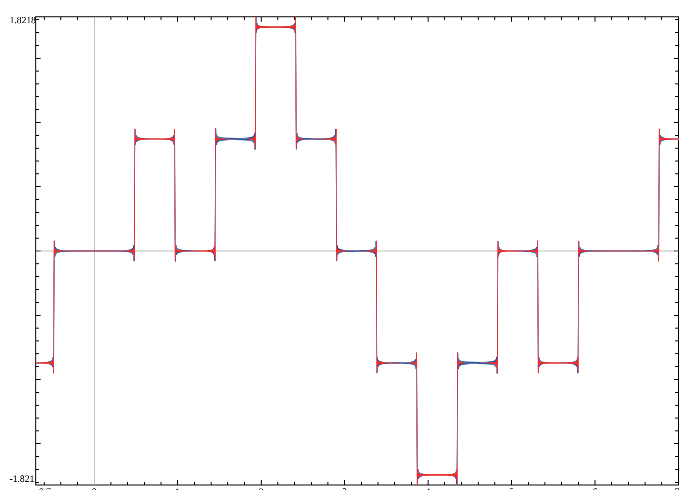
- 1, 1, 1, 1, 1, 1, . . . (1-periodic; for green primes);
- −1, 1, −1, 1, −1, . . . (2-periodic; for red primes).

so that the first number matches the value for N_p given by Step (b) on p. 60 (with a minor obvious modification of this recipe since now $\widetilde{N}_p^{\mathrm{res}}$ may be 2). Assign these values to N_{p^k}.

For “exceptional” primes (divisors of the discriminant, of the denominators of coefficients, of the leading coefficient, and possibly for $p = 2$) either one of “typical recipes” should be chosen, or the sequence below (we did not see it in degree 3):

- 0, 0, 0, 0, 0, 0, . . . (1-periodic)

(we postpone the recipe how to check which of 3 choices should be used until the section on p. 115).



Above, we wanted to write the recipe in the form similar to our recipe for degree 3 (see (c), (d) on p. 60). However, observing these 3 sequences, one can see that there is a remarkable shortcut (not possible in degree 3): $N_{p^k} = N_p^k$. In particular, the sequence N_m is “totally multiplicative”: [149] $N_{ab} = N_a N_b$.

Moreover, Quadratic Reciprocity shows that $N_p = \left(\frac{p}{D}\right)$, with D being the discriminant of the polynomial. (Here we use the Legendre symbol from p. 208.) By top-multiplicativity of Legendre symbol (see p. 208), $N_m = \left(\frac{m}{D}\right)$ *for every*

[149] Compare this with “ordinary” (non-total) *multiplicativity* which is the property of Step (e) on p. 60.
The chapter on p. 101 rewrites the conditions *added* to “ordinary” multiplicativity as $N_{p^{k+1}} = C_p N_{p^k}$ (with $C_p = N_p$) and reads it as “the sequence (N_{p^k}) satisfies a recurrence relation of length 1”.

m. **Conclusion:** the sequence N_m is D-periodic.
(Moreover, the periodic extension to $m \leq 0$ is either odd or even.)

This immediately implies that[150] $F(t)$ is a sum of δ-functions (with certain coefficients) at points proportional to $2\pi/D$. Unless D is a square, the integral over a period vanishes, and $F^{(-1)}(t)$ becomes a periodic step function. The plot above[152] is for the polynomial $4n^2 + 2n - 3$; its discriminant $4 \cdot 13$ has the square-free part $D = 13$, and since $D \equiv_4 1$, it coincides with the *fundamental discriminant* (the shortest period in the Euler' reciprocity; compare with Footnote 698 on p. 214).

Conclusion: in the case of degree 2, the "hidden symmetries" are "already exposed" in our sequence of colors (see p. 8, which are the periodic and mirror symmetric when colors are restricted to prime numbers; see the section on p. 15). Taking the Fourier transform converts this pattern not into *symmetries* of the graph (as in the case of degree 3), but into the fact that the $F(t)$ is a sum of δ-functions. Compare this with our discussion of motives on p. 75: every "flavor" of a distilled motive needs a *specific* approach to expose its pattern of (hidden) symmetries.[153] Above, we applied an approach which works with one type of motive to a motive "of wrong type" — and the result does not exhibit any symmetry.[154] (And, as we said before, applying such approaches to "a mix" of distilled motives leads to yet messier results. We consider two such examples in the section on p. 63.)

The fractality laws in a reducible case

As we discussed it on p. 55, the polynomials "$M \times$ tetrahedral numbers $+ 1$" with a whole number $M \geq 16$ have a positive discriminant, so may be used as "true" examples of the Langlands program (as opposed to the examples with negative discriminant, for which the fractal properties were already known before Langlands due to the Class Field Theory; see p. 82).

It turns out that for $M = 16$ the discriminant is $2^6 \times 13$, and experiments with the graph show that the conductor c happens to be[155] very small, 13. This is much smaller than $c = 148$ considered on p. 55. To see why we needed to deal with the harder case (one with larger conductor) observe how the graph of $F^{(-1)}$ behaves in this case; the plot of $F^{(-1)}(t)$ is in red, and the corresponding

[150] More precisely: either $F(t) \stackrel{\text{def}}{=} \operatorname{Re} F_{\mathbb{C}}(t)$ or[151] $\operatorname{Im} F_{\mathbb{C}}(t)$, depending on whether the Euler formulation of Quadratic reciprocity involves "even" or "odd" behaviour. (This happens since the Fourier series defining $F_{\mathbb{C}}(t)$ involves only N_n for $n > 0$.)

[151] The *other* graph would have a plot with log-spikes instead of jumps (we show such plots on p. 104 and p. 63). This may be explained since $F_{\mathbb{C}}(t)$ extends to $\operatorname{Im} t \geq 0$, hence Re and Im are related to each other by the *Hilbert transform*. But this operator is pseudo-differential (of order 0); hence the singularities of Re are determined by the singularities of Im (and vice versa).

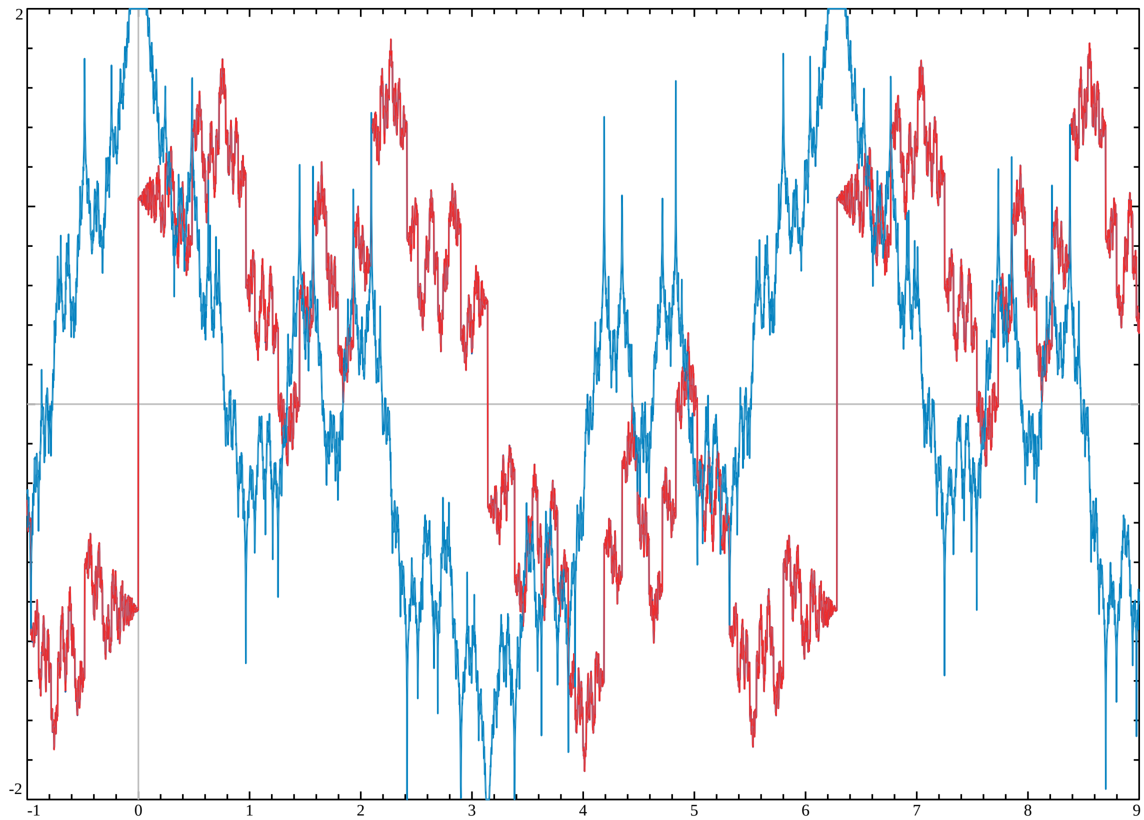
[152] Here to highlight the relevant features of the graph, we needed to use only 1,000 Fourier coefficients, instead of millions used for other graphs. However, because of this, the "Gibbs phenomenon" takes sufficiently wide zones around the jumps, and is very visible even without magnification. (Compare with Footnote 160 on p. 66.)

[153] There is a very nice (and more detailed) summary of relevant issues in the discussion *What is the Langlands Programme?* in **n-Cat Café**.

[154] Compare with the discussion at the beginning of the chapter on p. 101.

[155] For example, big upward jumps on the red graph happen at $x \approx 0.483/n$ with $n \in \mathbb{N}$. Observe that $2\pi/13 \approx 0.48332$.

imaginary part $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ is in blue:



Observe the principal properties of the blue graph which make it so different from what we saw before:

- The blue graph has a lot of "spikes"; this is due to the "logarithmic singularities"[156] at all points $2\pi{}^{R}/_{S}$.
- Only the widest "spikes" of the blue graph reach the top/bottom edges of the graph. In fact, the more narrow spikes are cut-off due to approximations in plotting. If we could increase the number of sample points for our graph by many orders of magnitude,[157] one could see that *all* these spikes actually go up or down to infinity!
- **Conclusion:** the low resolution of this plot hides another pathology: the function $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ is unbounded near these points. Since this points are dense, this means that the function is unbounded in every small interval — which means that it is impossible to plot it honestly![158]

For the red graph:

- The red graph has a jump at every point $2\pi{}^{R}/_{S}$.

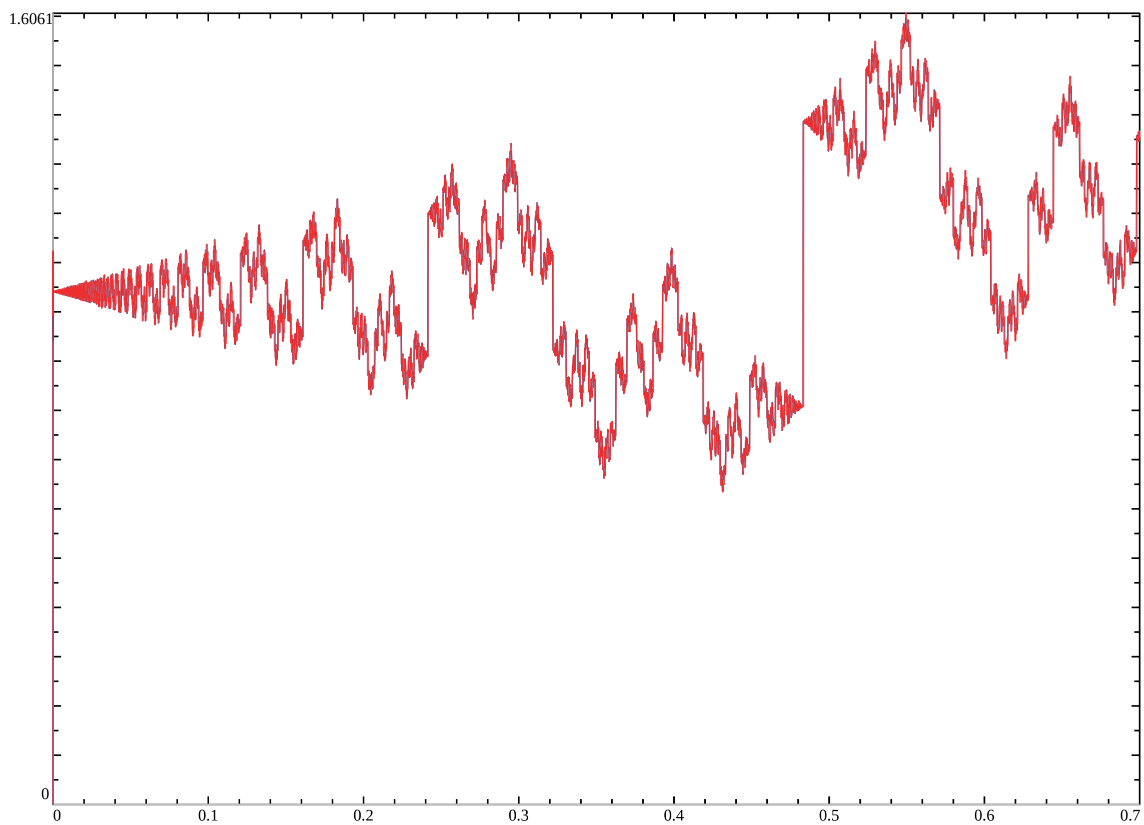
[156] These log-spikes are closely related to the jumps on the red plot. See Footnote 151 on p. 63.

[157] A singularity $y = {}^{1}/_{n} \log t$ becomes *exponentially* more narrow when $n \to \infty$. The corresponding jumps on the red graph are visible for n up to hundreds. One would need astronomical number of sample points to see similar number of spikes on the blue graph!

[158] While the "spikes" on the graph of $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ happen for t in an everywhere dense subset of $\mathbb{R}$, their projections to the t-axis happen to be a "meagre subset of measure 0", meaning that for a "random" value of t (such that t/π does not have "pathogologically good" approximations by rational numbers) $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ is close to the blue graph. (To have a plot for *every* t, one needs to take an extra antiderivative: $\operatorname{Im} F_{\mathbb{C}}^{(-2)}(t)$ has a honest plot.)

- Moreover, the graph above and its fragments shown below *suggest* that all the variation of the function $\Phi(t) := F^{(-1)}(t)$ "happens via jumps". In other words, $\Phi(t-0) - \Phi(t_0+0)$ is the sum of jumps of Φ between t and t_0 (if $t > t_0$).[159]
- On the right of every jump t_0, the red graph behaves as a Lipschitz function: $|\Phi(t) - \Phi(t_0 + 0)| \le C \cdot (t - t_0)$. (Likewise on the left.)

Recall what we saw for $M = 24$ (on p. 55): what was happening near 0 on the red graph was visually indistinguishable from the toy transform of the same graph. Now, near $t = 0$ the graph jumps from about -1.04 to about 1.04, *then* follows the "toy transform" pattern:



With a jump at $t = 0 = 2\pi{}^0\!/_1$ of magnitude $J \approx 2.08$, inspection of other jumps shows that the magnitude of the jump at $2\pi{}^R\!/_S$ is ${}^J\!/_S$; moreover, the direction of the jump depends only on $J \bmod 13$; one can recognize that the jump has the same sign as in $S^6 \equiv_{13} \pm 1$ (this is the Legendre symbol $\left(\frac{S}{13}\right)$ from p. 208). All this works for $13 \nmid S$.

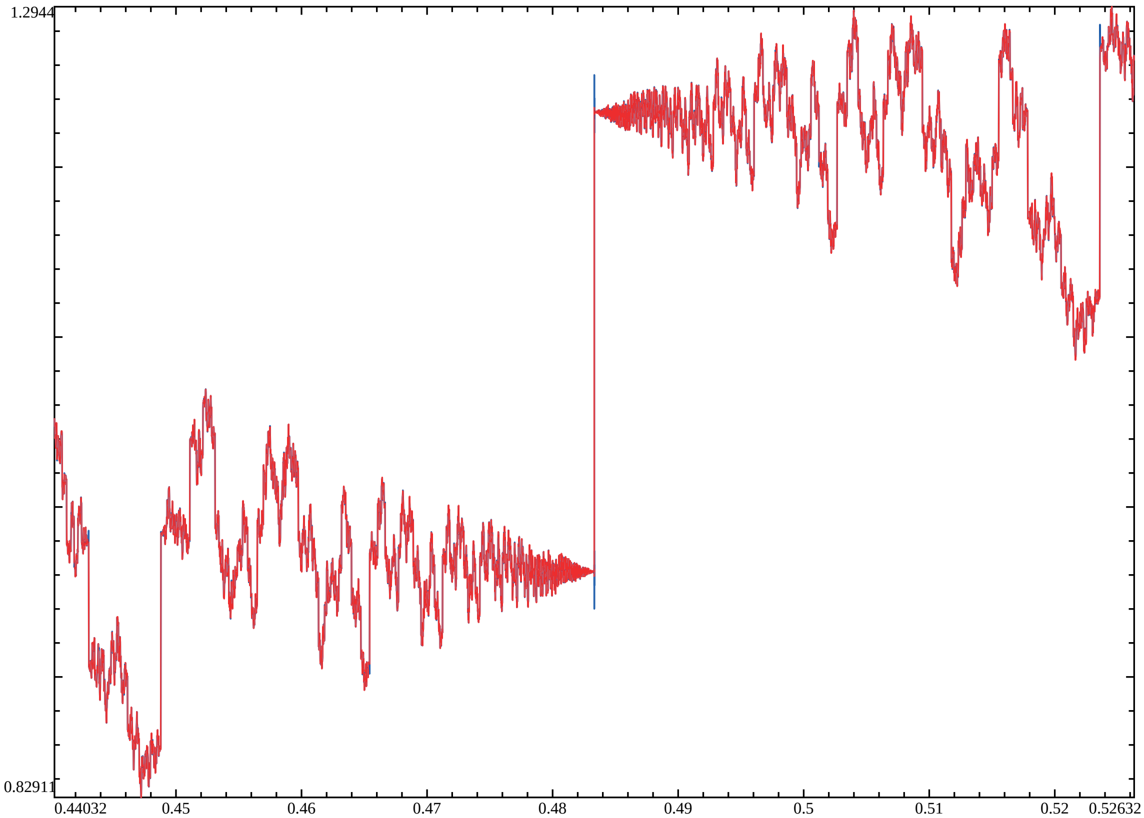
The jumps at the remaining points $2\pi{}^R\!/_S$ with $13|S$ behave differently: the magnitude of the jump is $\sqrt{13}{}^J\!/_S$; the sign of the jump coincides with $\left(\frac{R}{13}\right)$.

[159] One should be extremely careful with statements like this, since this sum is only conditionally converging. There is a way to overcome this (see Footnote 425 on p. 138). However, the result is strikingly unexpected: the sum of jumps is twice the variation $\Phi(t) - \Phi(t_0)$ of the function!

In short (we return to this in Remark 74 on p. 128): one can break jumps into two distinct types, depending on whether 13 divides Q for $2\pi{}^P\!/_Q$. It turns out that if one runs the sum above over *only one* type of jumps, this gives a correct answer! (In particular, this sum does not depend on *which* of two types we choose…) In other words: if one "forces" the correct jumps at one type of the points $2\pi{}^P\!/_Q$, the correct jumps at the other type of the points would be "spontaneously generated".

We know no heuristic which would explain this. … (This is an example of a situation when having a proof — in the section on p. 135 — does not lead to more understanding of what happens.)

For example, zoom in a lot[160] near $t = {}^{2\pi}/_{13} \approx 0.48332$:



The jump is by about $0.577 \approx 2.08/\sqrt{13}$, as predicted above.

When one focuses attention on the pattern of oscillations either on the left, or on the right of the points of jump, then (same as in the case $c = 148$) the shape of "one oscillation" matches the shape of the graph of one period of the function (compare with the first graph of this section). Under these restrictions, the shape behaves quite similarly to the shapes in horizon-self-similar points. However, if one wants to consider both left and right sides simultaneously, the jump between these sides breaks horizon-similarity completely (we discuss how to fix it in Remark 77 on p. 129).

How to explain the difference between what we see here (for $M = 16$) and what we saw for $M = 24$ (on p. 55)? The reason is very simple: the polynomial "16 × tetrahedral numbers + 1" *is reducible*: it vanishes at the point ½. In other words, $8m(m^2 - 1) + 3 = (2m - 1)(4m^2 + 2m - 3)$. So the zeros of this polynomial (including residues mod n at which the polynomial is divisible by n) break into two types: the zeros of $2m - 1$ and zeros of $4m^2 + 2m - 3$. Note that $2m - 1$ has zeros modulo any odd number n (at ${}^{(n+1)}/_2$). Therefore in the sequence of red/green colors (as those related to "tetrahedral numbers + 2", on p. 19) the color of primes $p \geq 3$ is going to be always green. Moreover, the number of solutions mod p (used by the transliteration rules on p. 59) is one more[161] than the number of solutions for $4m^2 + 2m - 3 = 0$.

[160] The "overshoots" on the jump(s) are examples of a phenomenon explained in the middle of 19th century: the "Gibbs phenomenon": they are due to sharp cut-offs in the low-pass filtering we use. Since we sum up millions of Fourier terms, these overshoots are very narrow (recall that *the height* of Gibbs' oscillations does not depend much on the number of terms, but the width does!), so the sample points for our plotting program miss the regions of these overshoots unless we use very high magnification.

[161] There is no collision between these solutions. This is due to the value of $4m^2 + 2m - 3$ at $m = {}^1/_2$ being -1 — which has no prime factors!

In the language of the section[162] on p. 75 and of Remark 65 on p. 119 "the corresponding motive is not fully distilled" — and the patterns corresponding to the factors are "overlayed on top of each other", contaminating these patterns.

In short: for a reducible polynomial the sequence of red/green colors *is a "mix"* of colors for the factors of this polynomial. Likewise for numbers N_k from the section on p. 59: they are determined by the corresponding numbers N_k^{quadr} for $4m^2 + 2m - 3 = 0$.

Recall that in the sections on p. 114 and p. 118, we introduced the reducible case $\boxed{\text{Reduc}}$ as one of the motivations of the notion of distillation. We *claimed* that distillation (or, in this particular case, factorization) simplifies the "hidden symmetries" a lot — but we did not provide the examples. Now we can give the example: factoring out $2m - 1$ changes the plots above to the plots for $4m^2 + 2m - 3$, — which we considered in the section on p. 62.

Conclusion: to see the results of factorization on the plots in the case of degrees of the factors $1 + 2$, compare the plots here to the plot in that section: fractality of Fourier transform is replaced by the periodicity of coefficients.

Decompability inverts distillation

The plots in the preceding sections show...[163]

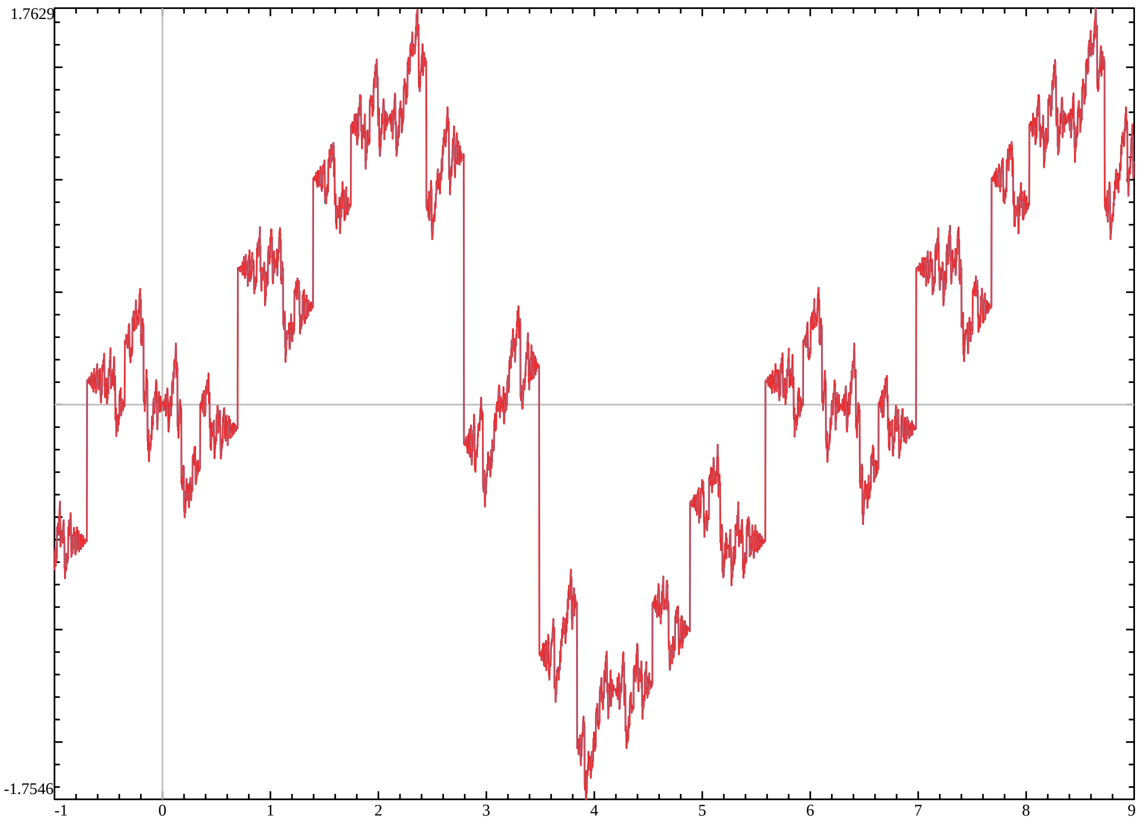
Abelian case of degree 3 and the "extra distillation"

Another case in which our naive procedure of distillation $N_p = \widetilde{N}_p^{\mathrm{res}} - 1$ does not result in a distilled motive is the case of an "abelian=cyclic" polynomial P of degree 3. While in this case the periodicity $N_p = N_p^{\mathrm{per}}$ still holds (here N_m^{per} is a certain periodic sequence), the identities $N_{p^k} = N_p^k$ and $N_{p^k}^{\mathrm{per}} = (N_p^{\mathrm{per}})^k$ and $N_{p^k}^{\mathrm{per}} = N_{p^k}^{\mathrm{per}}$ do not hold. The corresponding graph looks (again!) like an antiderivative of a sum of δ-functions (as we saw in the cases of degree 2 and D_4 without "extra"

[162] **N.B. (???) The next four paragraphs duplicate what is at the beginning of the section on p. 128. They also refer to sections in the future.**

[163] **N.B. (???) Better move the elementary parts of these discussions here.**

distillation):



(Here $M = 18$ in the the M-family of polynomials "$M \cdot$ tetrahedral numbers $+ 1$", see p. 18, and the discriminant is $D = 9^2$; compare with Footnote 139 on p. 59 and the discussion in the next section.[164])

While the jumps on this plot make it similar to the cases mentioned above, the reasons for these jumps seem to be completely different. For degree 2, the rank 1 of the sequence N_n was too small for our method of visualization (which "works" for rank 2); in the case D_4 the rank was 3 — which is too large. In both cases, the plots were not "exact fractals".

On the other hand, the rank for the plot above *is* 2, and the plot above *is an exact fractal*! However, the laws of fractality are *slightly* different from the plots covered in the chapter on p. 33 (since they need to take into account the jumps); we inspect what needs to be changed in Remark 77 on p. 129.

The reason for the jumps above is the sequence N_n being not "fully" distilled. More generally, motives for an abelian polynomial split into "motives of rank 1" — or permutation matrices $\mathscr{M}_p$ can be simultaneously diagonalized in an appropriate basis. In other words, the number of fully distilled parts is equal to the degree of the polynomial. In the case above the degree is 3, so we start with 3 distilled parts, and "what remains after the 'naive distillation' step" is still a fusion of two fully distilled parts.

The "extra" distillation in degree 4 and the "expected" behavior

Our construction of a graph associated to a polynomial P goes through 3 steps:

- Find numbers N_{p^k} following specific recipes (see below).
- Use the formula $N_{mn} = N_m N_n$ valid for mutually prime m and n (*multiplicativity*).

[164] It is easy to verify that this polynomial $3m(m^2 - 1) + 1$ is cyclic. Given a root x, consider $x' := 3x^2 + x - 2$ — which is also a root of P. Moreover, $x'' = -3x^2 - 2x + 2$, and $x''' = x$. Since any Galois symmetry should preserve the mapping ′, it can only make cyclic permutations of the roots x, x', x''. (This is similar to what we did in the D_4-case in the section on p. 122).

- Plot the antiderivative of the Fourier transform $F(t)$ of the sequence (N_n).

So far, we had two descriptions of what the numbers N_{p^k} *actually mean*: one in Footnote 179 on p. 74, and the other in the section on p. 115 (the compatibility of these desciptions was discussed in Footnote 319 on p. 117). While we illustrated these recipes only on examples of polynomials of small degree, it is straighforward[165] to generalize them to arbitrary polynomials.

However, in the preceding section we saw that in case of $\deg P = 4$, while the resulting graphs are *fractal-like*, they are not *exact fractals*. In particular, this *is not the way* to uncover “the hidden symmetries”. This happens because for rank > 2, the hidden symmetries predicted by the the Langlands program act on something which is much harder to describe than the function $F(t)$.

Here we do not try to explain[166] which symmetric object replaces $F(t)$ when rank > 2, but instead we focus on *the exceptional polynomials* P of degree 4 for which *there are* “hidden symmetries with rank $= 2$” (so the Langlands program describes symmetries of *an explicitly provided function*[167]). Recall that above we saw one case (on p. 103), of Galois type[168] D_4, where $F^{(-1)}(t)$ had jumps and “spikes” — and we claimed that they are due to the sequence (N_n) being “not fully distilled”. Moreover, recall that every step of distillation simplifies each sequence (N_{p^k}) (see Remark 45 on p. 75); this decreases the rank.

In the case D_4 the second distillation decreases the rank from 3 to 2 — which is related to exact fractals $F(t)$. If one believes this, plugging the simplified sequence (N_n^{dist}) into the construction above (instead of N_n) should give the Fourier transform $F_{\text{dist}}(t)$ which is an exact fractal — and the fractality laws are *exactly* the same as what we saw in degree three!

Moreover, in the case D_4 it is possible to treat the “extra distillation step” as a very simple “black box”: given the initial sequence (N_{p^k}), the following rules produce $(N_{p^k}^{\text{dist}})$:[169] [170]

(N_{p^k}) before the 2nd distillation	Switch(p)	$(N_{p^k}^{\text{dist}})$
$(3, 6, 10, 15, \ldots)$		$(2, 3, 4, 5, \ldots)$
$(-1, 0, 0, 1, -1, 0, 0, 1, -1, 0, \ldots)$		$(0, -1, 0, 1, 0, -1, \ldots)$
$(-1, 2, -2, 3, -3, 4, -4, \ldots)$	1	$(-2, 3, -4, 5, \ldots)$
	-1	$(0, 1, 0, 1, 0, 1, \ldots)$
$(-1, 1, -1, 1, \ldots)$		$(0, 1, 0, 1, 0, 1, \ldots)$

Note that in all the cases except for $(N_{p^k}) = (-1, 2, -2, 3, -3, 4, -4, \ldots)$ the sequence $(N_{p^k}^{\text{dist}})$ is determined by (N_{p^k}), however, in the exceptional case we also need to know one more bit of information; we call it Switch$(p) = \pm 1$. Anyway, for our example polynomial $x^4 - x^3 - x^2 + x + 1$ of type D_4, this bit may be determined as[171] $\text{Switch}(p) = \pm 1 \equiv_3 p \mod 3$.

Recall that this is the polynomial[172] of type D_4 with the smallest magnitude of the field discriminant, $D = 9 \times 13 = 117$ (and no real roots). The plot of the real and the imaginary parts of the antiderivative

---

[165] We cover the general case in the section on p. 115.

[166] We do this in the section on p. 169.

[167] **N.B. (???) Ref? Need to write explicitly earlier!**

[168] This notion is discussed in Footnote 299 on p. 103.

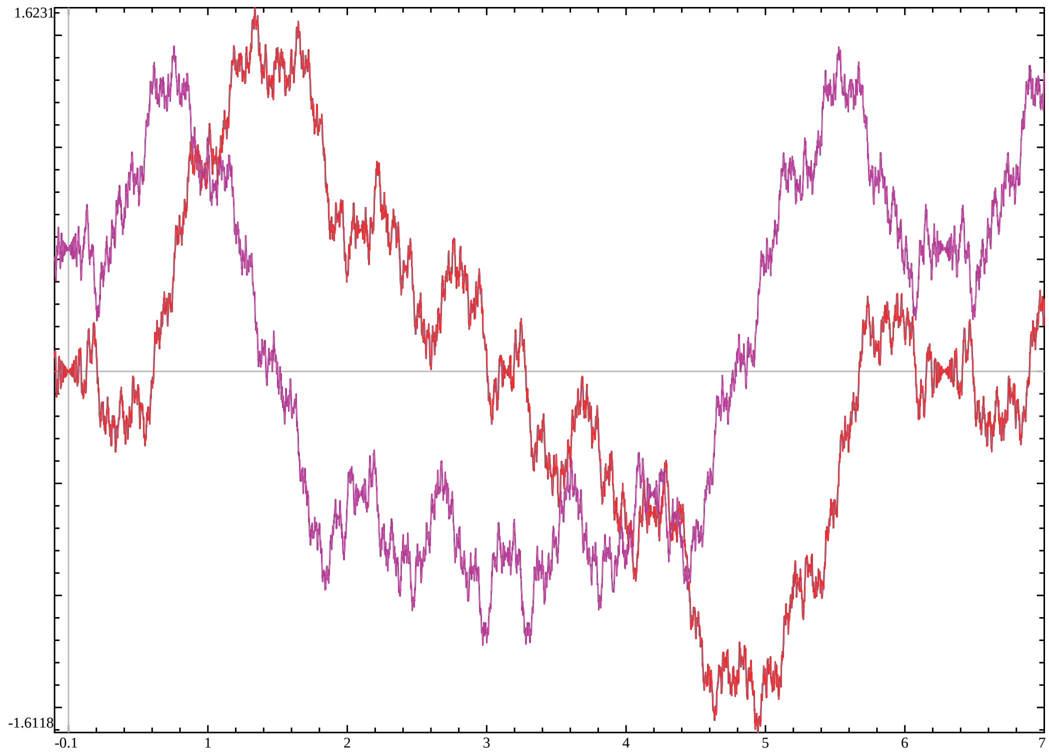
[169] We postpone discussion of these rules until the section on p. 114.

[170] 5 rows in this table correspond to 5 flavors (“conjugacy classes”) of symmetries of a square: rotations by 0°, or by ±90°, or by 180°, or reflections in vertical-or-horizontal or diagonal mirrors. Compare with the table on p. 122.

[171] In general, this is $\left(\frac{\Xi}{p}\right)$, and $\Xi = -3$ for our polynomial. Hence this expression simplifies to $\left(\frac{p}{3}\right)$. See Remark 69 on p. 123.

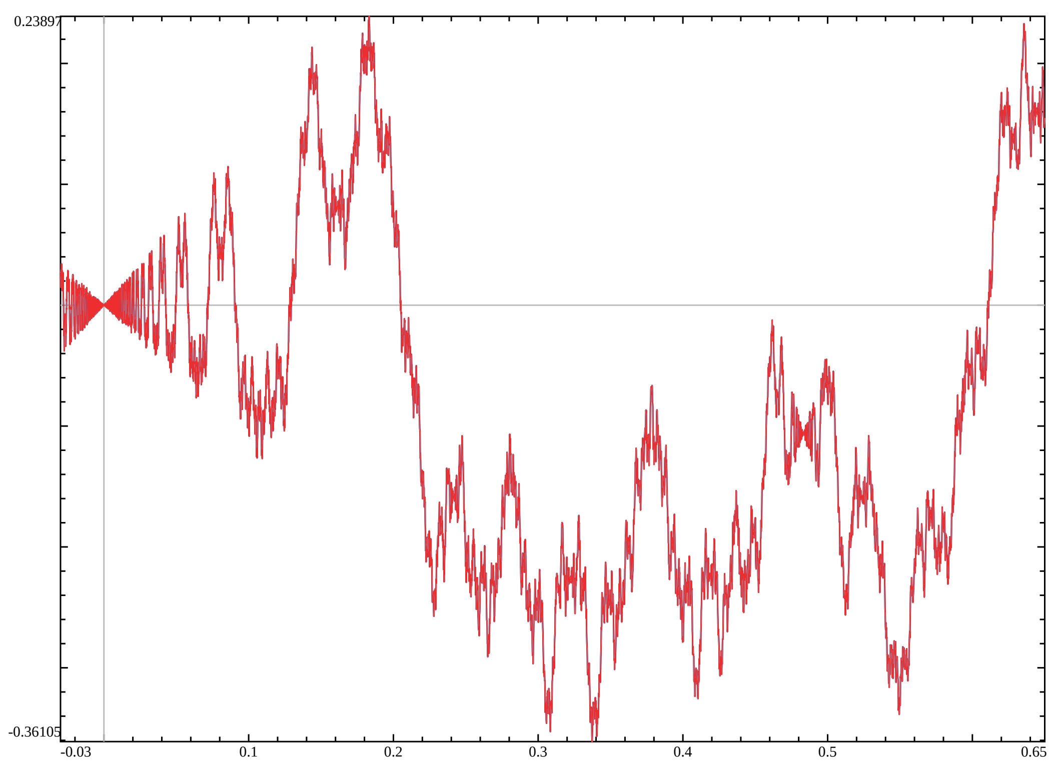
[172] **N.B. (???) Should we consider also $x^4 - x^3 + 2^x - 1$ which has not 2, but 1 signs “-” (2 real roots) in the field discriminant $D = -5^2 \cdot 11 = -275$? Should not this be combined with the sign of Ξ?**
Likewise, $x^4 - x^3 - 3x^2 + x + 1$ has 0 signs “-” (4 real roots), and $D = 5^2 \cdot 29 = 725$.

$F^{(-1)}_{\text{dist}}(t)$ of the Fourier transform $F_{\text{dist}}(t)$ of the sequence (N^{dist}_n):



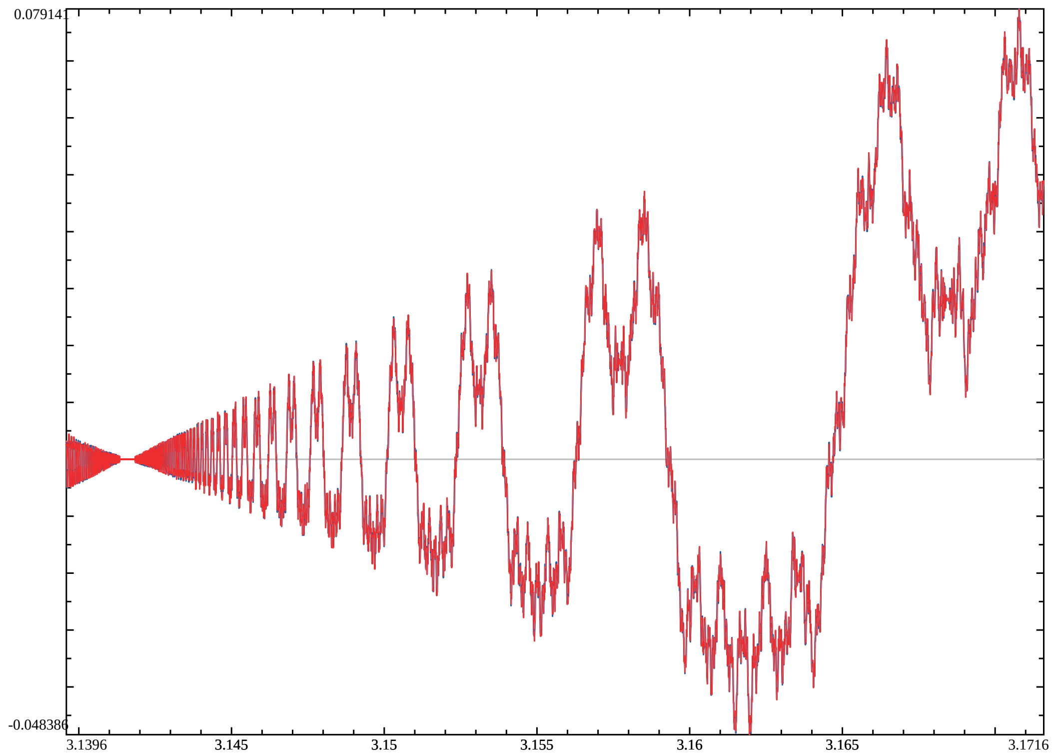
shows clear signs that it is indeed an exact fractal.

A moderate zooming near the origin demonstrates linear asymptotics near 0:

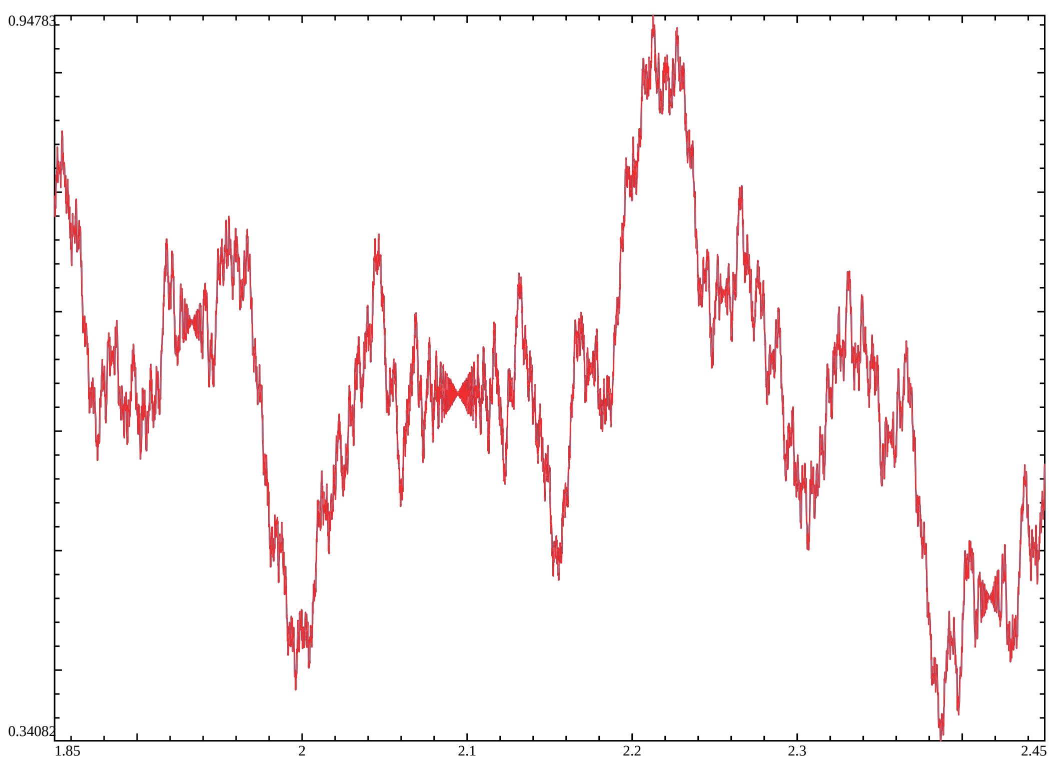


One can also see that all the oscillations have a very similar shape — and this shape coincides with the blue graph on the previous plot.

Near $t = \pi$ one can see the same shape of oscillations:

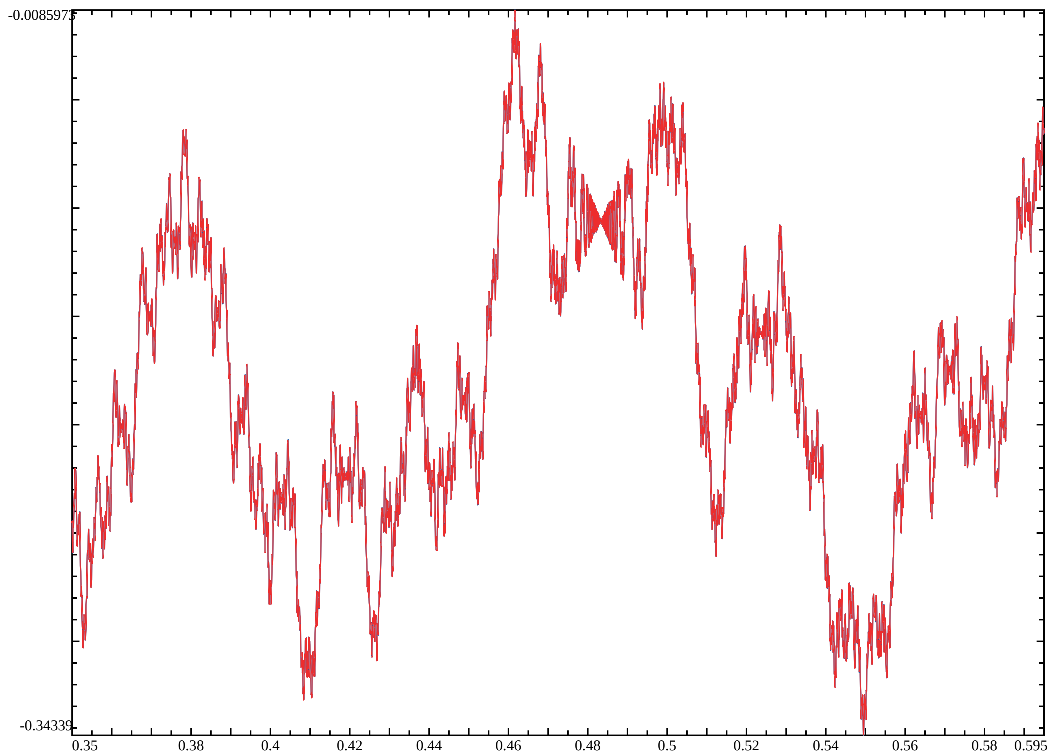


Finally, the behaviour near $t = {}^{2\pi}/_{3}$ shows a shape of oscillations where the top "half" is similar to the bottom "half":



It is not hard to recognize that this shape matches the red graph of the first plot.

Moreover, the zoomed-in graph above for the region of small t demonstrates a very prominent fractal-like feature near the point $t = {}^{2\pi}/_{13} \approx 0.4833$ (this point is in the Cantor hyper-family). Near this point one gets the same shape of oscillations as the blue graph above (only upside-down):



Conclusion: the "extra" distillation led to the antiderivative $F_{\text{dist}}^{(-1)}(t)$ which is horizon-self-similar: it satisfies our "actual" fractality law[173]!

In case D_4 the "extra distilled" $F_{\text{dist}}(t)$ is an exact fractal — in constrast to other cases of $\deg = 4$.

To preview the following discussion: only the abelian cases and the case D_4 allow an extra distillation — and the abelian cases lead to hidden symmetries of rank $= 1$ (as we will see in the section on p. 67 this is somewhat similar to quadratic reciprocity in Euler formulation; moreover, these symmetries "are not fractal").

Finer points of the transliteration rules

> ... Grothendieck constructed them, thus forever changing our understanding of the relationships between continuous and discrete.
>
> Yu. I. Manin, *Mathematics as Metaphor: Selected Essays*, 2008

What we discuss here is an immediate continuation of what we did in the last section of the preceding chapter.[174]

Remark 40: Note that we already know that $\widetilde{N}_p^{\text{res}} = 0$ if a prime number p is red. (As usual, we need to omit a few exceptional p⌟s.) When p is green, we need to decide whether $\widetilde{N}_p^{\text{res}} = 1$ or $\widetilde{N}_p^{\text{res}} = 3$. Above, we said that one should consider residues $\bmod p$ of our polynomial sequence of degree 3; count $0 \bmod p$⌟s among the first p of them. On the other hand, all we need is 1 bit of information to distinguish these two case.

[173] ... from the section on p. 42.

[174] The only reason we made a chapter break in the middle of this discussion was to signal the readers with less stamina that the remaining parts are just clarifications of the process outlined above.

In fact, already in the time of Gauss mathematicians knew how to get this extra bit of information. **Answer:** One should take a certain other sequence *of degree 2*, and color numbers into red and green according to *whether they are divisors of numbers in this second sequence.*

For example, for our sequence "tetrahedral numbers + 2" of degree 3 we should consider the sequence "squares + 971" of degree 2. Now we have *two colors* assigned to a number n: one according to whether n may divide numbers in the first sequence, the other according to whether it can divide numbers in the second sequence. Finally, for prime p one can find $\widetilde{N}_p^{\mathrm{res}}$ from the following table:

		Second color	
		🟩	🟥
First color	🟩	3	1
	🟥	0	✕

(with ✕ meaning "cannot appear").

Now we remind that Quadratic reciprocity says that the second color depends only on the position of p on the "conductor" 971-wheel, similarly to the coloring of the wheel on p. 14. **Conclusion:** One can find $\widetilde{N}_n^{\mathrm{res}}$ knowing the first color of the number n and the position of n on 971-wheel.[175]

Remark 41: The counts $\widetilde{N}_n^{\mathrm{res}}$ form a very fundamental mathematical object, leading to the notion of an L-function — another math tool as important as the functions $F(t)$ and $f(z)$ we discussed above. However, comparing our definition of $\widetilde{N}_n^{\mathrm{res}}$ with the formal definition of the "coefficients" of the corresponding L-function, one can discover that we oversimplified a bit; our definition is "correct" just for "about 61% of indices n"! (We explain it below.) Moreover, removing this "oversimplification" would allow replacing Steps (c)) and (d) on p. 60 above by something much easier to explain.

To see where we "cheated", inspect the particular case $n = 9$. Our (Gauss'!) 9-wheel is an example of a "new" arithmetic which has only 9 different "numbers" (residues mod 9). We can add/subtruct/multiply in this arithmetic; we may also divide by any "number" but 0 mod 9, 3 mod 9 and 6 mod 9. What we do above to find $\widetilde{N}_9^{\mathrm{res}}$ is we "replant" our polynomial to this arithmetic, and look how many times it takes the value 0 mod 9.

However, already in 1830 a French mathematician Évariste Galois (with *very* romantic biography; he was 19 when he published this) found out that there is *another* arithmetic with 9 "numbers" — and in this arithmetic one can divide by *every* number except 0. So Galois' arithmetic is, in a certain sense, "better" than Gauss'![176][177] In fact, to get the fractal behavior, and/or the remarkable properties of L-functions, one *must* use Galois' arithmetic in place of Gauss' when finding $\widetilde{N}_9^{\mathrm{res}}$.

So instead of finding the count $\widetilde{N}_9^{\mathrm{res}}$ of residues where the polynomial takes value 0, we do the same in the Galois arithmetic. Denote these counts $\widetilde{N}_n^{\mathrm{Gal}}$ (here n is a power of prime). However, as we already saw, the residues mod a prime number *already* have the required property: division by any non-0 residue is possible. This leads to $\widetilde{N}_n^{\mathrm{Gal}} = \widetilde{N}_n^{\mathrm{res}}$ provided n is prime; this also works if n is not divisible by any square (except 1^2).[178] It turns out that this holds for about 61% of numbers! (The exact fraction turns out to be $\frac{6}{\pi^2} \approx 0.6079$. We discuss this in the group F of exercises on p. 30.)

Remark 42: We said that to obtain fractal behavior, one must use the counts $\widetilde{N}_n^{\mathrm{Gal}}$ instead of $\widetilde{N}_n^{\mathrm{res}}$. How come that the recipe for N_n given above does not mention $\widetilde{N}_n^{\mathrm{Gal}}$?

---

[175] The cubic polynomial we considered above has discriminant $-3{,}884 = -2^2 \times 971$. This means that finding solutions is related to taking the square root of -971; since $-971 \equiv_4 1$, the conductor wheel related to this square root is the 971-wheel (see p. 213).

[176] Gauss' notebooks show that he also knew about this arithmetic — but he did not publish this.

[177] Essentially, Galois' "numbers" is the answer to the question: what are analogues of complex numbers if one starts with residues mod p instead of reals? (For others results of Galois we use in these notes see Footnote 322 on p. 118.)

[178] For example, $\widetilde{N}_{30}^{\mathrm{res}}$ is OK, but $\widetilde{N}_{60}^{\mathrm{res}}$ needs to be recalculated, since $2^2 = 4$ divides 60.

In fact, polynomials of degree ≤ 3 are very special: one can find $\widetilde{N}_{p^k}^{\mathrm{Gal}}$ for prime p provided one knows $\widetilde{N}_p^{\mathrm{res}}$. Moreover, this is almost exactly the process we apply on Step (c) on p. 60! **Conclusion:** Step (c) hides recalculation from $\widetilde{N}_{p^k}^{\mathrm{res}}$ to $\widetilde{N}_{p^k}^{\mathrm{Gal}}$. One can omit this step if one uses suitable formulas like $N_{p^2} = (\widetilde{N}_{p^2}^{\mathrm{Gal}} + (\widetilde{N}_p^{\mathrm{Gal}})^2)/2 - \widetilde{N}_p^{\mathrm{Gal}}$ etc.[179]

Remark 43: The recipes of Steps (c)) and (d) on p. 60 look coming out of a clear blue sky. In fact, they come from a very general principle:

$$\boxed{\text{The numbers } \widetilde{N}_{p^k}^{\mathrm{Gal}} \text{ satisfy a simple recursion relation in } k.}$$

(This relation is very similar to one for Fibonacci numbers: $F_n = F_{n-2} + F_{n-1}$; for examples, see Footnote 184.) In fact, the same holds for the sequence (N_{p^k}).[181] Additionally, instead of our polynomial of degree 3, one can take any polynomial; the same works for polynomials of any number of variables[182] — and even when one counts "common zeros": arguments where several polynomials all take value 0 mod p^k (or 0 in Galois' arithmetic).

The simplicity of these statements is completely deceptive. It turns out that they constitute another triumph of mathematics of 20th century. To make a long story short: in 1973 a Belgian/French mathematician Pierre Deligne finished his proof of Weil Conjectures (which were invented about 25 years before this). The conjectures (and the proof) are based on a revolutionary approach erasing boundaries between geometry and arithmetic. (Compare with the epigraph to this section: Yu. I. Manin writes that this "forever chang[ed] our understanding of the relationships between continuous[183] and discrete.") In fact, these recursion relations make a significant part of these conjectures.

In case of our polynomials of degree 3, the recursion relations simplify so much that we can write down all the possible solutions. This is what we did in Steps (c)) and (d) on p. 60.[184]

Warning: quite often in math, when there is a recursion relation between *counts* of objects of certain types, they come from simple "matching arguments": the relation between counts reflects "relations between *individual objects*". However, Weil relations between "counts of solutions" are much deeper: the *solutions themselves* have no relation to each other!

Unfortunately, with this topic our intuition can easily deceive us: what gets in the way is that *for residues* mod p^k, there is an obvious "connection" between *nearby* values of k: a particular residue mod p^k gives us a residue mod p^{k-1}. Contrarily, an analogous "connection" between Galois' arithmetics has very different properties:

$$\boxed{\text{The ``related'' powers } p^k \text{ and } p^l \text{ of } p \text{ are ``far away'': the connection works}^{185} \text{ only if } k|l.}$$

---

[179] The first term on the right-hand side is not as mysterious as it looks like. Denote it by $\overline{N}_{p^2}$. Then the general formula is $1 + \sum_k \overline{N}_{p^k} u^k = \exp \sum_k \widetilde{N}_{p^k}^{\mathrm{Gal}} u^k / k$ (equality of Taylor series in u).

The subtraction of the second term $\widetilde{N}_p^{\mathrm{Gal}}$ is harder to explain, since it is due to distillation process of Step (b) on p. 60. Essentially, in the formula above we may replace $\overline{N}$ by N if we replace[180] $\widetilde{N}_{p^k}^{\mathrm{Gal}}$ by $\widetilde{N}_{p^k}^{\mathrm{Gal}} - 1$. (One can also define $\overline{N}_n$ for composite n by multiplicativity, as in Step (e) on p. 60.)

[180] Since $\exp(u + {}^{u^2}\!/2 + {}^{u^3}\!/3 + \ldots) = 1 + u + u^2 + u^3 + \ldots$, this gives the formula $1 + \sum_k \overline{N}_{p^k} u^k = (1 + u + u^2 + u^3 + \ldots)\left(1 + \sum_k N_{p^k} u^k\right)$, or $N_{p^n} = \overline{N}_{p^n} - \overline{N}_{p^{n-1}}$. In the sections on p. 115 and on p. 120 we use this formula as $1 + \sum_k N_{p^k} u^k = (1 - u)\left(1 + \sum_k \overline{N}_{p^k} u^k\right)$.

[181] Moreover, the same also holds for $\widetilde{N}_{p^k}^{\mathrm{res}}$ — but this is trivial: for most p the numbers $\widetilde{N}_{p^k}^{\mathrm{res}}$ do not depend on k.

[182] Compare with the beginning of the section on p. 20.

[183] Unfortunately, the cases we consider here, of 1 equation with 1 unknown, result in the sets of solutions of dimension $1 - 1 = 0$. One needs to go to cases of dimension ≥ 1 to see the "clearly geometric"="continuous" facets of the Weil conjectures.

[184] The relations boil down to $N_{p^{k+2}} = aN_{p^k} + bN_{p^{k+1}}$, with (a, b) being $(-1, -1)$, $(1, 0)$, $(-1, 2)$, $(0, 1)$ and $(0, 0)$ in 5 cases of Steps (c)) and (d) on p. 60. (If we do not know which case is applicable, then the Weil conjectures do not predict anything better than "the merge" of these recursion relations $N_{p^{k+6}} = N_{p^{k+5}} + N_{p^{k+4}} - N_{p^{k+2}} - N_{p^{k+1}} + N_{p^k}$.)

(For example, Weil conjectures may claim that there is a relation between *the counts* of solutions of a certain (system of) equations in the Galois' replacements for $\bmod p^3$, $\bmod p^4$, $\bmod p^5$. On the other hand, the only things in common between these "arithmetics" are the p residues $\bmod p$ — hence one cannot "match" the solutions themselves.[186]) *This* is the reason why Weil conjectures are so deep (and, for many people, much deeper than they seem to be on the first sight).

Remark 44: Nowadays, one could consider Weil relations as the first tiny but general enough step in the direction of the Langlands approach. It looks like Weil arrived at these relations by doing many "numerical experiments".

To see how revolutionary all this was at the time, note that when Hasse conjectured what is essentially the next step,[187] Weil himself did not believe that Hasse conjectures can keep water.[188]

"Distillation" and Motives

The operation we did on Step (b) on p. 60 looks very innocuous: all we do is subtracting 1. In fact, an explanation of *why* this leads to appearance of fractal properties is related to very deep branch of mathematics of today, *Theory of Motives*. It is a very hot and not yet fully settled down theme in contemporary math.

Essentially, "the motive of zeros of our polynomial" can be "distilled" to two independent parts. Each "distilled" part has its own symmetries (maybe "hidden"), but these symmetries are so different that when they are "overlapped" on top of each other, no recognizable pattern remains. (This is similar to playing two very different pieces of music at the same time: if they are sufficiently dissimilar, no theme would remain recognizable. We will clarify these notions in Remark 66 on p. 120.)

This section turns out to be the most technical in the first half of this report.[189] I did not find a way to make it simpler; however, nothing else in this report depends on the explanations of this part, so feel free to skip it altogether.

Remark 45: One way to explain what happens in Step (b) on p. 60:

There is a hierarchy of "difficulty" of sequences, and:

"Distillation" means: "remove" from the given sequence any trace of "simpler" sequences.

The main idea is that the result of distillation is much simpler to deal with than the initial sequence.

Sequences simpler than degree=3 are sequences of degree 0, degree 1 and degree 2. **Conclusion:** in our sequence $\widetilde{N}_n^{\mathrm{res}}$, we need to

- find "the traces" of "$\widetilde{N}_n^{\mathrm{res}}$ for sequences of degree 2",
- find "the traces" of "$\widetilde{N}_n^{\mathrm{res}}$ for sequences of degree 1" (and degree 0), and
- subtract these traces from our sequence $\widetilde{N}_n^{\mathrm{res}}$.

Essentially, we want to write down our counts $\widetilde{N}_n^{\mathrm{res}}$ related to a sequence of degree 3 as

$$\widetilde{N}_n^{\mathrm{res}} \equiv T_n + (\widetilde{N}_n^{\mathrm{res}} - T_n),$$

[185] Another difference: this connection goes in "the opposite direction" comparing to one with residues: an element for smaller p^k induces an element for *larger* p^{km}.

[186] Moreover, the real show-stopper is that these 3 arithmetics have "interesting sets of symmetries" — but these symmetries are "not compatible". This alone breaks any attempt to "match" solutions between these arithmetics — except for the solutions which already exist $\bmod p$. (We discuss such symmetries in Footnote 322 on p. 118.)

[187] This probably happened before WWII. The simplest case of this conjecture was proven about 20 years ago — essentially, together with the proof of Fermat's Last Theorem.

[188] Later he changed his mind and confirmed the conjecture in a few cases — and now it is named "Hasse–Weil conjecture".

[189] Our calculations with Eisenstein series on p. 135 are yet much more technical. So is our *second* round of attack on the topic of distillation in the section on p. 120.

with T_n being a combination of counts related to sequences of degree 0, 1 or 2, and $\widetilde{N}_n^{\mathrm{res}} - T_n$ "having no similarity to counts $\widetilde{N}_n^{\mathrm{res}}$ related to sequences of degree 0, 1, or 2".

Remark 46: It turns out that it is easy to characterize sequences "having no similarity to counts $\widetilde{N}_n^{\mathrm{res}}$ related to sequences of degree 0, 1, 2":

The average value of such a sequence on primes in any arithmetic progression is 0.

(We must ignore progressions containing just one prime. This happens when the step is not mutually prime with the elements; otherwise the progression contains infinitely many primes.) In other words, the sequence C_k is of this form if the average value of the sequence C_{ak+b} restricted to prime values of $ak+b$ is 0 provided $a>0$ and a and b are mutually prime.[190]

Moreover, putting $T_n \equiv 1$ in the formula above achieves the goal:

The average value of the sequence $\widetilde{N}_n^{\mathrm{res}}$ on primes in any arithmetic progression is 1.

Indeed, numbers $T_n \equiv 1$ are counts of $0 \bmod n$s related to the sequence 1,2,3,... of degree 1. Indeed, in residues $\bmod\, n$ the shortest period of this sequence has length n, and the count of $0 \bmod n$s in this period is exactly $T_n = 1$. This leads to

There is no trace related to degree 2. The trace related to degree 1 is $T_n \equiv 1$.

Clearly, this immediately leads to the rule of Step (b) on p. 60. As a result of subtracting these T_n, the counts $\widetilde{N}_p^{\mathrm{res}} = 0, 1, 3$ at prime indices p become -1, 0, and 2.

Remark 47: It is not that hard to explain the meaning of the rule in the red frame.

First of all, degree 0 leads to $\widetilde{N}_p^{\mathrm{res}} = 0$ for most of primes p — so we may forget about it.[191] Note that degree 1 leads to $\widetilde{N}_p^{\mathrm{res}} = 1$ for most of primes p.

Next, recall the pattern we observed for "$\widetilde{N}_n^{\mathrm{res}}$ for sequences $a_n^{②}$ of degree 2": it appears when we write numbers in a suitable number of columns. Every column is an arithmetic progression with the step equal to the conductor for $a_n^{②}$, and:

The value $\widetilde{N}_p^{\mathrm{res}}$ on primes p in any such arithmetic progression is the same (for degree=2).

Moreover, one can show that this is "the whole pattern": a similar average in other arithmetic progressions is 0 unless the progression is related to the columns (which means: its step is not mutually prime with the conductor). And: the same rule works for degrees 0 and 1.

Conclusion: to "distill", all we need to do is to avoid the pattern in the box above. Note that any finite sequence can be written as "a constant sequence" + "a sequence with average 0", and these two parts are "orthogonal" to each other. Although we deal with infinite sequences, a similar approach still works — and this leads to the rule in the red frame.

---

190 As usual, we needed to over-simplify a bit. In fact the framed rule describes not the dichotomy "the degree is 0, 1 or 2" vs. "the rest", but a related dichotomy *abelian* (or even *cyclic*; it happens for degree up to 2, as well as "in some cases of higher degree") vs. *purely-nonabelian* (which may be restated as "covered by the Class Field Theory for $\mathbb{Q}$" vs. "needing the Langlands program"; compare with the section on p. 82). However, since anything of degree 0, 1, or 2 lives in the "abelian" realm, and we do not consider the abelian case of degree 3 (except for Remark 78 on p. 130), this is enough for our purposes.

191 On the other hand, tuning an equation of degree 0 (such as an equation $35 = 0$ in an unknown m — which does not enter the equation!) allows us to get "exceptional counts of solutions" in a prescribed list of prime (such as $p = 5, 7$ in the example above: any m is a solution modulo such p). This shows that when "removing traces of degree 0" allows to ignore a few "exceptional values" of p where the general approach gives "wrong answers" for the number of solutions.

Given a general sequence ν_n, how to "distill" it, making the average of ν_p on primes p in every arithmetic progressions to be 0 after distillation? It looks like for every arithmetic progressions we need to subtract the "averaged" value in this progression.[192]

This may look like a hopeless task: progressions with different steps may overlap, and dealing with such overlaps may lead to contradictions. Miraculously, this does not happen for N_n: the average in every progression is the same: and is equal 1 (see the green frame above!).

Remark 48: To illustrate the miracles which must happen to have the "average over primes" in all arithmetic progression to be the same, 1, note that the most interesting arithmetic progressions related to our example above ("tetrahedral numbers + 2") have step 971. Indeed, in Remark 40 on p. 72 we saw that the counts $\widetilde{N}_p^{\mathrm{res}}$ for prime p are controlled by our green/red colors, *and* by the position on 971-wheel. As the table in this remark shows, in some of these progressions only the count 1 appears, while in the others only the counts 0 and 3 appear.[193]

Obviously, the average of $\widetilde{N}_p^{\mathrm{res}}$ for prime p in the former kind of the progression is 1. It is very natural to expect that the average for the latter kind is going to be different (such as 1.5) — but this does not happen![194]

This was discovered about 150 years ago — it was the first precursor of the Langlands Program (a very remote precursor!). This is called Chebotaryov's density theorem.[195] In our case, it says that $\widetilde{N}_p^{\mathrm{res}} = 1$ for $\frac{1}{2}$ of primes p (this matches[196] our claim that $\frac{1}{2}$ of 970 arithmetic progressions have "the second color red" in the table of Remark 40 on p. 72), while $\widetilde{N}_p^{\mathrm{res}} = 0$ for $\frac{1}{3}$ of primes p and $\widetilde{N}_p^{\mathrm{res}} = 3$ for $\frac{1}{6}$ of primes p.

So *under the condition* "the second color is green" in the table of Remark 40 on p. 72, $\frac{2}{3}$ of the primes are going to have $\widetilde{N}_p^{\mathrm{res}} = 0$, and the remaining $\frac{1}{3}$ of them have $\widetilde{N}_p^{\mathrm{res}} = 3$. Obviously, this leads to the average being 1. (We return to this topic in Remark 105 on p. 157.)

Remark 49: The discussion above leads to another question: what happens *inside* one of 970 arithmetic progressions of the preceding remark? (We excluded the dull one, where the only prime is 971.) Half of them are not very interesting: only $\widetilde{N}_p^{\mathrm{res}} = 1$ appears there (for prime p) — so the all the primes there are green. In the other half, primes are red and green matching $\widetilde{N}_p^{\mathrm{res}} = 0$ and $\widetilde{N}_p^{\mathrm{res}} = 3$.

For example, in the progression starting with 1 (so it goes 1, 972, 1,943, 2,914, 3,885, 4,856, ...) the sequence of colors is

Here ○ stand for non-prime numbers. Removing even numbers (which are always ○) allows to see more colored circles (note that the step now is 2×971, so this progression goes as 1, 1,943, 3,885, ...):

To see yet more colored circles, remove all ○s, leaving just the prime numbers:

This did not help: still no pattern is visible![197]

Conclusion: inside this arithmetic progression there is no observable pattern.

---

[192] This is the limit of averages on longer and longer parts of this sequence.

[193] Actually, out of 971 possible progressions with this step, one does not have infinitely many prime numbers (just one prime 971). Out of 970 remaining progressions, half are of one kind, half of the other.

[194] We already mentioned this in Remark 38 on p. 61.

[195] In fact, for our question, the earlier version discovered by Frobenius is enough.

[196] Well, this also involves the Dirichlet theorem on primes in arithmetic progressions.

[197] ... except the approximate pattern we mentioned above: the ratio of green:red is close to 1:2. Indeed, the observed value is 26:57 — which is reasonably close to 1:2 for a sample of size less than 100. If we consider longer and longer chunks of our sequence, the ratio would go closer and closer to 1:2.

On the other hand, there are transliterations rules similar to those discussed above which are tuned to these sequences of colors •, • and ∘. They translate them to numbers M_n such that their Fourier transform $G(t) := \sum M_n e^{imt}$ *also* has fractal properties. (However, the fractality law for $G(t)$ is a bit more complicated than what we considered above: it has extra factors of the form $\exp i(at + b/t)$.)

Conclusion: to expose the patterns in colors in these arithmetic progressions one needs to go through the steps very similar to those for our initial sequence of red/green colors. There is no simplification due to restriction of attention to such progressions!

Fractional-linear transformations

In the section on p. 43 we constructed an example of a 2π-periodic function $g(t)$ which has a horizon-self-similar point $t = 0$.[198] As we saw, the non-smooth points of any such function $G(t)$ are images of $t = \infty$ under chains of transformations[199] $t' = -1/\gamma t$ and translations $t' = t \pm 2\pi$ (with a certain fixed γ; what we used was $\gamma = \pm 2/\pi$ — the choice of the sign is irrelevant since our seed function g_0 was odd).

Remark 50: In fact, chains of transformations $t' = -1/\gamma_0 t$ and $t' = t \pm 2\pi$ may be controlled to some extent: these chains may result only in transformations of a very specific form. Indeed, both transformations can be written as $t' = \frac{\alpha t+\beta}{\gamma t+\delta}$, one with $\alpha_0 = \delta_0 = 0$, $\beta_0 = -1$, the other with $\alpha_1 = \delta_1 = 1$, $\beta_1 = 2\pi$, $\gamma_1 = 0$. Since composition of such (*fractional-linear*) transformations is again a fractional-linear transformation, any chain of toy-transforms and shifts results in a fractional-linear transforms of t.

Moreover, if $4\pi^2\gamma_0$ is an integer,[200] then using the new variable $T = t/2\pi$, these fractional-linear transformations are going to have integer coefficients α, β, γ, δ.[201]

Remark 51: It turns out that if $\pi^2\gamma_0 > 1$ (as in the example in the section on p. 43, where $\gamma_0 = 2/\pi$, and as in all examples related to divisors of polynomial sequences), then there are other restrictions.

---

[198] Note that for self-similarity, we needed to use the imaginary unit i as a scaling factor. If we want to have real scaling factors, then what we constructed is a pair of functions $\operatorname{Re} G(t)$ and $\operatorname{Im} G(t)$ such that near $t = 0$, $\operatorname{Re} G(t)$ is horizon-similar to $-\operatorname{Im} G(t)$, and $\operatorname{Im} G(t)$ is horizon-similar to $\operatorname{Re} G(t)$.

[199] The $-$-sign is very convenient. With it, the transformation is (locally) increasing; moreover, it enables the relation $\alpha\delta - \beta\gamma = 1$ used below. (See also Footnote 222 on p. 85.)

[200] This is what happens in examples related to divisors of numbers in polynomial sequences, when $\gamma_0 = c/4\pi^2$ with c being the conductor.

[201] To understand the example graphs below, it is crucial that one can say more. Call a fractional-linear transformation $T' = \frac{\alpha T+\beta}{\gamma T+\delta}$ with integer coefficients, with $\alpha\delta - \beta\gamma = 1$ and with $c|\gamma$ "*congruence*", and with extra conditions $\alpha \equiv_c 1$ (then automatically $\delta \equiv_c \alpha$) "*strongly-congruence*". Then any transformation τ we may encounter in chains as above is either strongly congruence, or $\tau \circ (-1/cT)$ is strongly congruence. (Here $c = 4\pi^2\widetilde{\gamma}_0$, here the base transformation is written as $t' = -1/\widetilde{\gamma} t$.)

Actually, it is very important for us that the strongly congruence transformations form a "sufficiently small" collection of fractional-linear transformations: this makes the tesselations of the section on p. 87 possible. The Lobachevsky-rotations sending one "tile" of tesselation to another one coincide with the strongly-congruence transformations.[202]

The Langlands program predicts that *any* congruence transformation gives a fractal symmetry of $F(t)$ (possibly changing the sign of oscillations). About half of them (including all strongly-congruence) preserve the sign as well (compare with Footnote 206 on p. 80).

[202] Moreover, for $c > 4$ the arguments in Remark 51 on p. 78 show that just a tiny part of the collection of strongly-congruence transformations may be formed by chaining $T' = -1/cT$ and $T' = T \pm 1$.[203]

Indeed, the latter collection was already discussed in Remark 30 on p. 47; as we saw, the corresponding horizon-self-similar points avoid certain intervals. (There is no such avoidance when one considers *all* strongly-congruence transforms; see Footnote 204. We discuss such an example in Remark 52 on p. 80.)

With the pictures of the section on p. 87 one will be able to see that chaining $T' = -1/cT$ and $T' = T \pm 1$ corresponds to "walking" between the gray disks through the tangency points. Moreover, since the green lines separate these gray disks, from a particular gray disk one cannot walk to *all* the gray disks. (See Footnote 237 on p. 88.)

[203] Note that the transform $T' = -1/cT$ is not a strongly-congruence (and not even congruence!). However, combinations as above involving an even number of these transforms are going to be strongly-congruence.

The preceding remark restricts the transformation obtained by chaining both qualitatively (they should be fractional-linear) and quantitatively (see footnotes: there are divisibility properties related to the conductor c). However, there is another metric as well: look at how far the image of the point $t = 0$ (if it exists) can go from multiples of 2π.

Indeed, if a point near 0 is inside $|t| < \pi - \varepsilon$, then its non-trivial translations by multiples of 2π are in $|t| > \pi + \varepsilon > {}^{1}/_{\pi\gamma} + \varepsilon$, hence applying $t' = {}^{1}/_{\gamma t}$ to these points sends them again into $|t| < \pi - \varepsilon$ (with an appropriate choice of ε). (Compare with what we do on p. 46.) Hence starting with $t = 0$, shifting by multiples of 2π, and applying $t' = {}^{1}/_{\gamma t}$ (and combining these transformations in an arbitary order) would never get the point further than $\pi - \varepsilon$ from a multiple of 2π. Hence there is going to be a zone $(\pi - \varepsilon, \pi + \varepsilon)$ which the image of 0 cannot visit![204]

In fact, we already saw this effect in pictures of the section on p. 43, when such a "prohibited zone" appeared as a "smooth" zone in the graph near $t = \pi$. In the following section (on p. 45) we saw that going from a "family" to "super-family" to "super-duper-family" etc. could never extend these sets close to the boundary of $[-\pi, \pi]$.

The moment we know one such "prohibited" zone appears, one can proliferate this zone along $\mathbb{R}$ using the transformations above. This puts a "copy" of such a zone between any 2 given "possible images of 0", hence these copies "appear everywhere": near any point of $\mathbb{R}$, there is such a "prohibited zone". In fact, "possible images of 0" form what is called a "meagre" subset of measure 0.

Prime conductors and "Tetrahedral + 2" again

Recall that when discussing the graph for $F^{(-1)}(t)$ for the polynomial "Tetrahedral numbers+2", we eventually abandoned plotting this function near horizon-self-similar points: it is not computationally feasible. So we could not fully demonstrate that our description of the visual Langlands pattern works for this function. (We needed to switch to a polynomial $6 \times \text{Tetrahedral} + 1$ with a much smaller conductor to do so.) Recall that this description (stated on p. 41 before Remark 27) can be summarized as:

$$\boxed{\text{Near every } t \text{ there is a number } 2\pi{}^{R}/_{S} \text{ which is a horizon-self-similar point for } F(t).}$$

(Reminder: the "actual" transform for $F(t)$ implies "the honest transform for antiderivative" for $F^{(-1)}(t)$; see p. 42.)

However, the conductor $c = 971$ for "Tetrahedral + 2" is a prime number. It turns out that for prime conductors, there is a very simple and powerful generalization of this pattern. It is especially strong if the discriminant d is positive (in other words: if the polynomial has 3 real roots; the "Maass case" of Remark 18 on p. 36):

$$\boxed{\text{If } c \text{ is prime and } d > 0, \text{ then every number } 2\pi{}^{R}/_{S} \text{ is horizon-self-similar.}}$$

For negative discriminant (the "modular form case"), the situation can be described as

$$\boxed{\text{If } c \text{ is prime and } d < 0, \text{ then every number } 2\pi{}^{R}/_{S} \text{ is horizon-similar to either } F(t), \text{ or } \operatorname{Im} F_{\mathbb{C}}(t).}$$

We already saw indications of this in our plots of $F^{(-1)}(t)$ near $t = 0$: these plots were fractal transforms of $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$. Now we know that something similar is going to happen for every rational point: $F^{(-1)}(t + 2\pi{}^{R}/_{S})$ is going to be (up to additive constant) similar to[205] the toy transform either of a shift $F^{(-1)}(t + C_{2\pi^{R}/_{S}})$ of $F^{(-1)}(t)$, or to the toy transform of a shift of $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$.

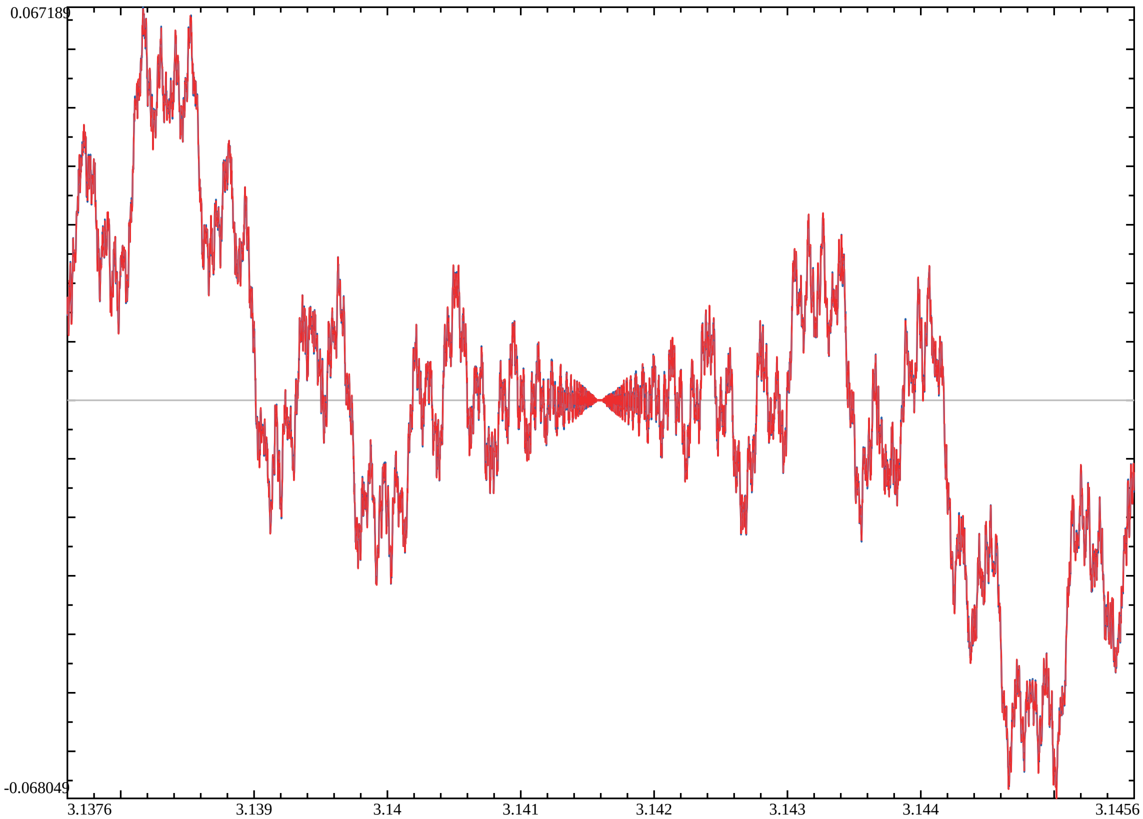
---

[204] Compare with strongly congruence transformations: it is not hard to see that for any c, one can make the image ${}^{\beta}/_{\delta}$ of 0 to be arbitrarily close to any given number.

[205] The first case happens when c divides S. Note how the transform $t \mapsto -{}^{1}/_{|c|t}$ exchanges this subset of $\mathbb{Q}$ and its complement.

(If c is not prime, this happens when c divides S, while *the other* case happens when S is mutually prime with c. In particular, there are yet other cases!)

Going back to the case of the prime conductor 971 with $d < 0$: while reaching horizon-self-similar points requires zooming about c^2 times, many points which are "horizon-similar to the imaginary part" may require much smaller magnification. For example, here is what happens for ${}^{R}\!/\!{}_{S} = {}^{1}\!/\!{}_{2}$:



Comparing to the graph near $t = 0$ on p. 50, one can observe 3 differences:

- The "oscillating zone" is half as wide for the new graph.
- The sign of oscillations is inverted.[206] Indeed, focus on the right half of the graphs; the minima on the new graph match in shape the maxima on the old graph.
- To match these two graphs, one needs a non-linear "transform of the variable t". Indeed, the outermost of the minima on the new graph is about 3 times as far from the "center" as the next minimum (and the next such ratio is about $1{}^{2}\!/\!{}_{3}$). For the maxima on the old graph, the corresponding ratio is about 2 (and the next one is about 1½).[207]

Remark 52: We want to stress that all the preceding examples of graphs of $F^{(-1)}(t)$ but one on p. 58 and the last one were for $t \approx t_0$ with t_0 for which the horizon-similarity could be explained by chaining the transformation $T' = -{}^{1}\!/\!{}_{cT}$ of Hecke's functional equation[208] and the translations by multiples of 2π (which preserve $F(t)$ by definition). This means that horizon-similarity at these points t_0 could have been discovered during the half-a-century between Hecke's discovery and the rise of the Langlands program.[209]

---

[206] It turns out that this is due to 2 mod 971 being not a square. With Legendre symbol from p. 208, $\left(\frac{2}{-971}\right) = \left(\frac{2}{3}\right) = -1$.

[207] This non-linear transformation is fractional-linear (see p. 78): $T \mapsto {}^{1}\!/\!{}_{2} + {}^{T}\!/\!{}_{2(971T+2)}$; here $T = t/2\pi$.

The reader may find it interesting that composition with non-linear transformation $T \mapsto {}^{T}\!/\!{}_{2(971T+2)}$ sends an odd function $F^{(-1)}(t)$ to an *odd* function $F^{(-1)}(t+\pi)$. This cannot happen for an arbitrary odd periodic function; this reflects extra "fractal" symmetries of F.

[208] See the section on p. 82 for details.

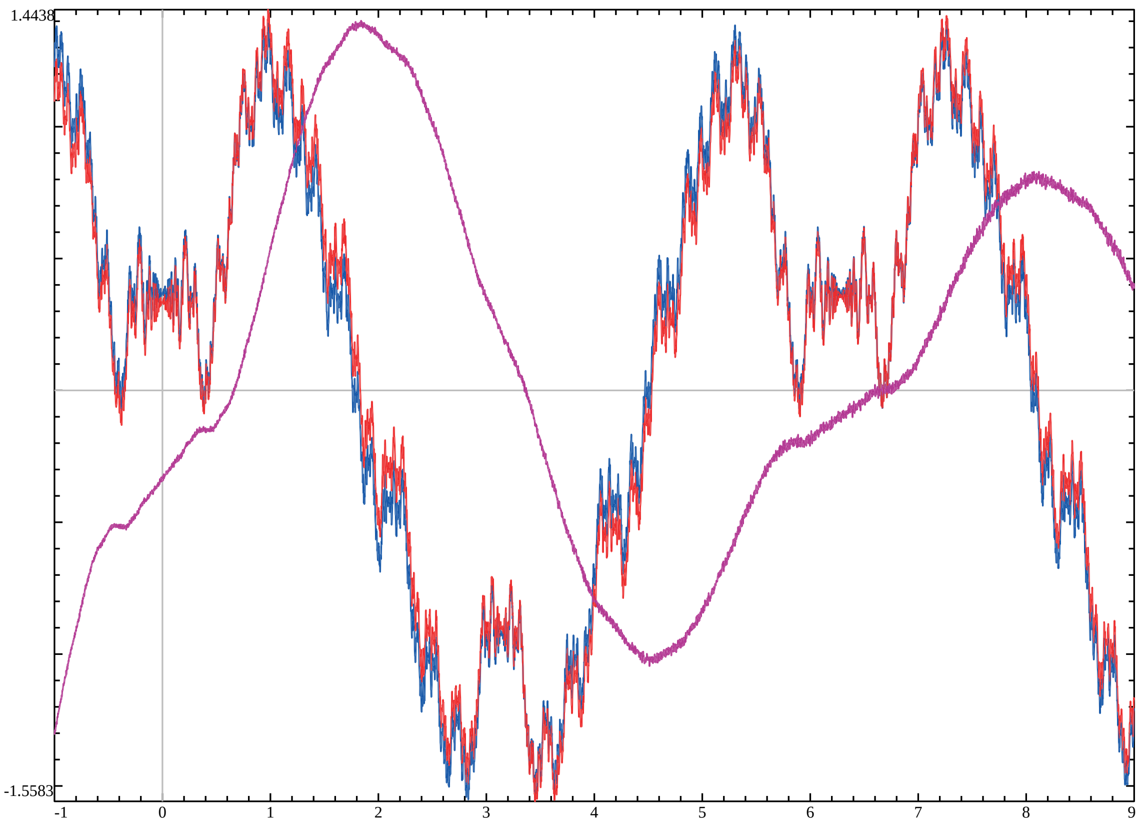
[209] I do not know whether such observations were actually made during this period.

However, the last graph illustrates a “hidden symmetry” which (as far as I know) cannot be explained by Hecke’s result alone.[210] (In fact, the majority of points $2\pi^{R}/S$ are of this type; see Footnote 202 on p. 78.)

The honest fractality law for $F^{(-1)}(t)$

Above, on p. 49, we claimed that the fractality law for the antiderivative $F^{(-1)}(t)$ is “almost visually indistinguishable” from the toy fractality law. In particular, $F^{(-1)}(t)$ is very similar to a toy transform of a suitable function.

Example: (matching the discussion on p. 53): the red curve is the plot of the toy transform of $F^{(-1)}(t)$, the blue curve plots $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$ (for conductor 23),[211] and violet plots the difference:[212]



The graph for difference is scaled up 10 times;[213] it is, obviously, completely “negligible”. Moreover, it is much smoother than the functions we subtract. Obviously, without plotting the red and blue graphs “on top of each other” there would be no way to tell them apart.

[210] We need to repeat: since this is the case of negative discriminant, it is covered by the Class Field Theory for imaginary quadratic fields. So this particular case of horizon-similarity could have became known about a decade after Hecke (but it is doubtful people noticed it before 50s). For more details, see the section on p. 82.

[211] Compare with Footnote 128 on p. 54.

[212] The “thickness” of the graph of difference is a result of numerical errors due to ignoring the higher Fourier coefficients. It decreases roughly as the inverse of the number of terms to sum. The actual graph is quite smooth — but even 16,000,000 Fourier coefficients are not enough to demonstrate this! (Recall that this is *the simplest case*, with very small conductor, 23. To do a similar graph with higher precision, or a larger conductor would require prohibitive time for computation, of order of magnitude of weeks — or I would need to add features like FFT to the software I use.)

On this graph one can also recognize that $F^{(-1)}(t)$ is proportional to the derivative of the “negligible” term — as it should be, due to “integration by parts” (see below).

[213] To visualize this scale of magnification: observe that where the violet graph is positive, the red graph is (slightly) above the blue one. — And likewise for where the violet graph is negative.

Calculation: Assume that F and $\widetilde{F}$ are related by the "actual" fractality law, so $F(t) = \widetilde{F}(1/t)/t$. Integrate by part, denoting the antiderivatives of F and $\widetilde{F}$ by g and $\widetilde{g}$ (so $F(t) = g'(t)$ and $\widetilde{F}(t) = \widetilde{g}'(t)$). Then $g(t) = -t\widetilde{g}(1/t) + \mathrm{Rest}(t)$ (here $\mathrm{Rest}'(t) = \widetilde{g}(1/t)$). These three terms are exactly what is plotted above. These relations explain the observations above.[214]

Conclusion: the fractality law for $F^{(-1)}(t)$ has two terms — and the principal one is exactly the "toy fractality law". The remaining term is "negligible": on our graph, its contribution cannot be seen — with one exception.

Indeed, all that the "negligible" term does is "moving" the features of the graph up and down a bit. The reason for this is that this term is much more smooth than the principal term. Essentially, comparing to wild variations of values of $F^{(-1)}(t)$ in any small region in t, this extra term is practically constant. Hence adding this term would just move the graph up or down.

Remark 53: Of course, moving the features up or down *too far* may make the "visual pattern of toy transform" harder to recognize. Compare with the small plot on p. 57.

Historical approach: cases that *only* the Langlands program can explain

In these notes we illustrate one application of the Langlands program: based on the list of divisors of values of a cubic polynomial, we construct a sequence of numbers N_m. The Langlands program predicts that the Fourier transform of this sequence has fractal symmetries.

However, if we want to investigate this application in historical settings, instead of the Langlands program we could have used its two precursors. These precursors became known about half a century before the Langlands program. While they are not as powerful as the actual Langlands program, in our application *all* easiest-to-reach fruits may be obtained using just the precursors.

The newer of two precursors was finalized about 90 years ago: the Class Field Theory. In general, it works by "splitting the complexity of a given polynomial P" into two parts: recall that solving $P = 0$ may be "relatively uncomplicated"[215] if we already know roots of a certain other, much simpler polynomial P_0 (in other words: P is "cyclic", or, at least, "abelian" *relative* to P_0). If $\deg P = 3$, then P_0 has $\sqrt{D}$ as a root; here D is the discriminant of P. The Class Field Theory converts many delicate questions about solutions of $P = 0$ to (rather involved) questions about P_0.

Two most useful (and most completely investigated) cases when the latter questions may be fully answered are when $\deg P_0 = 1$ (for $\deg P = 3$ this means that D is a complete square[216]), and when $\deg P_0 = 2$ and it has no real roots (for $\deg P = 3$ this means that $D < 0$). In the former case one gets a complete description of numbers N_m very similar to what we saw with Quadratic Reciprocity: there is a periodic sequence N_m^{per} such that $N_p = N_p^{\mathrm{per}}$ for prime p. (Compare with Remark 78 on p. 130.) The latter case is what we called the "even" (or "modular form") case in Remark 18 on p. 36. In this case the description of numbers N_m is less direct, but it is nevertheless sufficient to deduce all the fractality properties we use in these notes.[217]

Conclusion: to expose cases which are *not covered* by the Class Field Theory, our cubic polynomial should have $D > 0$ which is not a complete square. This is the "even" (or: the "Maass form") case.

---

[214] A very observant reader would note that with the formula above, $\mathrm{Rest}(t)$ would be very singular at 0. To avoid this singularity, we cheated and shifted the argument t in the graph by 2π.

[215] How to do this was discovered about two centuries ago.

[216] We discuss this case in a lot of details in the section on p. 67 and the section on p. 125.

[217] Apparently, the first example (in our terms, $M = 6$) was investigated by van der (den?) Blij in 1952. He (essentially) identified $F_{\mathbb{C}}(2\pi z)$ with $\eta(z)\eta(23z) = q\prod_{m=1}^{\infty}(1-q^m)(1-q^{23m})$; here $q = \exp 2\pi i z$ and we use the Dedekind modular form η.

However, he did not mention the (known) connections of his approach with polynomials of degree 3 (it looks like he, essentially, uncovered a very simple particular case of the result of Hecke of 1927). In an example in his 1975 lectures in Durham, Serre stresses this connection (and says that most of his examples came from Tate's letters of 1973/74 — but probably not this one...). Don Zagier's chapter in the book The 1-2-3 of Modular Forms exposes these connections directly.

The other precursor is the Hecke's functional equation for the Dedekind ζ-function discovered a century ago. (While it is usually not stated this way, in our setup) it claims exactly that our fractality law *works at* $t = 0$.

Recall that we already investigated what happens if the fractality law works at $t = 0$: by chaining our fractal transformation with periodicity, one obtains a giant "Cantor hyper-family" of other points at which the fractality law works as well (see Remarks 30 and 31, as well as Remark 51 on p. 78). Since this hyper-family avoids a lot of intervals, and we expect that horizon-self-similar points "appear everywhere", it should not be hard to pick up a horizon-self-similar point which cannot be explained by such chaining.

However, we have been working under a severe constraint: the zooming factor needed to expose the "fractal pattern" should not be prohibitively large (we do not want to spend weeks computing these graphs!). It turns out that many of the simplest points with "reasonable" zoom factors are in the hyper-family! This leads to the situation when most of our graphs can be explained by the Hecke's result.

Conclusion: to expose cases which are *not covered* by the functional equation, our graphs should show the fractal pattern about a point $t = t_0$ which is not in the "Cantor hyper-family". However, of the graphs in these notes, the only graphs not related to the hyper-family are one on p. 50 for $D = -23 < 0$, and one on p. 79 for $D = 2^4 \times 37 > 0$.

Combining two restrictions above, we need to provide a graph for the Maass case (so $D > 0$ and not a square) at a point which cannot be obtained from 0 by a chain of integer translations in T and applying $T \mapsto {}^1\!/_{cT}$ (here $t = 2\pi T$). The only graph which satisfies both restrictions is one on p. 79 with $c = 37$. (Compare with Remark 52 on p. 80.)

(Another educating facet of this paper is that the sequence he works with is a "mix" of our N_m with a Fourier coefficients of a certain Eisenstein series — compare with Remark 76 on p. 129. So this gives a very different example of a need to "distill" to see the patterns.)

On Lobachevsky geometry and zones of self-similarity

The groups of symmetries

In the preceding chapters we used a relatively new (about 25 years old) approach where we consider the Fourier transform $F(t)$ as a generalized function, and plot its antiderivative $F^{(-1)}(t)$. Note that "taking the antiderivative" is a regularization in the sense of Remark 21 on p. 37 — however, it is a very "mild" regularization: it replaces the sequence N_n (this is a sequence of integers, hence not decaying!) by a sequence N_n/n which decays, but rather slowly.

However, in this chapter we are going to ignore this approach (and $F^{(-1)}(t)$) until p. 90. Instead, we start by introducing geometric methods suggested by the other, older approach. That approach applies a very strong "regularization" making the Fourier transform *much* smoother. Such a regularization was described in Remark 21 on p. 37. One of the disadvantages of this approach is that one needs to use *different* regularizations in the odd and the even cases (introduced in Remark 18 on p. 36) — so with the older approach the difference between these two cases appears much earlier than necessary. (The exact form of these regularizations was described in Remark 35 on p. 58.)

On the other hand, using these particular regularizations has amazing corollaries. Indeed, they depend on the parameter s ("strength", which for $s \in \mathbb{N}$ may be thought of as a "repetion count"). For example, in the "odd" case the regularization replaces N_n by N_n/e^{ns}; now the Fourier transform of N_n/e^{ns} is a function of two variables t and s with $s > 0$. Writing $t + \mathrm{i}s =: z$ converts the Fourier transform of (N_n) into a function $f(z)$ defined on the upper half-plane $\mathfrak{H} := \{\operatorname{Im} z > 0\}$.

It turns out that if one considers the complex Fourier transform (as in $F_{\mathbb{C}}$) then

- The function $f(z)$ is complex-analytic.[218] The "boundary trace" of this function is $F_{\mathbb{C}}(t)$.[219]
- Every transformation we saw preserving the function $F(t)$ would preserve $f(z)$ too — when we write z instead of t in the formula for the transformation.

Moreover, in Remark 22 on p. 38 we claimed that (with Lobachevsky geometry!)

these transformations of z become just "rotations" if one equips $\mathfrak{H}$ with a certain curved geometry.

Essentially, the conditions on $F(t)$: periodicity and horizon-similarity (the latter makes a match between the "behavior on horizon" and the "behavior at finite points t") become to *symmetries* of $f(z)$ *in Lobachevsky geometry*. A geometric description of these symmetries allows us to *detach* the properties of these symmetries from the properties of $F(t)$ and $f(t, s)$.

So, in this chapter, we inspect these symmetries as "separate entities". Then we use the results to deduce a much more detailed information about regions of self-similarity for $F^{(-1)}(t)$.

Remark 54: In this approach, all the arithmetic information about the polynomial of degree 3 we started with boils down to one integer c: the conductor. Recall that conductors for cubic polynomials have a tendency to be very large, leading to hard-to-visualize situations. However, in the context of symmetries, small conductors c make perfect sense — and lead to much nicer pictures.

So while c in our pictures is too small to be related to *any* polynomial, these pictures still illustrate the general trends on manageable examples with very small conductors.

Remark 55: Already in Remark 18 on p. 36 we saw that the behaviour of horizon-self-similarity may be different in the odd and the even case (even if the conductor is the same[220]). The even case would

[218] This should be replaced with real-analyticity in the "even" case.

[219] Recall that $F_{\mathbb{C}}(t)$ has "no values at points". The "boundary trace" coincides with the boundary value *when values at points are defined* — but the trace makes sense for generalized functions as well.

[220] The smallest conductor for which both even and odd cases are possible is 756 with the corresponding "even" and "odd" polynomials $m^3 - 6m - 2$, $m^3 - 6m - 12$.

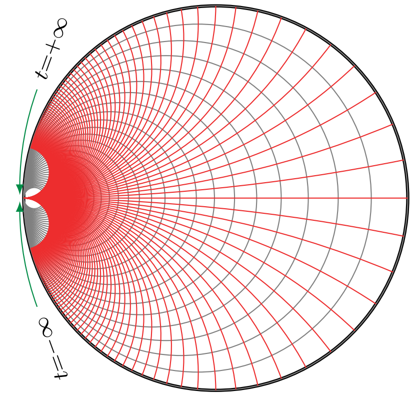
have more regions with such self-similarity (for example, a region near $t = 0$); in a certain sense, there are twice as many of them. Likewise, there are also twice as many symmetries of $f(t, s)$.

To simplify our pictures as much as possible, given c we start with the smallest "reasonable" collection of symmetries (which "works" for both even and odd cases), and postpone more complicated cases until the section on p. 93.

Lobachevsky-symmetries: the case $c = 1$

The easiest way to deal with a collection of symmetries is to find a picture such that these symmetries coincide with the symmetries of the picture. A particular case is when the picture consists of cut line which tessellate ("tile") the plane into pieces of the same shape and the same size. Moreover, when the collection consist of symmetries of a function $f(z)$, then if we know its values in one of these pieces, then we know its values everywhere:

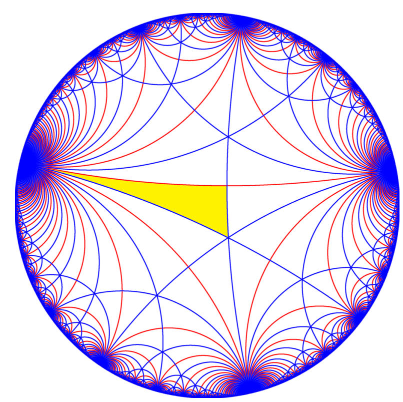
$$\boxed{\text{A Lobachevsky-symmetry which sends one piece to the other preserves } f(z).}$$



Of course, the same holds for the boundary trace $F(t)$ of $f(z)$. **Conclusion:** given such a coloring, one can discuss symmetries of $f(z)$ (and of $F(t)$) *without mentioning* f whatsoever. *This* is what we are going to do: *after* we describe the colorings, we won't need to mention $f(z)$ anymore. We would just apply the symmetries of the colorings to describe the symmetries of $F(t)$.[221]

However, it turns out that to simplify visualization of these examples, it is convenient to be creative with the *interpretation* of t and s.

While the function f from the preceding section takes arguments (t, s) in the upper half-plane $\{(t, s) \mid s > 0\}$ — which can be naturally identified with "the half-plane model" of the Lobachevsky plane, it is much easier to visualize the Lobachevsky moves using the "other flat-geometry model" of the Lobachevsky plane: the model inside a disk. (Geometrically, these two models — half-plane and disk — differ by inversion.)[222] In this model t and s become curvilinear coordinates in the disk; we show several coordinate lines on the right (t is in red, s is in gray).



Start with the simplest tesselation of Lobachevsky geometry (on the right; on Wikipedia, it is in the article Truncated triapeirogonal tiling[223] together with a few other examples, some of which are for small conductors). Every piece of tesselations we consider is made of several copies of "an elementary tile". This tile (in yellow on the right)[224] is marked as "index 1" in the Wikipedia article above. How the piece is made of these elementary tiles

[221] Essentially, the purpose of introducing $f(z)$ was to *lead us to* the Lobachevsky geometry. The interpretation of $F(t)$ as a boundary trace of something "as symmetric as" $F(t)$ is sufficient for our purposes: we do not care about finer details of $f(z)$.

[222] In our context, the major advantage of the disk model is that our toy/actual transforms have $-1/t$ as the argument; this means they, essentially, exchange points $t = 0$ and $t = \infty$. In the disk model, *both* $t = 0$ and $t = \infty$ make perfect sense as points on the boundary of the model. Compare with the picture in Remark 23 on p. 38.

In fact, if the point i of the half-plane $\mathfrak{H}$ matches the center of the disk, then the transformation above becomes just the rotation of the disk by 180°. (By the way, this is the main reason why we prefer writing $-1/t$ in the argument — as opposed to just $1/t$ — which would lead to a mirror symmetry of the Lobachevsky plane. See also Footnote 199 on p. 78.)

[223] I do not know anybody using such bizarre names in real life, or in real math.

[224] Note that in Lobachevsky geometry it makes sense "to pull a vertex of a triangle to infinity". When we pull, the angle at this vertex goes to 0°. The yellow piece is such a triangle with angles 90°, 60°, and 0°.

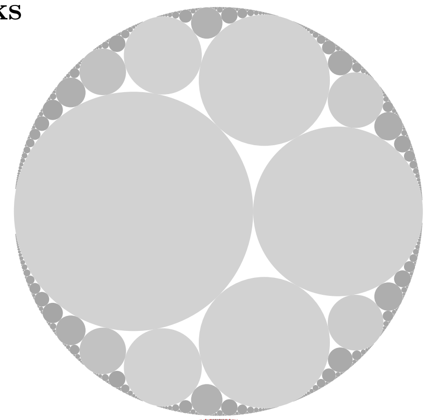
depends on the conductor only.[225] The yellow tile "combined" with any one of its neighbor tiles forms the piece good for[226] $c = 1$.

Remark 56: *On the picture,* the elementary tiles have different shapes and different sizes. Actually, in the sense of Lobachevsky geometry, these tiles have the same shape and the same size.[227]

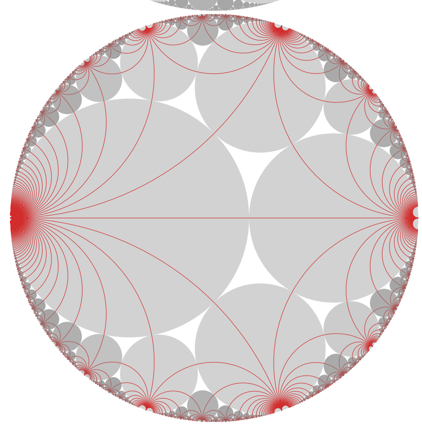
The observed difference is just a defect of our "visualization" of the Lobachevsky plane. Similarly to how the surface of Earth cannot be mapped exactly onto a flat surface, the features of Lobachevsky geometry cannot be rendered without defects on flat images.

Conclusion: What is drawn above is *just a hint* of what is going on in Lobachevsky geometry. In fact, it takes a lot of training to be able to interpret these hints fully! Below, the reader may need a long leap of faith with our recurring claims like "that picture demonstrates this symmetry".[228]

Enhance the picture: the gray disks



For what follows, it is convenient to add extra "features" to the coloring above (made of the "cut lines" of the tesselation). Note that on the picture above every red line meets one blue line; this marks a point on every red line. Look at these meeting points for the red lines which "emerge" from a given point of the boundary of the disk ("the absolute") — they all lie on a particular circle tangent to the absolute. In fact, these circles are the circles from so-called *Apollonian gasket* which touch the boundary (on the right, we shade the insides of these circles gray[229]).



By construction, any (Lobachevsky) symmetry of the picture above is also a (Lobachevsky) symmetry of the white/gray coloring on the right. Moreover, the opposite is also true.[230] **Conclusion:** two pictures above have the same symmetries; moreover, if one combines these two pictures, the result still has the same symmetries.[231] (In the combined picture on the right, we keep only the red lines from the preceding picture of the cut lines.)

This picture fits $c = 1$. For larger c, the group of symmetries is going to be a subgroup of the group of symmetries of this picture. This leads to this picture being *a template* for the pictures for larger c. We would need to omit some of the gray disks, and modify the

[225] Keep in mind that for a large conductor c, one may need about $4c$ elementary tiles to make the shape needed above. Since conductors have a tendency to be quite large, most examples would lead to shapes made of monstrously huge number of tiles.

[226] Without doubling the yellow tile is a "piece" if we allow mirror symmetries. Compare with Footnote 271 on p. 94.

[227] In particular, there is a (unique!) Lobachevsky-symmetry of the picture above sending any "elementary tile" to any other tile.

[228] There are videos visualizing geometry and movements of the Lobachevsky plane. Google for `movement OR visualizing hyperbolic demo OR projections video`.

[229] We use darker gray for smaller disks to make them easier to see. This tint has no math significance.

[230] Indeed, one can reconstruct the red lines on the gasket: take two tangent gray disks, and connect the points where they touch the absolute. Likewise, any blue line is a common Lobachevsky-straight tangent to such a pair of disks.

[231] In fact, the disks make it easy to describe these symmetries. One can find a Lobachevsky-rotation sending any disk to any other disk. Moreover, note that the disks touching a given disk make a "necklace" surrounding the disk. Now given a disk, there is a unique Lobachevsky-rotation which keeps this disk in place, and sends a particular disk in this necklace to another such disk. (Finally, there is a unique reflection keeping two touching disks in place.)

(Compare with Footnote 227.)

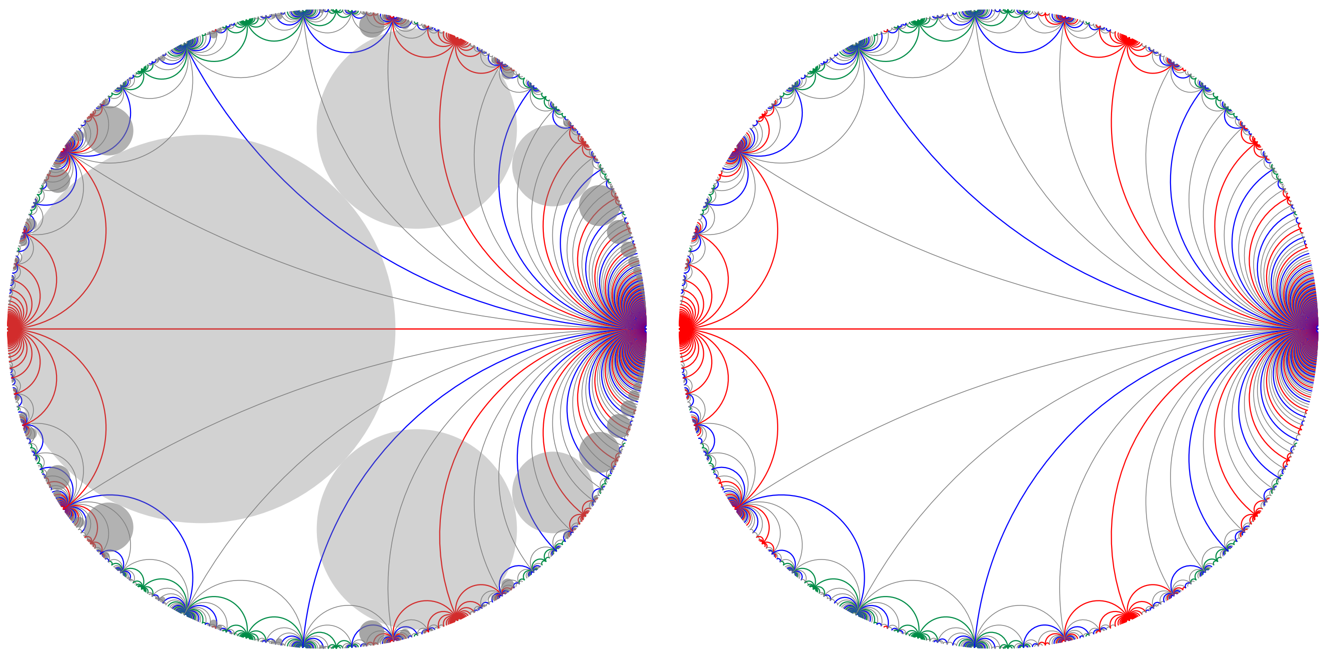
colored lines. (Additionally, we would need to rescale the coordinate t on the disk — and its boundary.)

The case $c = 5$

For larger conductors c one can draw pictures which are very similar in spirit. Here we consider the case of $c = 5$.[232] What one needs to do is:

- Remove some of the gray disks and the red lines;
- Change some red lines into green;
- Add suitable gray and blue lines.

(After this, there is still a lot of tangencies between the gray disks!) This gives the picture on the left (on the right, we remove the gray disks to make the colored lines easier to see):



It is easy to imagine yet another picture with gray disks only, and no lines. All three ways to color (gray disks only, and two colorings above) have the same collection of symmetries. Moreover:

- The gray and colored[233] lines on the right picture match the red lines on the picture on p. 85.[234]
- These lines cut the picture into "triangles". Every such triangle matches a red-sided triangle which on the picture on p. 85 is made out of 6 "yellow elementary tiles".
- For any two of these triangles, there is a Lobachevsky-rotation or Lobachevsky-translation sending one to the other (one can even send a given corner to a given corner). In other words, in Lobachevsky geometry these triangles have the same shape and the same size.[235]
- Ignore the gray lines. Then the colored lines cut the picture into 6-sided pieces having 2 red sides, 2 green and 2 blue. Each piece is made of 4 triangles.
- These larger pieces are also the same shape and the same size (in Lobachevsky sense).

[232] As we discuss it in the section on p. 93, there are several different analogues. Until then, it is enough to say that here we consider the "smallest useful" collection of symmetries.

[233] Here and below "colored lines" means "non-gray" lines.

[234] For this and other matches below, it is better to Lobachevsky-move the picture on p. 85 ("squeeze it to the left").

[235] Well, the Lobachevsky geometry is not scaling-invariant: if shapes match, the size *should also* match!

- Moreover, these pieces are *fundamental domains*: they have no symmetries,[236] and any symmetry of the whole picture moves a piece to a piece.
- Therefore, these pieces are closely related to the gray disks. For example, every one contains exactly one tangency point of the gray disks.[237]

  Moreover, the gray disks color every piece with 2 colors: gray and white. One can Lobachevsky-overlay any two pieces so that they match, moreover, the colors match as well.[238]

Remark 57: For any gray disk, the unit "necklace rotation" of Footnote 231 on p. 86 repeated c times is a symmetry of the whole picture. (This shows that 2 of every 5 consecutive "beads" in such a necklace for $c = 1$ remain in the picture for $c = 5$.)[239]

The gray disks and the "special zones"

Above, we constructed a coloring of the Lobachevsky plane such that its symmetries coincide with the symmetries[240] of the (generalized) function $F(t)$ on the absolute and of the function $f(t, s)$ on the Lobachevsky plane.[241] [242] Here we reap the fruits, using the symmetries of the picture of gray disks to inspect the fractal transforms[243] which preserve $F(t)$.

Since there is a symmetry of the picture moving any gray disk to any other gray disk, and these symmetries preserve $F(t)$:

The function $F(t)$ "behaves the same" near any two points where a gray disk touches the absolute.

Following Remark 23 on p. 38 we identify the leftmost point of the absolute with $t = \infty$ (as on the picture on p. 85). Recall that the absolute is essentially the t-axis on which a periodic function $F(t)$ is defined, and the behaviour of a periodic function "near $t = \infty$" is its behaviour "near horizon"— which is what matters for our fractal transforms. This immediately implies:

The points where a gray disk touches the absolute are horizon-self-similar points of $F(t)$.

(The horizon-self-similar zones are as on the graph on p. 55.)

[236] Indeed, there is a Lobachevsky-reflection of a piece which preserves the coloring of its edges, — but it does not preserve the gray diagonals drawn in the piece. (This implies that there are no symmetries of the picture which are Lobachevsky-reflections; compare with Footnote 271 on p. 94.) Anyway, in this chapter we ignore reflections!

[237] Moreover, the corner of a "piece" where two red sides meet is contained inside a gray disk. Likewise, the other 5 corners are contained in the *disks which were removed* when changing the picture for $c = 1$ to one for $c = 5$; hence these corners do not meet *the remaining* gray disks. Hence every "piece" meets only two gray disks, and two green sides of the piece completely avoid the gray disks.

[238] The same holds if we also take into account the coloring of the edges of the pieces.

[239] **Warning:** this match of the disks does not extend to a match of triangular tiles and/or the coordinate t on the absolute. As c grows, the "triangles" above the horizontal diameter become squeezed closer and closer to this diameter, and the range of t covering "the right half" of the absolute decreases (approximately as $[-3\pi/c, 3\pi/c]$).

[240] Here it helps to interpret the functions F, $F^{(-1)}$, and f as tensor-valued, as in Footnote 378 on p. 130. Then the toy/actual transforms become just coordinate-changes applied to tensor fields, and horizon-self-similarity may be interpreted as "being symmetrical".

[241] Recall that the key reason why F and f have the same symmetries is that our constructions of "continuation into the plane" and of "taking the boundary value" were *intertwining*: Lobachevsky-moving one of them would Lobachevky-move the other in exactly the same way. See Remark 23 on p. 38.

[242] The function $f(t, s)$ can also be considered as a coloring of the Lobachevsky plane: its value at (t, s), which is a real number, may be considered as a color assigned to this point. So the idea of gray disks is that we can replace this infinity of colors with only 2 colors!

(Well, to take into account that f is a tensor field, one can consider $|f|$ as a color. Otherwise f colors not the Lobachevsky plane, but its tangent bundle.)

[243] See p. 34.

As we explained above,[244] Euclidean-rotations of our pictures of the Lobachevsky plane are also Lobachevsky-rotations, but there are many more Lobachevsky moves. They lead to more complicated "skewed rotations" of the absolute ("fractional-linear transforms", see p. 78). The non-linearity of these transforms may shrink some parts of the absolute and expand the others. Hence if a graph of a function has a visible pattern, such a "non-linear" coordinate transform may distort this pattern—although it would remain visible in a smaller region, where the non-linearity is not "too strong".

Therefore, if this transform is not "too non-linear" near one tangency point, then

It sends the zone of "visual horizon-self-similarity" near this point to another such zone.

Doing a similar thing with the leftmost point $t = \infty$ of the absolute gives:

The zones of "visual horizon-self-similarity" are transforms of a certain region near $t = \infty$.

Conclusion: If we can identify this region, then the zones above are images of this region under Lobachevsky-symmetries of the picture with the disks!

Loosely speaking (and there is no other way to discuss this, since "the zones of *visual* horizon-self-similarity" depend on our *visual shape-recognition*[245]), use as "the unit of measure" "the projection of the gray disk near $t = \infty$ to the absolute".[246] This leads to a reformulation:

The zones above are certain central parts of the projections of gray disks to the absolute.

(... *except* for the zone near $t = \infty$ itself: then the transform is not a fractal-transform, but identity!)
The answer: The region in question is the central $4/c$ of the projection of the leftmost gray disk (recall that $c = 5$ in the example above). Call the corresponding zones inside the projections of other gray disks *the $4/c$-central zones*.[247]

To understand why this recipe works, we need to

- visually inspect the zone of "visual horizon-self-similarity" for a toy transform of a sample periodic function,
- identify the matching range near $t = \infty$;
- Find which part of the absolute on the pictures above matches this range of $t \approx \infty$.

We will address the last item in the next remark, and the first two in the remark which follows it.

Remark 58: Following Remark 23 on p. 38, on the pictures above the leftmost point is "the infinity" of the absolute, and the rightmost point is $t = 0$. Moreover, any Lobachevsky-rotation which exchanges these two points is $t \mapsto -1/\gamma t$ on the absolute, for a certain $\gamma > 0$ (the "toy transform"!).

Now observe the "pieces" next to the leftmost point $t = \infty$; as we described it above, one can Lobachevsky-rotate one of them to overlay it on top of its counter-clockwise neighbor. Moreover, by Footnote 238 on p. 88 the edge colors must match. Observing the red edges near $t = \infty$ shows that this move ρ sends every red line starting at the leftmost point to its counter-clockwise neighbor.

---

[244] See Footnote 222 on p. 85.

[245] Indeed, "mathematically" the fractal transform is defined *everywhere*. However, it is not everywhere "visually recognizable": the non-linearities hide the similarities.

[246] Making this rigorous requires choosing the center of projection. However, there is no "best" way to do this. Different choices would result in "slightly different" regions—but for us just the *approximate* size of regions is important.

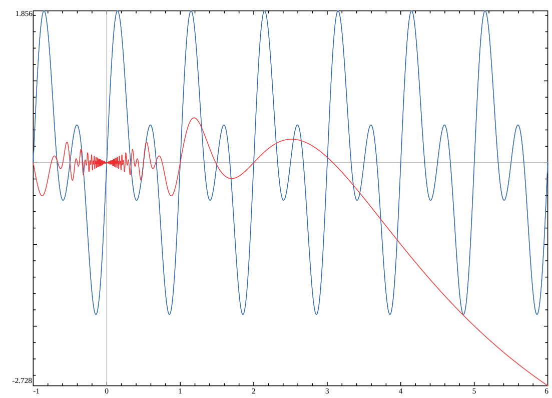
[247] In fact, we could have replaced every gray disk by a $c/4$ times smaller disk ("the $4/c$-disk"), and then all the properties of our coloring discussed above would be still preserved, *and* the projections of the $4/c$-disks would match exactly the sizes of the regions of visual horizon-self-similarity. However, then (even with our small conductor $c = 5$) the disks would be yet harder to see clearly. Moreover, the facts that the original disks are tangent to each other, and that they match the case $c = 1$ (so that on our pictures just the *quantity* of the disks depends on c, not their sizes and placement) are sufficiently interesting for us to prefer the picture with larger disks.

Later (in the section on p. 96), when we work with harder-to-understand pictures, we are going to have *both* the "original" and the $4/c$-disk marked on the picture.

Recall that the whole idea of the coloring above, on p. 87, is that its symmetries are also symmetries of $f(z)$ (hence, automatically, of its "boundary trace" $F(t)$). This immediately identifies the action of ρ on the absolute with the translation of t by the period 2π of $F(t)$. **Conclusion:** the ends of the red lines starting at $t = \infty$ are at $t = 2\pi k$ with integer k.[248]

For general c the picture would still contain a "necklace" accumulating at the rightmost point $t = 0$. The k steps of the corresponding "necklace rotation" of Footnote 231 on p. 86 can be recognized as the strongly-congruence transform $T' = {}^{T}\!/_{kcT+1}$. Hence the disks of this necklace touch the absolute at points $T = {}^{1}\!/_{kc}$, or $t = {}^{2\pi}\!/_{kc}$.

So two disks in this necklace next to the leftmost one are at $t = \pm{}^{2\pi}\!/_{c}$, and the edges of the projection of the leftmost disk are twice this, at $t = \pm{}^{4\pi}\!/_{c}$. This shows that the regions between $t = \infty$ and points $t = \pm\pi$ (from the next remark) take ${}^{4}\!/_{c}$ of this projection — leading to the answer above.



Remark 59: To quantify the effects of non-linearity of $-{}^{1}\!/_{T}$, on the right we consider a typical example of a function[249] $\Phi(T)$ with period 1, and graph $\Phi(T)$ and ${}^{1}\!/_{4}T\Phi(1/T)$ for T in $[-1, 6]$. One can immediately see that for $|T| > 2$ the red plot "does not look as" following its pattern clearly visible for $|T| < 1$ (although "mathematically", it is "the same" pattern).[250] In other words, the visible pattern "exists" in $[-2, 2]$ (the "narrow flavor"), or maybe even up to $[-4, 4]$ in the "wider flavor" which stresses our imagination.

To make our description work equally well with transformations $T = {}^{1}\!/_{cT'}$ with different cs, one can rewrite the estimate we obtained for $c = 1$ in terms of T'. This is $|T'| > {}^{1}\!/_{2}$ for the "narrow flavor" (or $|T'| > {}^{1}\!/_{4}$ for the "wider flavor"). One can restate this as

$$\boxed{\text{The pattern on the graph of } T\Phi({}^{1}\!/_{cT}) \text{ is visible when } T = {}^{1}\!/_{cT'} \text{ with } T' > {}^{1}\!/_{2}.}$$

Note that $T' > {}^{1}\!/_{2}$ means that we remove one period of Φ around 0.

Conclusion: "the pattern is visually recognizable" on the image of all the periods of $\Phi(T)$ except one[251] period around 0.

---

248 Doing similar arguments at the rightmost point $t = 0$ shows that ends of all lines starting at $t = 0$ are at $t = {}^{2\pi}\!/_{k}$ with integer k. (Moreover, for the piece immediately above the horizontal diameter and bounded by the colored lines, one can find that its corners are at t being 0, π, ${}^{6\pi}\!/_{5}$, ${}^{4\pi}\!/_{3}$, 2π, ∞. For the piece to the right of it, the values are 0, ${}^{2\pi}\!/_{5}$, ${}^{\pi}\!/_{2}$, ${}^{2\pi}\!/_{3}$, ${}^{4\pi}\!/_{5}$, π.)

249 We use the "same" letter T for the variable as before, when we had $t = 2\pi T$ and 2π-periodic functions of t.

250 The situation does not improve for $|T| > 6$, where $T\Phi(1/T)$ quickly converges to a certain limit.

251 This gives just an estimate "of the order of magnitude" of the zone to delete. Moreover, this estimate is sensitive to the shape of the graph of $\Phi(t)$.[252]

However, in practice, we need not the toy transform, but the "honest law for antiderivative", see p. 49. It turns out that the extra term in this law already messes up (a little) what happens near the edges of "the narrow flavor" of this zone, and its contribution breaks up the visual pattern in the "wider flavor". (We already mentioned this in Remark 34 on p. 57.) So to get a recognizable pattern, the narrow flavor (or maybe even it is a bit more narrow) could be a better fit.

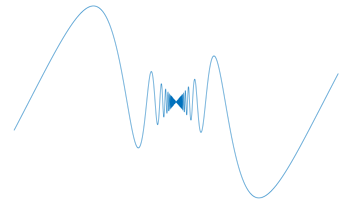
Compare with what happens near the rightmost point $t = 0$ on the pictures above (on p. 87). There is a "necklace" of gray disks converging to $t = 0$; their projections fill the whole neighborhood of this point. Taking ${}^{4}\!/_{c}$-central zones gives zones "converging to 0" with gaps of relative width about $c - 4 : 4$ (or maybe $c - 8 : 8$). Now observe the plot on p. 54 (for $M = 6$, $c = 23$): this is exactly what happens there! (Likewise for plots near $t = 0$ for other values of M. — Unfortunately, these zones become invisibly narrow if the conductor is in the hundreds!)

252 … as we have seen in the section on p. 55, with $\Phi(T)$ having "extra" symmetries.

Covering properties of the zones of horizon-self-similarity

Up to now, we were somewhat vague about visual patterns in $F^{(-1)}(t)$, avoiding the question:

Given t_0, can we zoom into the graph of $F^{(-1)}(t)$ and see $t = t_0$ in a region of horizon-similarity?



Recall what the preceding section started with a picture with gray disks and established:

- The periodicity[253] of $F^{(-1)}(t)$ "observed near 'the horizon point' $t = \infty$ on the absolute" makes "the hourglass" pattern (as on the right);
- We can consider this pattern as "a template" in a certain zone near the tangency point $t = \infty$ of the corresponding gray disk.
- Lobachevsky-symmetries of the picture "distribute" this template near every other gray disk.
- On the *other* disks (not at infinity) "the hourglass" patterns become the toy transform patterns.

Hence every gray disk leads to:

- A special point on the absolute (the tangency point).
- A special (although "approximately defined") region about this special point (the $^4/_c$-central zone).

Conclusion: The special point shows *where* we should zoom in, and the zone shows *how much* to scale to see the zones of horizon-*self*-similarity.[254]

Now the question above can be reformulated as:
Question: which part of the absolute is covered by the $^4/_c$-central zones?

It turns out that while the claim

Every small piece of the graph of $F^{(-1)}(t)$ looks like a fractal transform of the whole graph.

does not "work 100%", one needs[255] to allow just a tiny amount of exceptions.[256]

In terms of the gray disks, this means that the projections of these disks to the absolute should cover "almost" the whole absolute. (Moreover, they would overlap strongly enough so that the $^4/_c$-central zones would also cover it "almost completely".) Contrary to this, on the picture above with the gray disks for $c = 5$, one can clearly see big regions near the absolute where there is no gray disks — even if one tries to zoom into the picture (this is possible in the electronic copy).

Indeed, on the picture above on p. 87, note the "worst offenders": the points of the absolute where the green lines join together. (Below, we focus on one of them, a bit left of the top point, matching $t = ^{4\pi}/_5$.) Near such points there are no gray disks drawn!

However, it is just an artifact of computer plotting. It is not possible to create a PDF graph into which one can zoom forever; but if it were possible, and one was patient enough to zoom *deep enough*,

[253] … together with $F^{(-1)}(t)$ being actually a tensor field! (See Footnote 240 on p. 88.)
(The difference between $t = \infty$ and $t \neq \infty$ is due to extreme non-linearity of the coordinate t near $t = \infty$.)

[254] Furthermore, note that to simplify our pictures, so far they were related to the "smallest possible" flavor of various groups of symmetries we may consider (compare with the section on p. 93). This means that we do not yet list *all* the possible zones. We complete this list later, in the section on p. 96. In the same section we also discuss horizon-similar but non-*self*-similar zones.

[255] To see that there are exceptions, take a gray disk which is in the picture for $c = 1$, but is removed in the picture for $c = 5$. It touches the absolute at a rational multiple of π (for example, $t = 0$ is such; for $c = 5$ another example is $t = ^{4\pi}/_5$), and (for example, on the picture with gray disks on p. 95) it is not hard to see that every nearby disk has much smaller diameter than its distance to this point t.

(When we add more symmetries later, in the section on p. 94, we will see that the point $t = ^{4\pi}/_5$ *is also* a point of horizon-self-similarity. However, $t = ^{2\pi}/_3$ is not, and a similar argument works there too.)

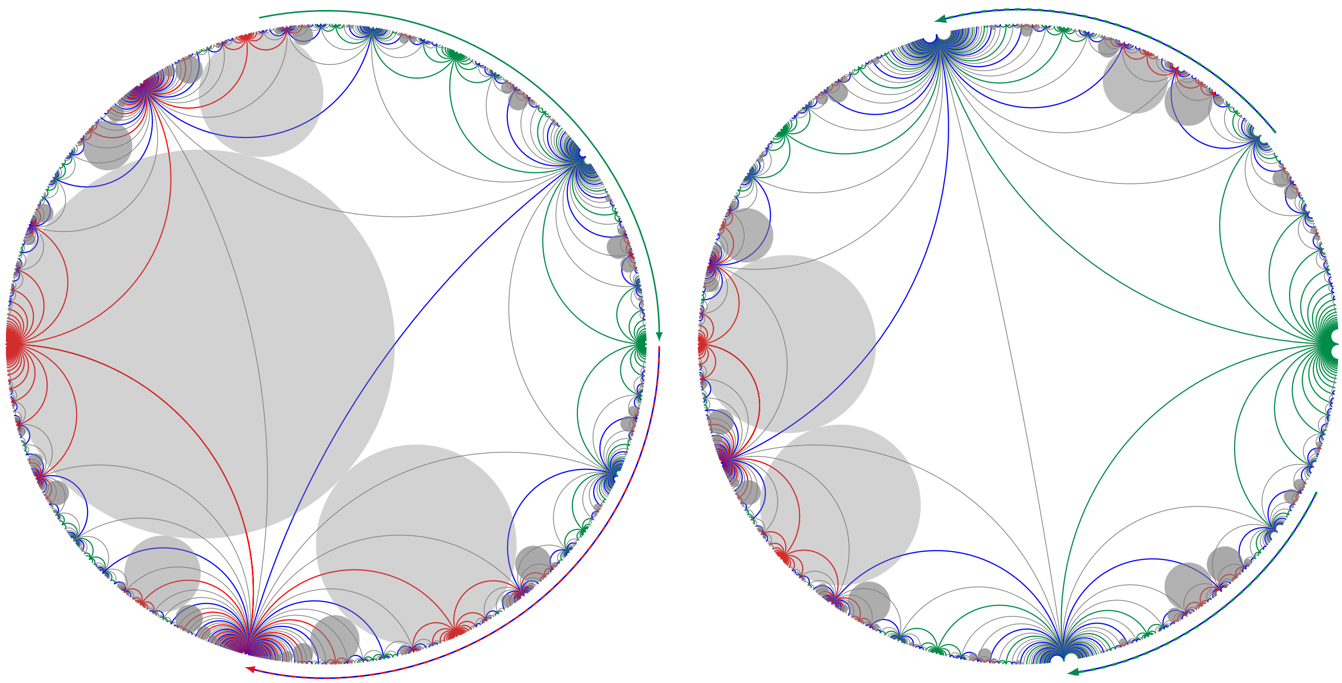
[256] The uncovered set is a "meagre" subset of measure 0.

one would see that the framed statement above holds. However, one would need to zoom scaling up hundreds or thousands times — even with the tiny conductor $c = 5$ we discuss here!

To substantiate the framed claim above, it helps if we can zoom near the "worst offender" point: where green lines join (on the left of the topmost point). Unfortunately, the more we zoom into our Euclidean picture so that this point is visible, the smaller is going to be the relative size of the disks in our field of view![257]

Fortunately for us, some Lobachevsky-*moves* look like *zooming* in pictures drawn in "our" geometry. For example, a Lobachevsky-translation along a Lobachevsky line looks like *zooming in* at the "tail" end of this line, and *zooming out* near the "head" end of this line. So a Lobachevsky-translation to the left along a horizontal diameter of our disk would zoom in at the rightmost point. Therefore, this way we can zoom near our point of interest while keeping the whole picture visible, and without breaking the "spirit of the picture".[258]

Before we can zoom this way near our "worst offender" point on the picture on p. 87 (where the green lines join), we must apply a Lobachevsky-rotation "about" the leftmost point of the absolute to make "the worst-offender" point into the rightmost point; the result is below on the left. (This already has a side effect of zooming in near the point of "green convergence". Note also that the point which *was* rightmost ends on the left of the bottom of the picture):



Finally we can make the horizontal Lobachevsky-translation to the left which we discussed above. This results in the right picture (and now we zoomed a lot into the rightmost point and have a much more clear picture of the green lines).

To see how the gray disks behave near what is now the rightmost point, note that the green lines cut the disk into "*slices*". Moreover, there is a Lobachevsky-symmetry of our picture which keeps the righmost point intact and sends a slice into the next slice clockwise. (This is similar to what we did in Remark 58 on p. 89.) **Conclusion:** all slices have the same shape and the same size (in Lobachevsky sense) and are "colored in the same way".

[257] As in Footnote 255.

[258] This happens because Lobachevsky-symmetries are conformal maps when considered in our, Euclid geometry. Near every point, such a map always looks like zooming and/or rotating. So unless it is linear, a conformal map would zoom into some points, and zoom out of some others.

In particular,

$$\boxed{\text{The gray disks in every slice are positioned the same way.}}$$

Moreover, the green lines also cut the *absolute* into chunks. Observing the largest slice shows that

$$\boxed{\text{In any chunk of the absolute, the projections of two largest gray disks to the boundary cover about } 1/3 \text{ of it.}}$$

Conclusion: although we cannot see tiny disks near the rightmost point, nevertheless even if we count just the two largest disks in every slice, together their projections cover about $1/3$ of the space near this point. (Indeed, this claim holds *in every chunk* of the absolute near this point, so it must also hold if we *join* the chunks together.)

In the original picture with the gray disks "before zooming", it is easy to see that the point we considered is the "worst" point with respect to having gray disks nearby. **Conclusion:** near *every* point, at least $1/3$ of the absolute is covered by the projections of gray disks — provided we include the tiny "invisible" disks as well.

Furthermore, it is easy to improve this estimate $1/3$ above. Indeed, to obtain the estimate $1/3$ we considered only the projections of *two largest* disks in a slice — but now we know that at least $1/3$ *of the rest* is also covered by projections of tiny disks. This means that at least $1/3 + 2/3 \times 1/3 = 5/9 > 1/2$ is covered by the projections. Repeating this argument again improves the estimate first to $1/2 + 1/2 \times 1/2 = 3/4$, then to $3/4 + 1/4 \times 3/4 = 15/16$ etc. Continuing like this, we can get as close to 1 as we want to — however, it is clear that to get close to 1, we *need* to consider incredibly small disks!

(Of course, a similar argument works if we consider just the $4/c$-central zones of each projection — only one would need more steps.)

More symmetries

In fact, out of several possible flavors of fractal symmetries the example above deals with the smallest one: in terms of Footnote 201 on p. 78 these symmetries are both the "strongly-congruence" type, and the "keeping sign" type (these types of "congruence" transforms coincide[259] for $c = 5$). Because these symmetries keep sign of $F(t)$,[260] in the zones considered above every oscillation of the graph of $F^{(-1)}(t)$ matches the shape of the period of the graph of $F^{(-1)}(t)$ *without flipping its sign.*[261]

One can also show[262] that in the "odd" case these zones exhaust *all* the "keeping sign" regions of horizon-self-similarity. (In the even case one needs to take into account Remark 61 on p. 94 too. We do it in the following section.)

However, if we do not mind the extra "minus" signs, we need to consider a larger group of symmetries. The spirit of the pictures above was that our symmetries send one gray disk to another; so if we want to switch to a larger collection of symmetries, we should increase the number of the gray disks likewise.

Conclusion: for $c = 5$, there is a similar picture with twice as many disks — and with this new picture the arguments above work as well. (Below, in the section on p. 94, we illustrate this by adding the red disks to the gray ones.) Hence there are twice as many $4/c$-central zones too, and in the "newly added" zones every oscillation of the graph of $F^{(-1)}(t)$ matches the shape the period of this function with the opposite sign.[263]

Moreover, the "sign-flipping" symmetries can be described geometrically, as symmetries of the right picture on p. 87 which exchange red and green lines. So the newly added disks are tangent to

[259] Compare with Footnote 266 on p. 94.

[260] In pedantic mode: ... *would "keep"* the sign — if F with such a tiny conductor existed.

[261] In terms of formula of Footnote 102 on p. 42, this means $\varepsilon > 0$.

[262] **N.B. (???) Check!!!**

[263] Compare with Footnote 206 on p. 80; a similar thing happens for $c = 5$ near $t = 4\pi/5$ — this time for *horizon-self*-similarity (as opposed to "similarity to what happens at $t = 0$", or to "horizon-similarity to $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$").

the absolute at the points where the green lines meet[264], and the newly added ${}^{4}\!/_{c}$-central zones are the central regions inside projections of these disks.[265]

Example: the point $t = {}^{4\pi}\!/_{5}$, where the green lines meet, is not covered by the projections of "the old" gray disks; however, there is a "sign-flipping symmetry" which sends $t = {}^{4\pi}\!/_{5}$ to $t = \infty$ (where the *red* lines meet). This shows that the function *is also horizon-self-similar* at $t = {}^{4\pi}\!/_{5}$ — but with the sign-inversion.

Remark 60: Theoretically, for large conductors one could investigate yet another picture (but we are not going to do it here!): the sign-keeping flavor of symmetries has a very natural strictly smaller sub-collection of "strongly congruence" transforms.[266] This leads to three different arrangements of disks: one for "only strong symmetries", one for all sign-keeping symmetries, and one for all "congruence" symmetries.[267]

Remark 61: To add insult to injury, on our graphs we saw still *other* zones of fractality, for example the zone near $t = 0$.[268] As we already mentioned, the corresponding transformation $t' = {}^{1}\!/_{\gamma t}$ is directly related to Hecke's functional equation (see the section on p. 82 for details).

Before, we connected the horizon-self-similarity in the zones we saw with existence of "good" moves of the Lobachevsky plane which send a neighborhood of $t = \infty$ to such a zone (here a "good" move preserves f and F). Likewise, this zone near $t = 0$ is also an image of a neighborhood of $t = \infty$, however this time the effect of this move $T' = -{}^{1}\!/_{cT}$ on f and F depends on the "parity": in the "odd" case it would multiply $F_{\mathbb{C}}$ by the imaginary unit i, in the "even" case it preserves F.

Obviously, the images of *this zone* under "good" moves would have exactly the same fractality pattern as the pattern in this zone. Hence these images are also horizon-similar!

Adding the $T' = -{}^{1}\!/_{cT}$ to any flavor of "congruence" symmetries doubles this class (one can consider the "old symmetries", as well as their "combinations with $T' = -{}^{1}\!/_{cT}$").[269] So this provides 3 more classes ("as above, but possibly combined with $T' = -{}^{1}\!/_{cT}$") to consider.[270]

We investigate the largest of these augmented types in the section on p. 96.[271]

Conclusion: we already illustrated the "sign-keeping" symmetries above. Below, we first add "sign-flipping" symmetries; then we double the class of symmetries once more by adding $T' = -{}^{1}\!/_{cT}$.

---

[264] . . . (but not the points where the green and the blue lines meet! The size of the disks is determined by them "filling the void" between the gray disks

[265] For this larger arrangement of disks, four out of any five consequent disks in a necklace would be included — as opposed to two-out-of-five of Footnote 247 on p. 89. Compare with the picture on p. 95 — where we color the added disks red.

[266] This does not happen for $c = 5$ since in this case $\{\pm 1 \bmod c\}$ includes all non-0 (or invertible) squares mod c. (Compare with Footnote 206 on p. 80. Such cs are divisors of $2^3 \times 3 \times 5 = 120$.)

[267] The symmetries of the first and third arrangements have names: $\Gamma_1(c)$ and $\Gamma_0(c)$.

[268] While we saw that the behaviour of $F^{(-1)}(t)$ in such zones is different in "even" and "odd" cases (see Remark 18 on p. 36), the geometry of the *zones themselves* are the same. So here we treat these cases uniformly.

[269] The transform $T' = -{}^{1}\!/_{cT}$ is not a congruence transform (unless $c = 1$)! A possibility of "adding it" like we did above is due to its being a "*normalizer*" of the "old" group of symmetries.

[270] This is related to the fact that the suitable symmetries live in $\mathsf{PGL}_2\mathbb{Q}$ and not in $\mathsf{PSL}_2\mathbb{Z}$.

[271] Since $F(t)$ is even, it has another symmetry $t' = -t$. This leads to a mirror symmetry of the Lobachevsky plane — however, it does not add extra info about fractality properties.

To avoid proliferating our symmetries yet more, in this chapter we focus only on non-mirror (orientation-preserving) symmetries.

(On the pictures above for the case $c = 1$, allowing mirror symmetries leads to a very nice and useful kaleidoscope — "the yellow piece" fills the whole plane using only reflections in its sides. — However, I do not see any similar simplification for cases of higher c. So it looks like avoiding mirror symmetries has only positive effects. Compare this to our choice to consider only solutions to $\alpha\delta - \beta\gamma = 1 > 0$ in Footnote 201 on p. 78 — the reflections correspond to $\alpha\delta - \beta\gamma < 0$.)

Warning: if one consider $F_{\mathbb{C}}(t)$, this mirror symmetry changes its values by complex conjugation.

Adding "sign-flipping" zones

Considering larger groups of symmetries would lead to yet more disks in our pictures. If we continue as above, the pictures would become very crowded. Before we proceed, we need to modify our infrastructure.

First, we want to visualize $4/c$ times smaller gray disks (this follows Footnote 247 on p. 89) to take advantage of their projections to the absolute matching in size the horizon-similar zones. However, we do not want to abandon the convenient features of larger disks (see the same Footnote). So we are going to draw them both: a smaller disk inside a larger circle.

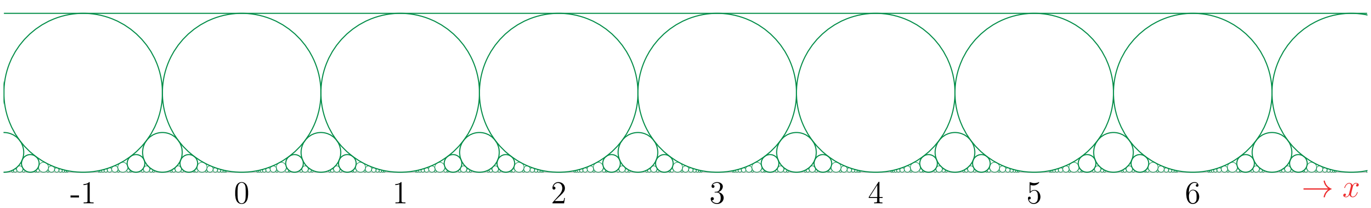
With this modification, the recipes we used before become:

- Take the "outside" circles in the Apollonian gasket (those touching the boundary).
- Introduce an appropriate coordinate on the boundary=absolute.
- Remove all the circles except those matching the "sign-preserving" horizon-self-similar zones.
- In the remaining circles, shade sub-disks of $c/4$ times smaller radius.

The projections of the shaded sub-disks to the absolute are approximations to the "visually" horizon-self-similar zones.

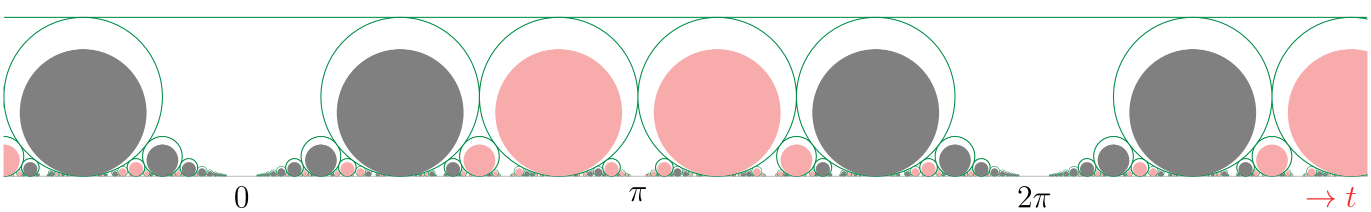
Second, we want to use a different model of the Lobachevsky plane. While the model in the disk used above simplifies visualization of Lobachevsky moves, the required book-keeping is too complicated.[272] The half-plane model of the Lobachevsky plane allows us to state a more explicit description of the pictures.

In this model the "outside" circles of the Apollonian gasket turn into the the Ford circles: the circles tangent to the boundary at points with rational coordinates $x = R/D$ with the diameter $1/D^2$:



(The Apollonian circle tangent to the absolute at $t = \infty$ is an exception; it becomes the horizontal line at height 1.)

For our purposes, the suitable coordinate on the absolute is $t = 2\pi x/c$. **From this moment on, we mark the horizontal axis with this rescaled coordinate.** With a prime c (below $c = 5$ again), it turns out that to get the correct picture of the gray disks, we must omit the disks with the numerator R of $t = 2\pi R/D$ divisible by c:



This still leaves twice as many disks — but it turns out that the "extra" disks (marked in red) are exactly what we wanted to add:

$$\boxed{\text{The red disks "match" the "sign-flipping" Lobachevsky-symmetries of } f(t,s).}$$

In particular, the "sign-keeping" symmetries send gray and red disks to the disks of the same color, while the "sign-flipping" symmetries exchange the colors. The projections of the gray disks are

[272] For example, we could not state explicitly which of the Apollonian circles are omitted on our pictures for $c = 5$.

the horizon-self-similar zones "keeping the signs", and the projections of the red ones are for the "sign-flipping" zones. (The color depends on $\left(\frac{R}{c}\right)$.)[273]

The new process may be summarized as:

$$\boxed{\text{Remove some Ford circles, and “inscribe” smaller disks in the remaining circles.}}$$

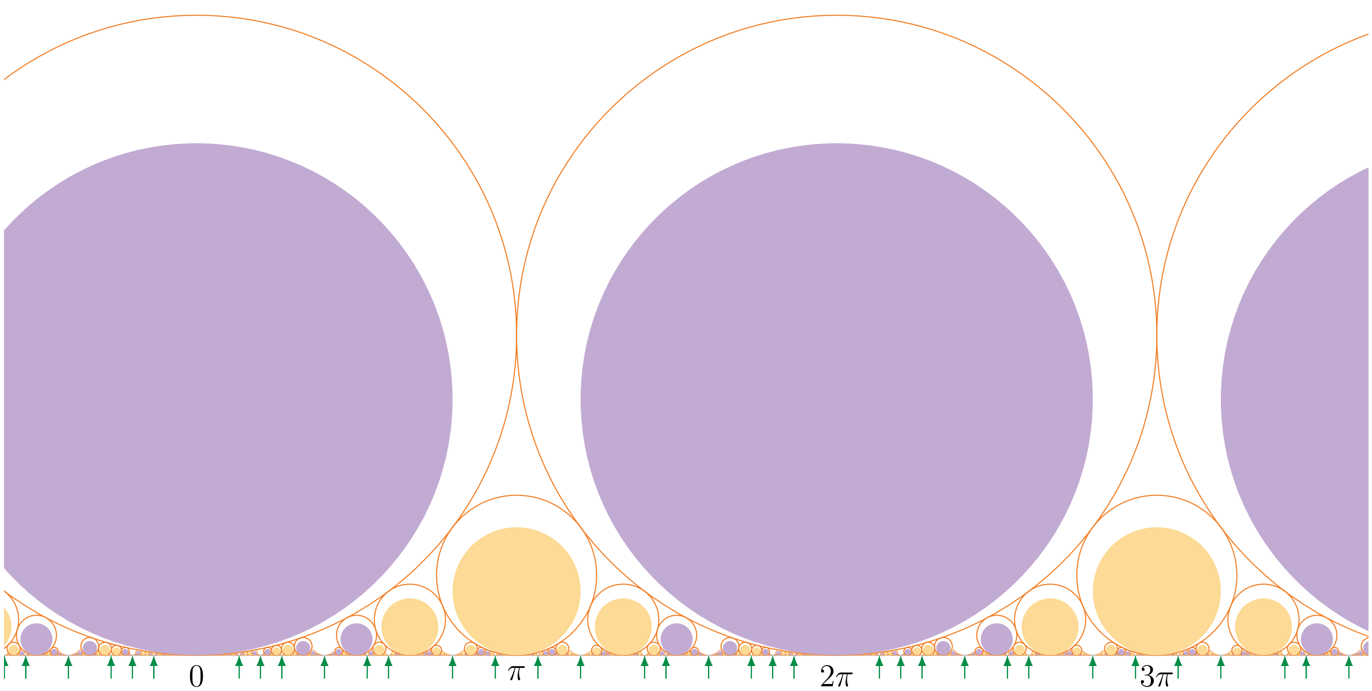
(The scaling factor for the smaller disks is ${}^{4}\!/\!{}_{c}$.)

As c grows, the coordinate t rescales, so the height of the strip goes down as ${}^{2\pi}\!/\!{}_{c}$, likewise for the step/pitch between the largest green circles. Moreover, the shaded disks would shrink (relative to the green circles); as a result, the horizon-self-similar zones become (relatively) more and more narrow. (This matches the behaviour we saw on our plots of $F^{(-1)}(t)$.)

All horizon-similar zones

Above, we already doubled the collection of disks we consider by adding red disks to the gray ones. However, in Remark 61 on p. 94 we introduced yet another way to double: via adding the transformation $T \mapsto -{}^{1}\!/\!{}_{cT}$ (of Hecke's functional equation; here $T = {}^{t}\!/\!{}_{2\pi}$.) This transformation would multiply[274] $f_{\mathbb{C}}$ and $F_{\mathbb{C}}$ by a (complex) constant (which may be 1). In coordinate x this transformation becomes $x \mapsto -{}^{c}\!/\!{}_{x}$. **Conclusion:** to account for these additional zones, we need to add to the picture of Ford circles above its transform under $x \mapsto -{}^{c}\!/\!{}_{x}$.

However, one can immediately see that $z \mapsto -{}^{1}\!/\!{}_{z}$ preserves the Ford–Apollonian gasket. (Here we extend the coordinate x on the horizontal axis to a coordinate $z := x + \mathrm{i}y$ on the upper half-plane with $\operatorname{Im} z \geq 0$.) Hence a transform of the Ford–Apollonian gasket by $z \mapsto -{}^{c}\!/\!{}_{z}$ is the same gasket upscaled c times. What remains is to shade the corresponding disks (purple and yellow, depending on whether the preimage of the Ford circle contains a gray or a red disk):



[273] Note that in the disk model, we had a gray disk tangent to the absolute at $t = \infty$. In half-plane model it becomes a half-plane $\operatorname{Im} x > \text{const}$. We do not shade it, since it does not contribute to the zones in question anyway!

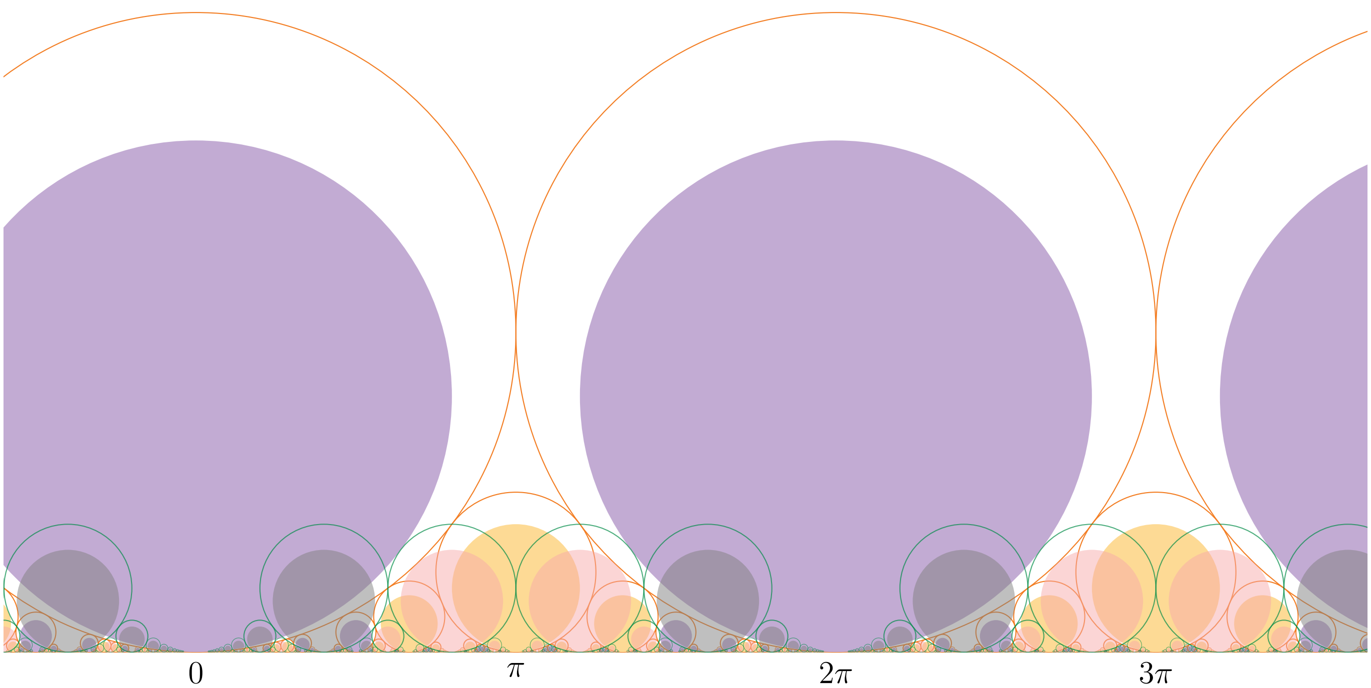
[274] Here again we need to consider F and f as *tensor* fields. (See Footnote 240 on p. 88.)

This time we must omit circles with tangency points $t = 2\pi R/D$ with $c|D$ (we mark a few of them with green arrows), and the color depends on $\left(\frac{D}{c}\right)$ (violet is for $\left(\frac{D}{c}\right) = 1$). **Conclusion:**

The orange circles and the green circles are tangent to the absolute in two complementary subsets of $2\pi\mathbb{Q}$.

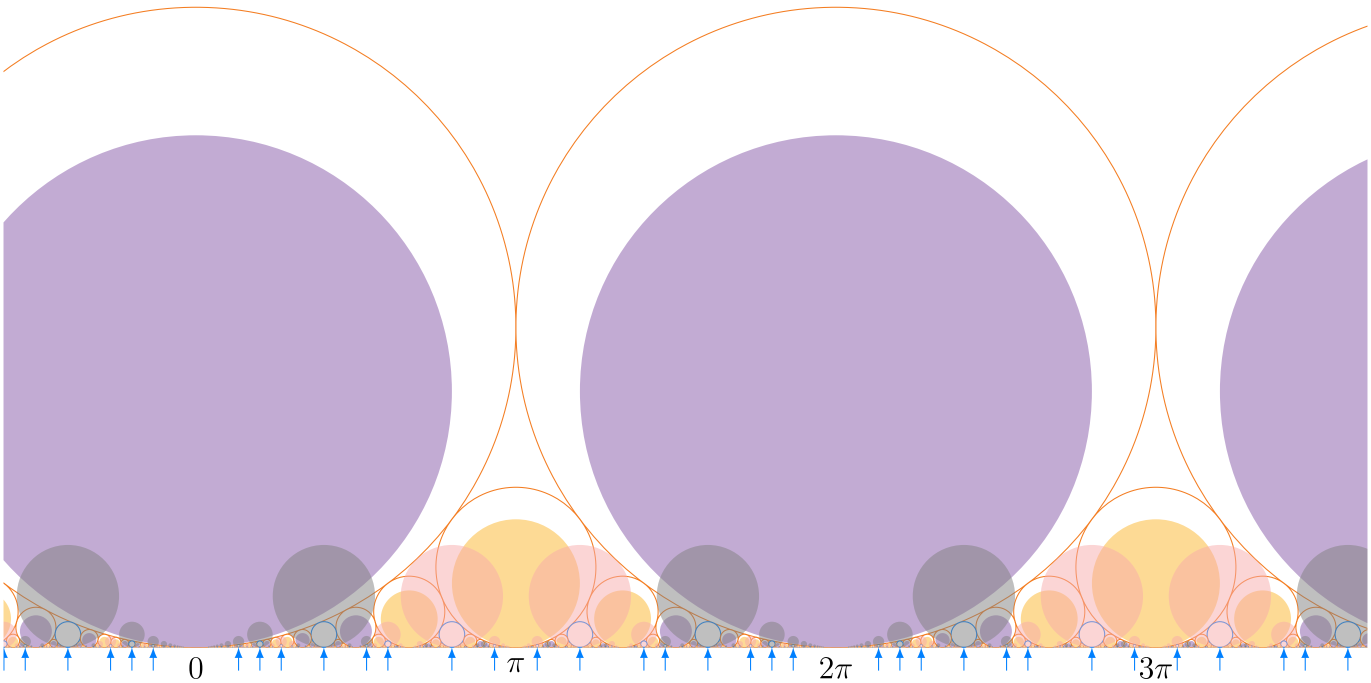
In other words: the tangency points of green circles on the picture with gray and red disks on p. 95 coincide with positions of "omitted" Ford circles in the pattern of orange *circles*.

Overlaying the last two pictures on top of each other gives:



Even if we ignore the *disks*, the orange and green *circles* look like a mess. But we can fix this!

Indeed, now, as in the Ford arrangement, every rational multiple of 2π on the boundary is the tangency point of exactly one orange or green circle — but while the diameters of orange circles are given by Ford's rule ($1/D^2$ on p. 95), the diameters of the green ones are c times too large. The fix is to replace every green circle by a blue one with the same tangency point and c times smaller diameter:



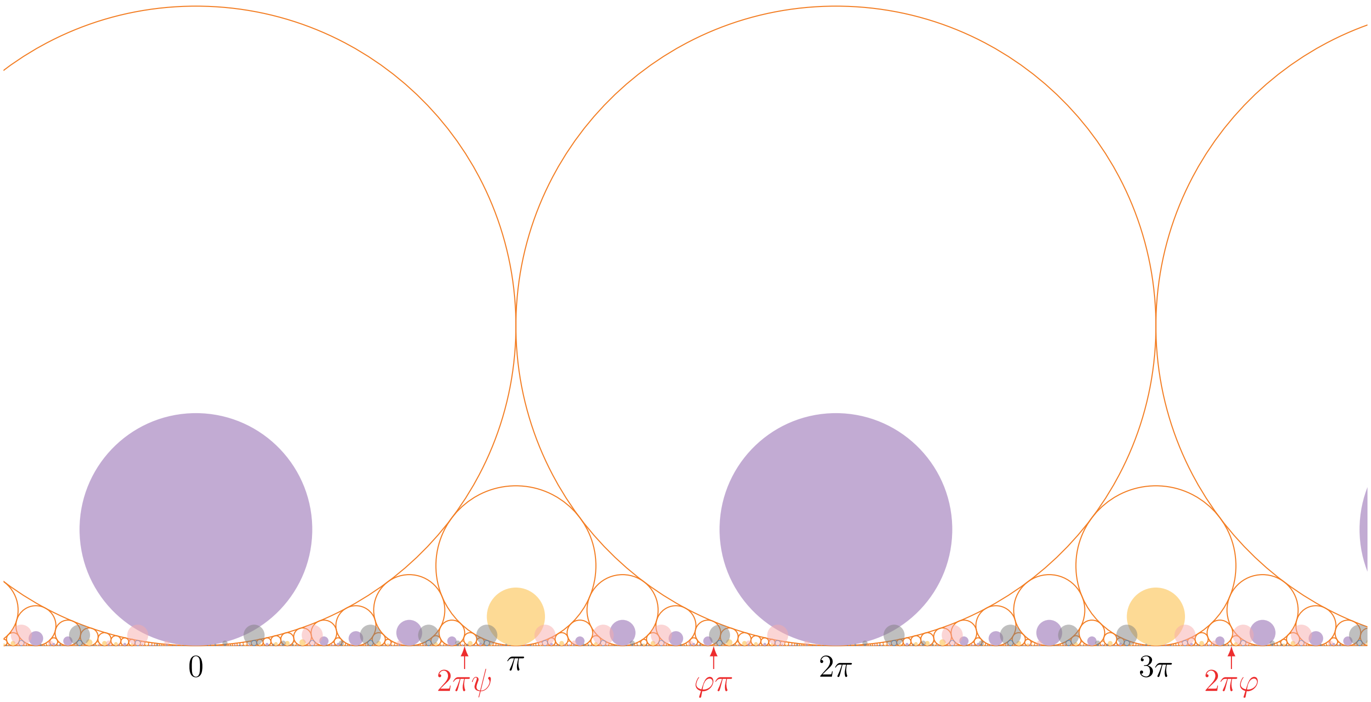
This way, the orange and (tiny) blue *circles* form a perfect Ford pattern. Moreover, the remaining visual mess *of the disks* can be clarified by a simple recipe for their diameters:

“Inscribe” smaller disks into the orange Ford circles. For every blue circle, “outscribe” a larger disk.

(The scaling factors are ${}^{4}/_{c}$ and 4 for orange and blue circles correspondingly. The color is choosen depending on $c|D$ in $t = 2\pi{}^{R}/_{D}$.) Summarize the relation of this picture with the fractal properties of $F^{(-1)}(t)$:

- Projections of gray and red disks are zones of visual horizon-self-similarity.
- Projections of violet and yellow disks are zones of visual horizon-similarity (self- for “even” case, to-Im $F_{\mathbb{C}}^{(-1)}(t)$ otherwise).
- In projections of yellow and red disks the horizon-similarity “flips” sign (but not for violet and gray disks).

When c grows (but remains prime), blue circles become more scarse (they match denominators D divisible by c) and smaller. So although the size of gray and red disks relative to blue circles is c-independent, their possible sizes go down with c, — and the rate of going down is similar to one for violet and yellow disks. For example, below we illustrate $c = 11$:



Here $\varphi := (1+\sqrt{5})/2 \approx 1.6180$ and $\psi := 1/(2 + 1/(1+\varphi)) \approx 0.4198$.

As above, orange and blue *circles* form the Ford pattern, violet and yellow *disks* are incribed in orange circles, gray and red are outscribed "in" (tiny!) blue circles. Four different colors of disks correspond to the types of symmetry described in the bullet list above.[275]

Complement to zones

One can restate our construction of the disks above for a prime c this way: to find a disk whose projection contains a given number $t = 2\pi\alpha$, we need to solve $|\alpha - R/D| < 2/c \cdot 1/D^2$, or to solve $|\alpha - R/D| < 2 \cdot 1/D^2$ with $c|D$ (which is equivalent to solving $|c \cdot \alpha - R/D| < 2/c \cdot 1/D^2$ with $c \nmid R$). Hence any number which is "not badly approximable" (can be approximated by rationals with more than quadratic precision) is in such a projection. (It is well-known that badly approximable numbers are "very rare": they form[276] a "meagre subset of measure 0".)

To give an example of such an exceptional number, we need α such that both α and $c \cdot \alpha$ are "sufficiently" badly approximable. However, for $c = 11$, while φ is the usual suspect for an example of badly approximable numbers,[277] the number $11\varphi = 17 + 1/(1 + 1/(3 + 1/(11\varphi + 6)))$ has infinitely many continued fraction coefficients being $17 + 6 = 23$ — hence it has many approximations good for $c < 46$. Because of this, $2\pi\varphi$ is in the projection of a gray disk for $c = 11$.[278][279]

---

[275] If c is not prime, one may need to repeat this process of addition of new colors of disks. For example, if $c = p_1^{a_1} \dots p_l^{a_l}$ with distinct p_l, then instead of one extra symmetry $-1/cT$, it is possible to define l "independent" fractional-linear symmetries $w_{p_1^{a_1}}, \dots, w_{p_l^{a_l}}$ (see Lemma 9.24 of Knapp's *Elliptic curves*). This would increase the number of colors for the disks to 2^{l+1}.

[276] It is still the same formulation as we had in Footnote 256 on p. 91. However, now we can relate our set of exceptions to a classical problem in number theory. In particular, any upper bound on the Hausdorff dimension of the set of solutions to $|\alpha - R/D| \ge 2/c \cdot 1/D^2\ \forall R, D$ works as an estimate for our exceptional set as well.

[277] It cannot be approximated by rationals with the required precision for $c > 2\sqrt{5} \approx 4.472136$.

[278] This is *almost* visible on the picture above — but one may need to zoom in *a lot*.

[279] Moreover, it is way easier to see that $\varphi\pi$ is in a projection. This is not surprising since it is much easier to approximate $\varphi/2$ (and also $11\varphi/2$) by rationals.

On the other hand, both ψ and $11\psi = \varphi + 3$ are extremely badly approximable by rationals — and $2\pi\psi$ on the picture above behaves correspondingly.[280]

Remark 62: Here we want to get a very rough heuristical estimate of which part of the absolute is covered by the projections of the disks. First, focus on the projections of disks with the given denominator D; they cover the fraction $\approx \varepsilon_D \varphi(D)/D^2$ of the absolute; here $\varepsilon_D = 4$ if $c|D$, and $\varepsilon_D = 4/c$ otherwise. The averaged value ε of ε_D is about $8/c$. Heuristically, it looks reasonable to assume (as the 0th approximation!) that for different D, the intersections of zones behave as if the zones were "independent". This leads to the estimate $\prod_{D=1}^{d} \left(1 - \varepsilon_D \varphi(D)/D^2\right)$ for the relative size of what is not covered by projections of disks with $D \leq d$.

This product decreases as $\mathrm{const} \cdot d^{-6/\pi^2\varepsilon}$. We can estimate that to decrease the uncovered part by half, we need to increase d about $10^{c/16}$ times; here we use $48/\pi^2 \log_2 10 \approx 16.16$.

So for $c = 23$, to see half of the graph of $F^{(-1)}$ covered by the horizon-similar zones, we need to zoom so that we can see zones of width $1/27$ of the period. On the other hand, for $c = 971$, one would need to zoom about 10^{60} times.

Numerical experiments (easily done up to $c = 59$) show that this estimate gives quite a good match. For example, for $c = 23$ it turns out that to cover about half of absolute, one needs $d = 23$ (instead of 27 above).[281]

---

[280] Similar examples of badly approximable numbers Ψ and $c \cdot \Psi$ (those with the tail of continued fraction coefficients being 1,1,1,...) exist for a prime $c = p$ when p has a quadratic residue mod 5. (For $c < 50$, this gives $c = 5, 11, 19, 29, 31, 41$.) We sketch a very rough scheme of the proof below.

First, one can immediately see that both Ψ and $c \cdot \Psi$ should have the form $(\alpha\varphi + \beta)/(\gamma\varphi + \delta)$ with integer coefficients and $\alpha\delta - \beta\gamma = 1$. From this it is easy to deduce that the condition above is necessary.

Moreover, if $\gamma, \delta > 0$ and $\Psi = (\alpha\varphi + \beta)/(\gamma\varphi + \delta) > 0$, then the continuous fraction for Ψ has the required form. Try to solve $c \cdot \Psi - \varphi \in \mathbb{Z}$; this equation can be reduced to $\gamma\delta - \gamma^2 + \delta^2 = c$ having integer solutions. Indeed, given such a solution, one can find α, β with $\alpha\delta - \beta\gamma = 1$ and put $\Psi := (\alpha\varphi + \beta)/(\gamma\varphi + \delta)$; then $\delta | c\beta - \gamma$ and $c \cdot \Psi = \varphi + (c\beta - \gamma)/\delta$ as required.

Furthermore, any solution to $\gamma\delta - \gamma^2 + \delta^2 = c$ leads to another solution $\gamma' = 2\gamma + \delta$, $\delta' = \gamma + \delta$; moreover, if $\delta > 0$, then $\delta' > 0$ and $\gamma' > \gamma$. Iterating this, one can immediately see that if a solution exists, there must be solutions with $\gamma, \delta > 0$.

(The rest requires more esoteric math. Existence of a solution to $\gamma\delta - \gamma^2 + \delta^2 = c$ can be investigated via Hasse's local-global principle; in the case of indefinite binary form $\gamma\delta - \gamma^2 + \delta^2$ it says that it is enough to find solutions $\bmod p^k$ for all $k \geq 1$ and *all* prime divisors p of $2c|D|$; here D is the discriminant, so $|D| = 5$. Moreover, the Product Formula for Hilbert symbol shows that one can replace "*all p*" above by "*all but one*". Skipping $p = c$ leaves just $p = 2, 5$ — which implies that the answer depends only on $c \bmod 2^a 5^b$ with $a, b \gg 0$. A simple check improves this to $a = 0$, $b = 1$, and the criterion above.)

[281] In fact, this change from 27 to 23 is "as expected" with a bit more precise analysis of the product above. Indeed, note that the change of the product when d goes from $kc - 1$ to kc is approximately as large as the change between kc and $(k+1)c - 1$. Because of these, the answer for "when projections cover ½ of the absolute" tend to "be attracted" to multiples of c.

With such a correction, our estimate is reasonably good already for $c = 7$, and the total length of projections with d given by this formula tends to have only a tiny systematic error: it is close to $1/2 + 1/c$ instead of ½.

Degrees higher than 3: the same "hidden symmetries" as for degree 3 appear only if an "extra distillation" is possible

In construction! (Not yet optimized — but more or less complete — exposition.)

So far, in the context of the key question punctuating our notes:

Which prime numbers p appear as divisors of elements P_m of a polynomial sequence?

we encoded the answers into a sequence $(\widetilde{N}_n^{\mathrm{res}})$ of numbers, and demonstrated two situations (each "working" for certain $P_\sqcup$s) when *distillation*[282] to sequences N_n can uncover "hidden symmetries" in these answers:[283]

Rank=1 The sequence N_n is periodic (and either even, or odd[284]). Periodicity and even/oddness are conditions of symmetry![285]

Rank=2 The Fourier transform $F(t)$ of the sequence N_n has a rich collection of fractal symmetries.

We saw the case of rank 2 appearing for "non-abelian" (irreducible) polynomial sequences of degree 3. For P of degree 1, our "distilled sequence N_n" is[286] 0. The remaining irreducible[287] cases of degree ≤ 3 lead to sequences of rank[288] 1.

In fact, we saw that for a prime p, the sequence $(a_k) := (N_{p^k})$ satisfies a certain linear recurrence relation $a_{k+r} = C_{r-1}a_{k+r-1} + \ldots + C_0 a_k$ of length r. Moreover, the minimal possible value of r is the same for all p (with only a finite number of exceptions) — and it coincides with the rank.[289]

This bird's eye view leads to two most important conclusions:

- The "hidden symmetries" in different ranks looks completely unrelated. (Compare with Footnote 47 on p. 19 and Remark 49 on p. 77.)
- When one considers a "not fully distilled" sequence N_n, it is "merged" from separate "fully distilled parts", hence the parts have "unrelated" symmetries. Therefore, these symmetries disappear when these parts "are merged together".[290]

Finally, note that the abelian case for degree 3 is much more rare than the non-abelian: it happens when the discriminant is a square.[291] In particular, the hidden symmetry in the *general* case of a polynomial of degree 3 has rank = 2; so it cannot be exposed in the same way as in the degree 2.

The purpose of this chapter is to demonstrate that for larger degrees, the same effects continue to hold:

---

282 . . . as in the section on p. 75.

283 Here we want to use the new notion of "rank". We explain this on p. 119.

284 Meaning that the periodic continuation to $n \leq 0$ either satisfies $N_{-n} = N_n$, or satisfies $N_{-n} = -N_n$.

285 In fact, this case has an extra symmetry: the sequence N_n turns out to be *totally* multiplicative: $N_{nm} = N_n N_m$ for *any* m and n. (We also use this notion in the section on p. 62. Note that being even/odd is a particular case of total multiplicativity — compare with Footnote 332 on p. 120.)

286 Moreover, the "undistilled" sequence $\overline{N}_n \equiv 1$ is the "trivial" case of rank 1.

287 Recall that for polynomials sequences which are products of two polynomial sequences, the "key question" may be reduced to the factors. So it makes sense to consider only irreducible polynomials.

288 For degree 2, the sequence N_n itself is of this form. In the "abelian" case of degree 3, the sequence N_n "distills" into two components ζ_n and $\overline{\zeta}_n$ of rank 1. (See the section on p. 67.)

289 **N.B. (???) Mappings to other groups? (Take a minimal strict representation (to G). Then any other is a sub in a tensor power, hence its image is included in the image of $G \subset \mathbf{GL}_n\mathbb{C}$.)**

290 Recall that the sequence N_n is *already* a result of the ("naive") distillation corresponding to breaking of $\overline{N}_{p^k}$ into two parts: 1 and N_{p^k}.

291 We put more details in Footnote 340 on p. 122.

Rank > 2 The hidden symmetries in the general case of degree ≥ 4 cannot be exposed in the same ways as in degrees 2 or 3.

Rank $= 2$ New cases For special (non-abelian) cases in degrees 4, the sequence N_n can be distilled into a sequence of rank 1 and a sequence of rank 2. Hence, one "distilled component" is periodic, the other gives $F_{\text{dist}}(t)$ which is an exact fractal. The hidden symmetries in the general case of degree ≥ 4 cannot be exposed in the same ways as in degrees 2 or 3.

Rank $= 1$ New cases In abelian cases of degree > 2,[292] the sequence N_n can be distilled into several sequence of rank 1 (possibly with complex coefficients).

Degree 4: the surprising (counter)examples

The main arc of these notes is the discussion of fractal properties of (the graphs of antiderivatives of) functions $F(t)$ "corresponding" to counting modular solutions of cubic equations. Here we demonstrate what happens for equations of larger degree *if one follows the naive (direct) analogues* of the construction of $F(t)$ working for degree 3. We are going to keep the notation $F(t)$ for such a result.

As annouced above, the Langlands program predicts that such cases are of higher rank 3 than what we considered before. In particular, one should not expect the resulting function $F(t)$ to have the same fractality law as what we considered above.[293]

Summary: while in degree 4 it is not a problem to produce the sequence N_n (related to the number of roots of our polynomial $\bmod n$), its Fourier transform $F(t)$ *is not expected* to have fractal properties. In particular, there is no reason to expect that *the same* patterns of fractality as what we saw for degree 3. On the other hand, the actual plots below show that:

- Surprisingly, the graphs show *some* of characteristic features of symmetry w.r.t. the toy transform: for example, observe the "hourglass" shapes appearing near $t \in \pi\mathbb{Q}$. The most prominent of these shapes is near $t = 0$ (with a more symmetric hourglass).
- However, the graphs *do not* look like a toy transform of a *periodic* function. (The "shape" of oscillations changes when we get closer to "the waist of the hourglass".)
- Furthermore, zooming near the waist of the hourglass shows the "flattened" zone: , provided we calculate the Fourier series by abruptly cutting the sum off at a certain place[294]. (In section on p. 145 we saw that these zones suggest a presence of discrete Fourier spectrum for[295] $F(1/t)/t$.)
- The fact that these features appear not only at $t = 0$, but also at other rational multiples of π is really surprising. (Recall that for degree 3, the horizon-self-similarity at $t = 0$ is due to Hecke's functional equation — which has direct analogues in every degree. So having something analogous at $t = 0$ — and, as a corollary, at points of the Cantor hyper-family[296] — is not *extremely* surprising. However, the validity at any $t \in \pi\mathbb{Q}$ is quite analogous to going from Hecke's functional equation to the whole Langlands program.)

I have no idea how to explain the appearence of these features. Below, the plots are provided without any explanation!

Recall that in plots for degree 3, we would start by showing the graphs near $t = 0$, but our final aim was to expose points t where horizon-*self*-similarity was present. For non-Maass cases,

[292] **N.B. (???) Not yet in this chapter!**

[293] Moreover, the "hidden symmetries" uncovered by the Langlands program are symmetries of a certain "new" object having very indirect relationship to $F(t)$. In fact, it seems that Langlands program does not say anything about $F(t)$ in these cases.

[294] … as opposed to doing something like Cesàro summation, where instead of abruptly cutting the terms, one uses a smoother cut-off function.

[295] Recall that for degree 3, it is periodic, hence has *only* discrete spectrum, at frequencies $1/4\pi^2 c\mathbb{Z}$.

[296] **N.B. (???) Ref!**

this required very strong zoom factors (quadratic in the conductor — which coincides with the field discriminant — compare with Footnote 440 on p. 144).

Since discriminants of irreducible polynomials of degree 4 cannot be very low, it is not computationally feasible to show such regions of horizon-self-similarity. Instead, we proceed as on p. 79 and show what *would be* a *non-trivial*[297] region of *just* horizon-similarity (in fact, it would be a horizon-similarity to $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$; see Footnote 205 on p. 79).

Moreover, recall that the conductor in the case of degree 3 was controlled not by the discriminant of the polynomial, but by its divisor, the *field discriminant*. (Compare with Footnote 440 on p. 144. For non-abelian case of degree 3, the conductor was equal to the field discriminant; for abelian, the conductor is the square root of this number.) So in the examples below we describe the field discriminant; moreover, it still seems that this number controls the size of the hourglass regions in very similar way to how it was working in degree 3.

In addition to what is described above, in the graphs below it makes sense to pay attention to:

- Near the waist of the hourglass, the amplitude of oscillation decays a tiny bit slower than in degree 3. This suggests that it is a toy transform of a function of very slow growth.[298]
- Absense of jumps and/or log-spikes (compare with our Eisenstein plots on p. 63; they also appear in the first — not-fully-distilled — example below) indicates that what we have may be related to something *cuspidal* (see Footnote 373 on p. 128).

This leads to

Conjecture: *for sequences N_n corresponding to polynomials of degree 4, the Fourier transform of N_n/n behaves as described above, unless the Galois group is abelian or dihedral.*

In turn, this may be amplified this way (compare with the section on p. 171):

Question: *Does this hold for all "Artin cases" of "components" of actions of finite Galois groups?*

Finally, recall that discriminants of "random" polynomials of high degree have a tendency to be very high. So instead of taking "beautiful" polynomials (such as our M-family, see the section on p. 144), below for every "type" we choose a special polynomial with the field discriminant as low as possible.

This finishes the introduction, and finally, we can provide the examples themselves.[299]

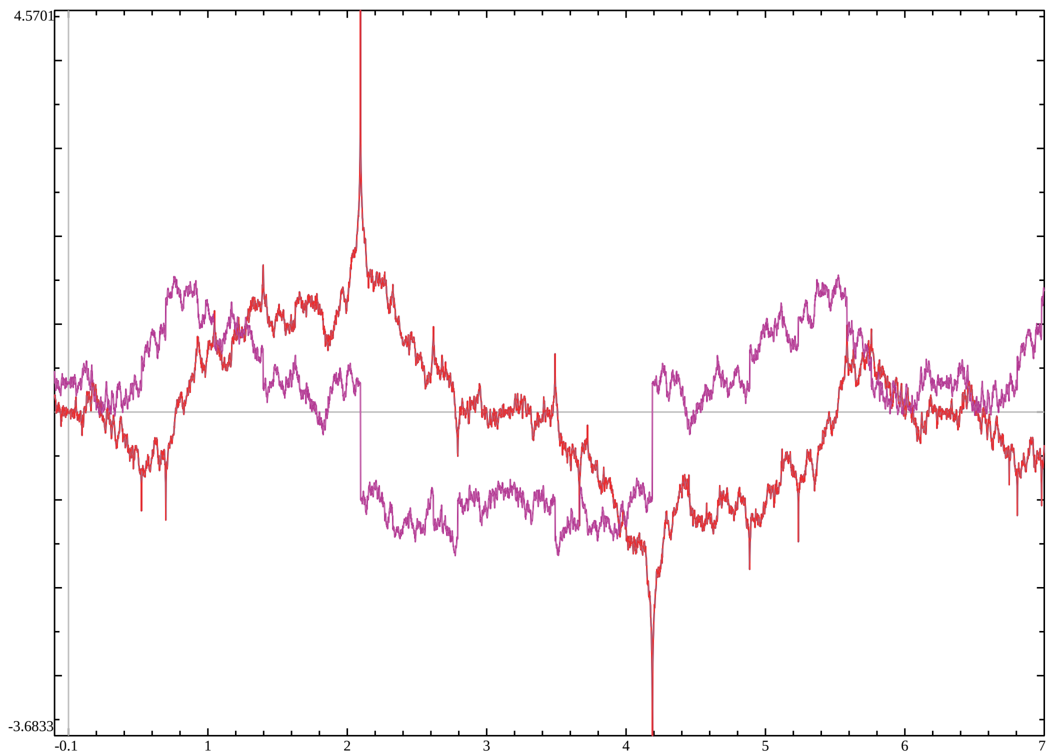
We start with the polynomial $x^4 - x^3 - x^2 + x + 1$ with the smallest magnitude of the field discriminant: $D = 9 \times 13 = 117$. It has no real roots and is not abelian, and the Galois symmetries form the dihedral group D_4. This implies that the corresponding motive (of rank 3) is not fully distilled (it breaks into two, of ranks $1 + 2$; compare with Footnote 551 on p. 172), so it is not surprising that the corresponding graphs of $F^{(-1)}(t)$ "change via jumps"! This shows a bit more than one period of

[297] Recall that for degree 3, the "trivial" points of self-similarity are those in Cantor hyper-family. If the function $F(t)$ has extra symmetries (as in the section on p. 55), the images of trivial points under these symmetries should be also considered "trivial".

[298] Examples below suggest logarithmic growth; see the plot on p. 108.

[299] The polynomials in these examples may be found using the "Online tables of number fields" (see the notes on p. 215). The gory details on Galois group for polynomials of degree 4 are summarized in Keith Conrad's notes *Galois groups of cubics and quartics (not in characteristic 2)*.

the real and imaginary parts:

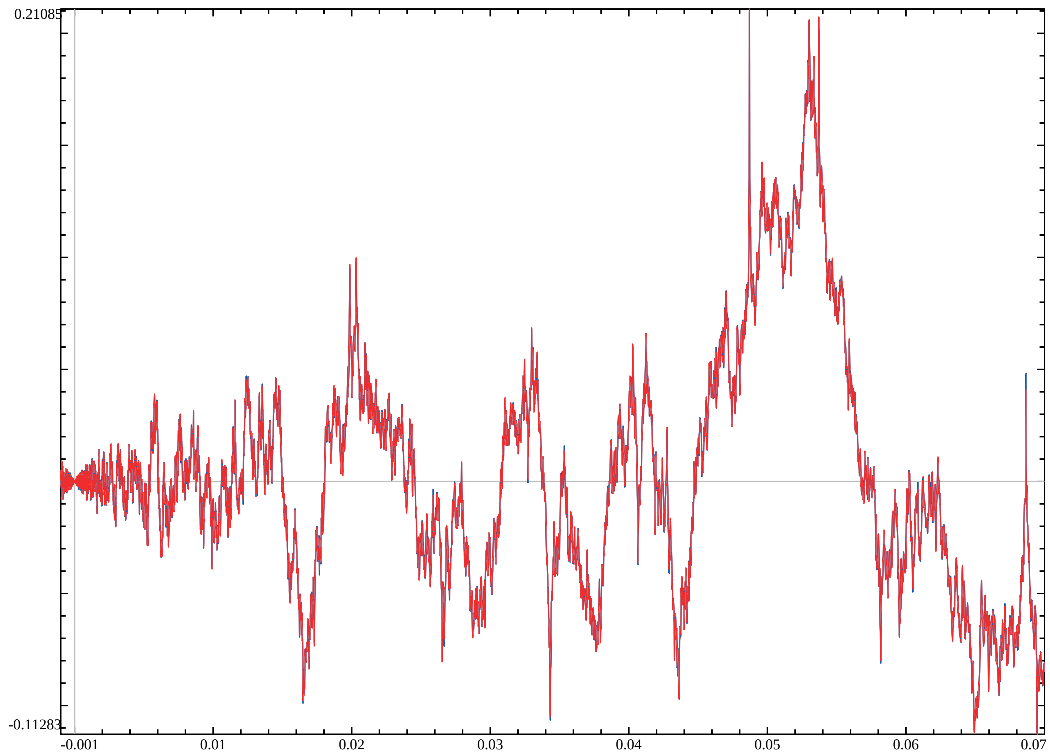


The real part has log-spikes, while the imaginary parts has jumps[300] (this is similar to what we already saw in other examples of non-distilled motives[301]).

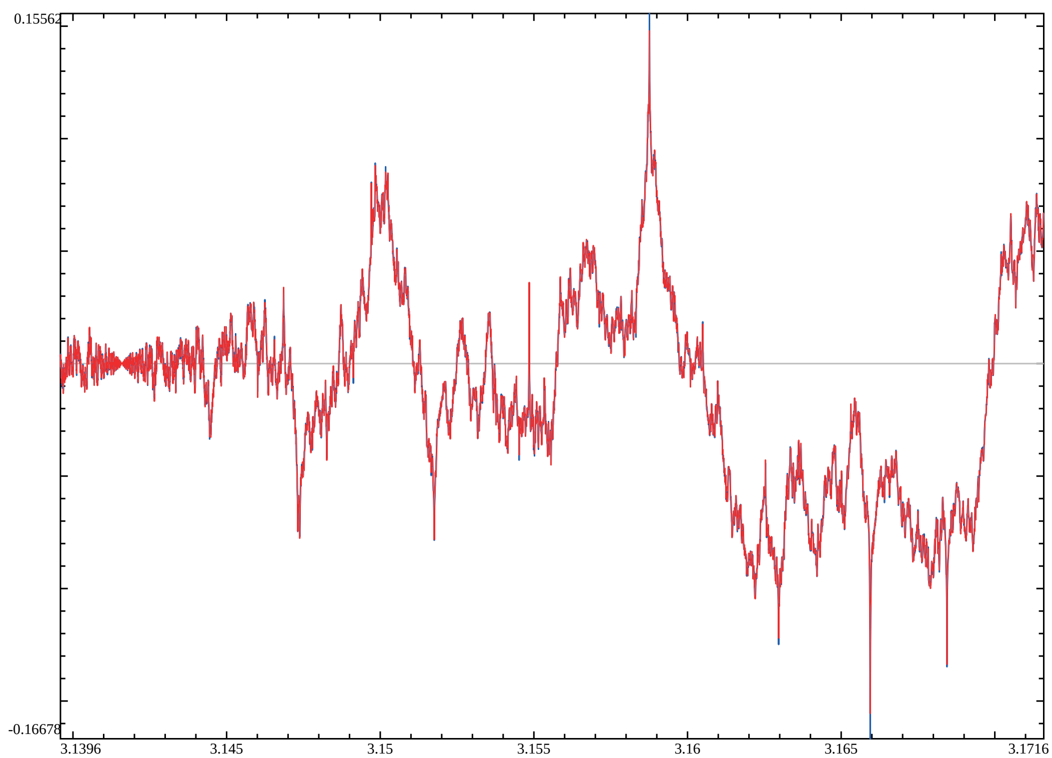
[300] These two features are closely related. See Footnote 151 on p. 63.

[301] Note that this is a much more complicated example than the non-distilled motives we demonstrate in these notes: they either mix a constant sequence N_n with a periodic N_n (for reduciblel cubic polynomials; see the section on p. 63), or two periodic N_ns (for abelian=cyclic cubic polynomials). *This* example mixes a "periodic" motive with a "modular form" motive.

Here is how it looks near 0:[302]



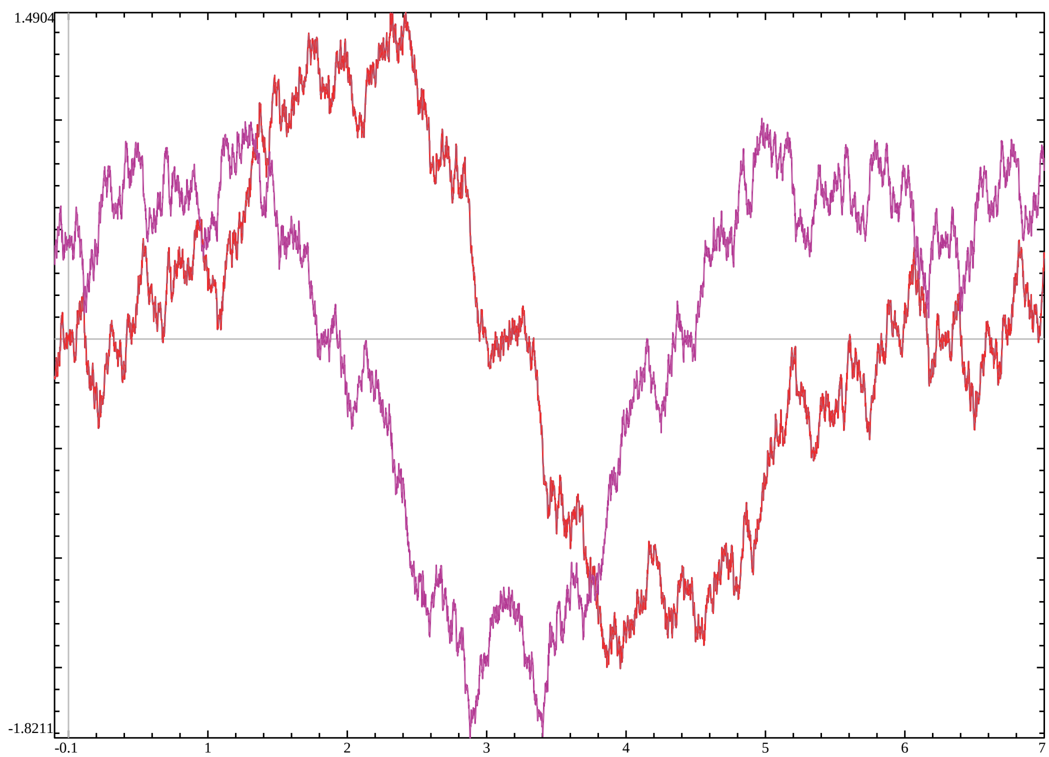
Here is the magnified view near $t = \pi$:



[302] Zooming into this graph shows that the top/bottom asymptotic near 0 are not linear, and follow the $C|t\log(t/K)|$-law we discuss below for field discriminant $D = 229$.

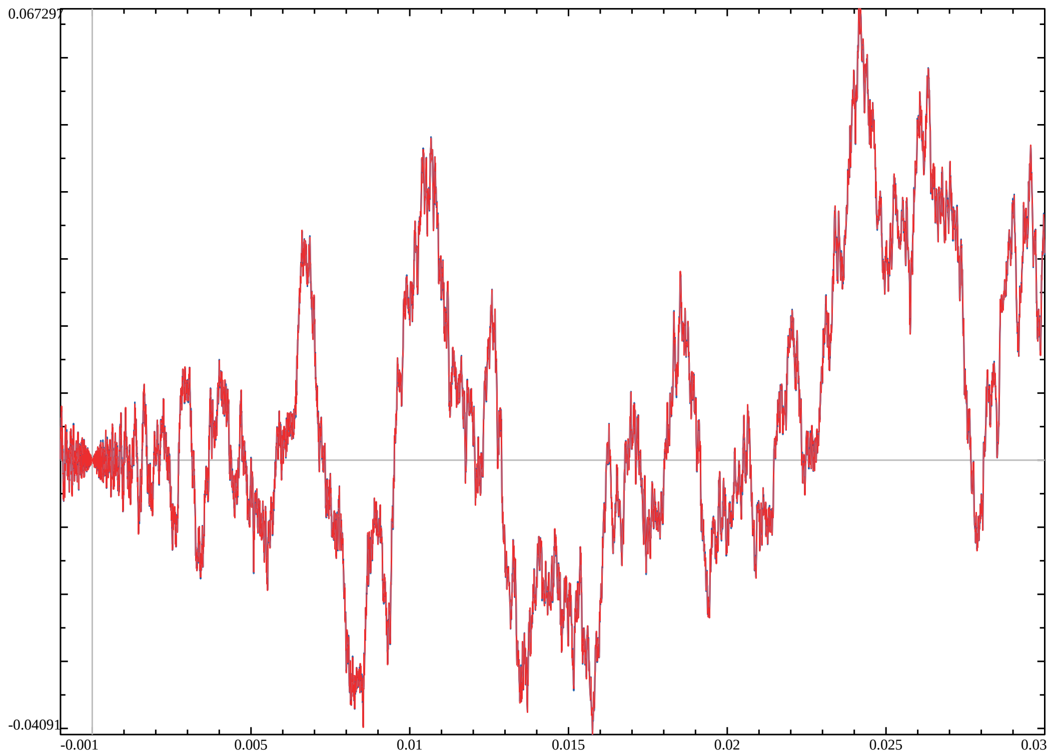
While the behaviour on the sides of the jump does not look like toy transform of a *periodic* function, the "hourglass" shape shows that it is a toy transform of an oscillating function *of very slow growth*! (Another interesting feature is the presence of noticable "spikes" even that close to $t = \pi$. In examples we saw before the widths of spikes were exponentially decreasing, so there were very few "high" spikes visible with our discretization.)[303]

Our next example is the polynomial $x^4 - x + 1$ with the smallest magnitude of the field discriminant for the case of a fully distilled motive: $D = 229$ (which is a prime number). It again has no real roots, and the Galois symmetries form the symmetric group S_4. Observing about one period of the real and imaginary parts:

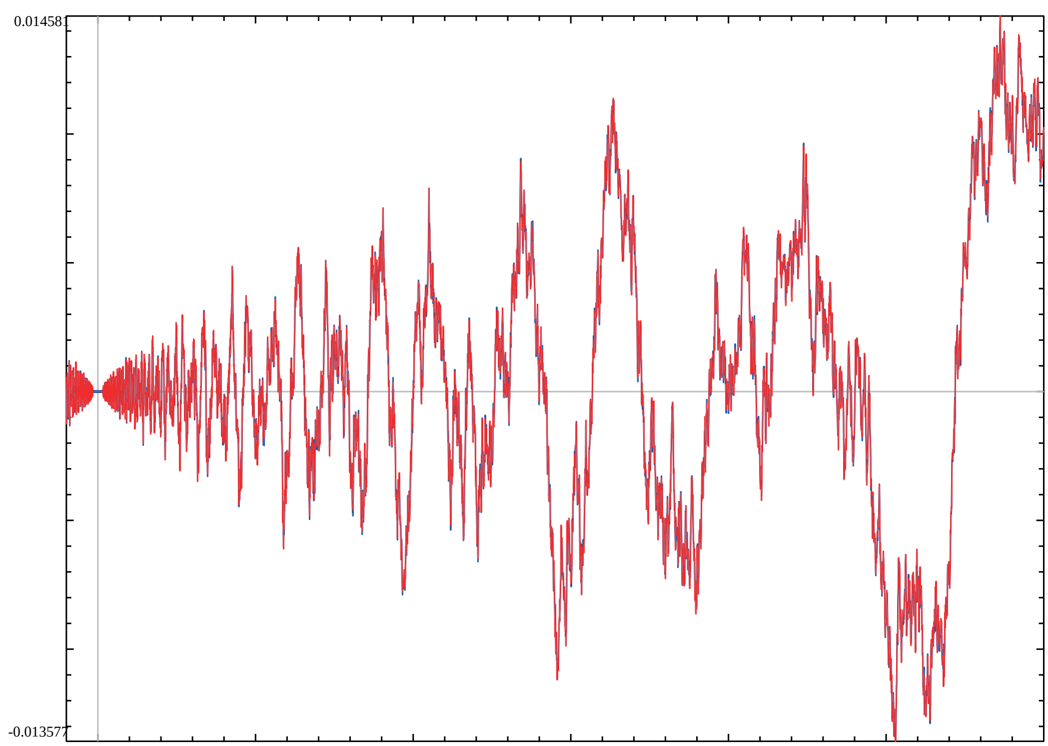


[303] **N.B. (???) Plot also $F^{(-1)}(t)$ for the irreducible 2-dimensional representation of the Galois group.**

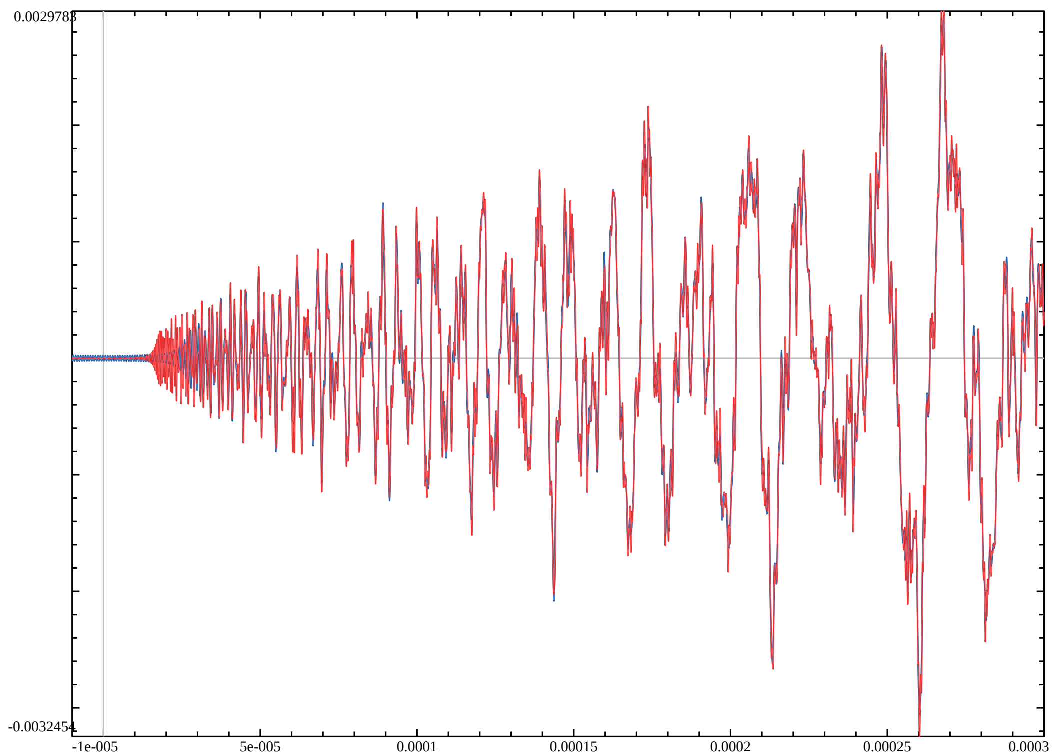
shows neither jumps/spikes, nor other types of discontinuity. When we looks near 0:



we again can see a "hourglass": the behaviour resembling two top- and bottom-asymptotes as $t \to \pi+0$ which are "almost straight". Moreover, zooming in 10 times shows that these "asymptotes" become steeper when we get closer to the "waist":

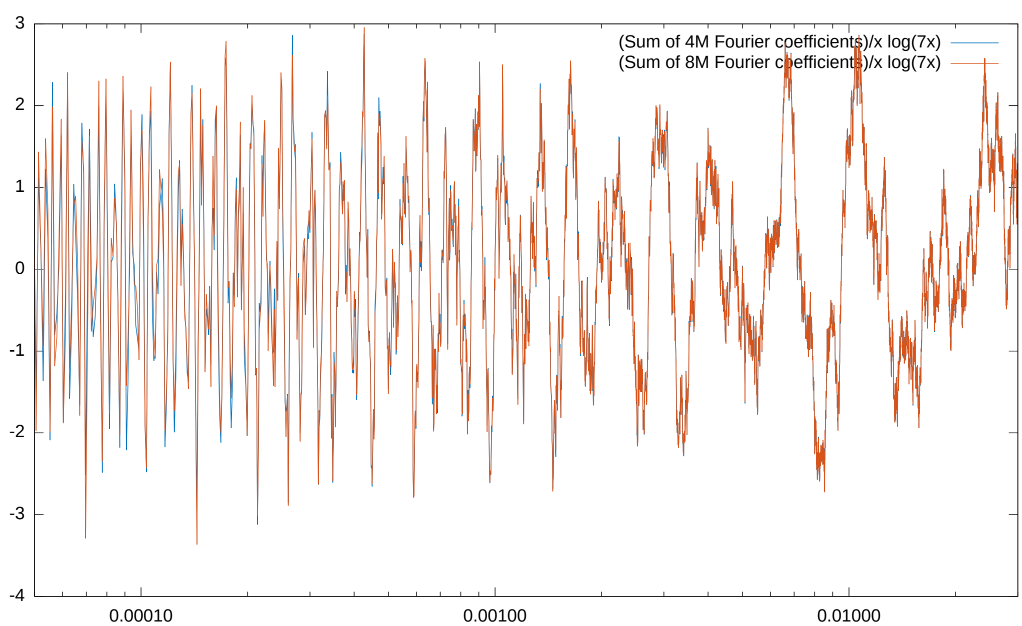


(So we can start to suspect that the "hourglass" is actually non-linear!) Zoom in 10 times more:



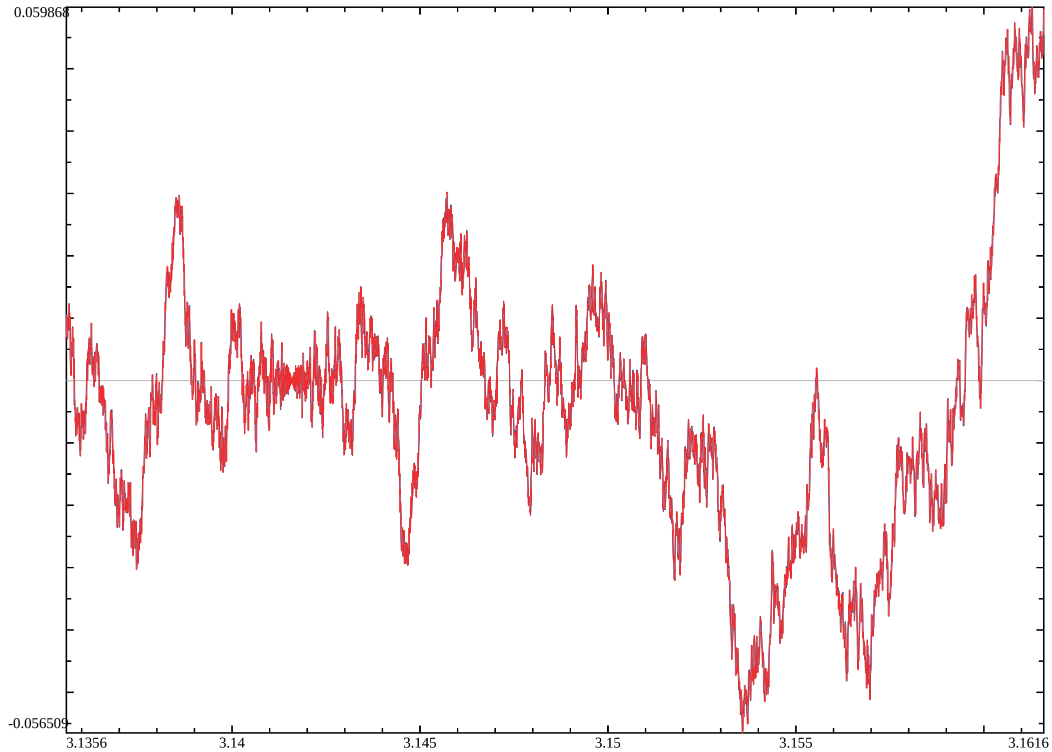
Here the "hourglass asymptotes" look almost linear now.[304]

In fact, the non-linearity of the "asymptotes" of the "hourglass" suggests that $C \cdot t \log t$ may be a better approximation for the asymptotic behaviour. This looks very plausible: on this plot we divide by $t \log t$ (the $\log t$ horizontal coordinate allows better view of what happens on different scales; the precision is abysmal near the left edge of the plot):

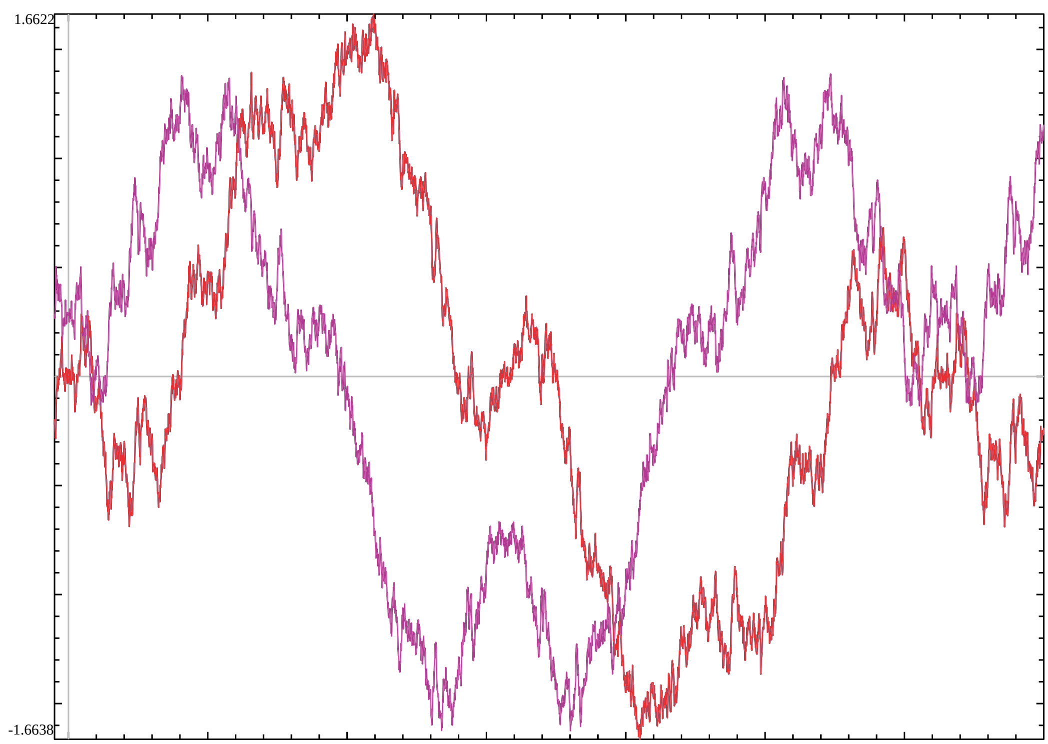


[304] However, a flattened zone becomes very visible. See Footnote 122 on p. 51.

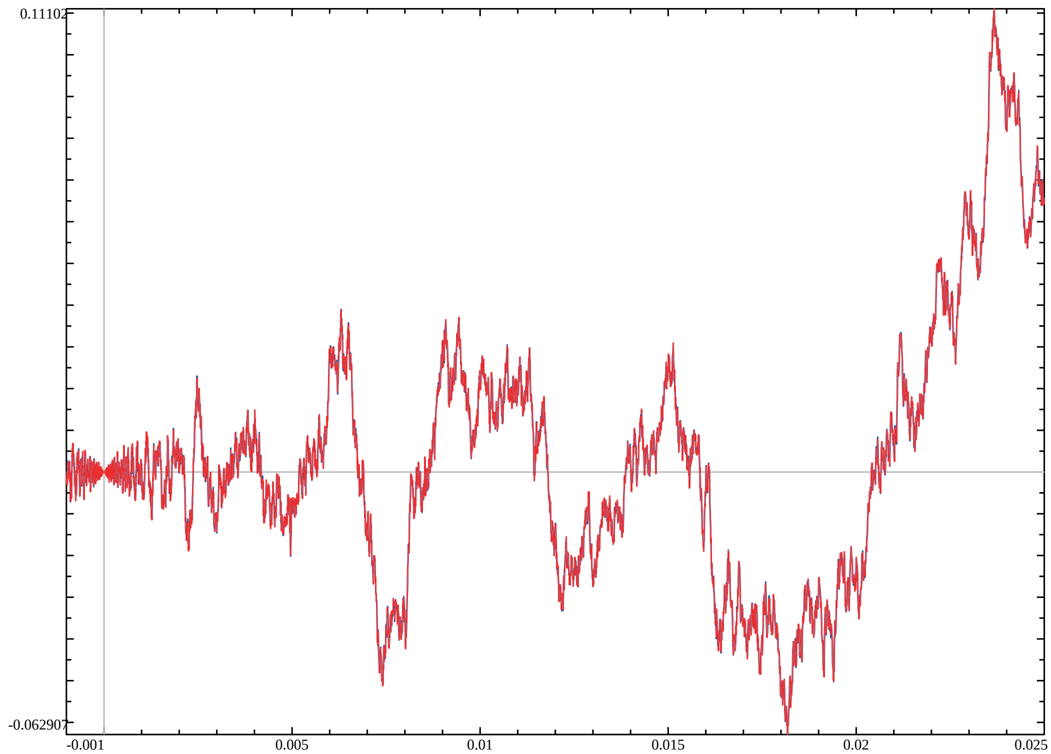
One can check that near $t = \pi$ very similar effects appear (we do not show the pictures with higher zooms — but they behave as above):



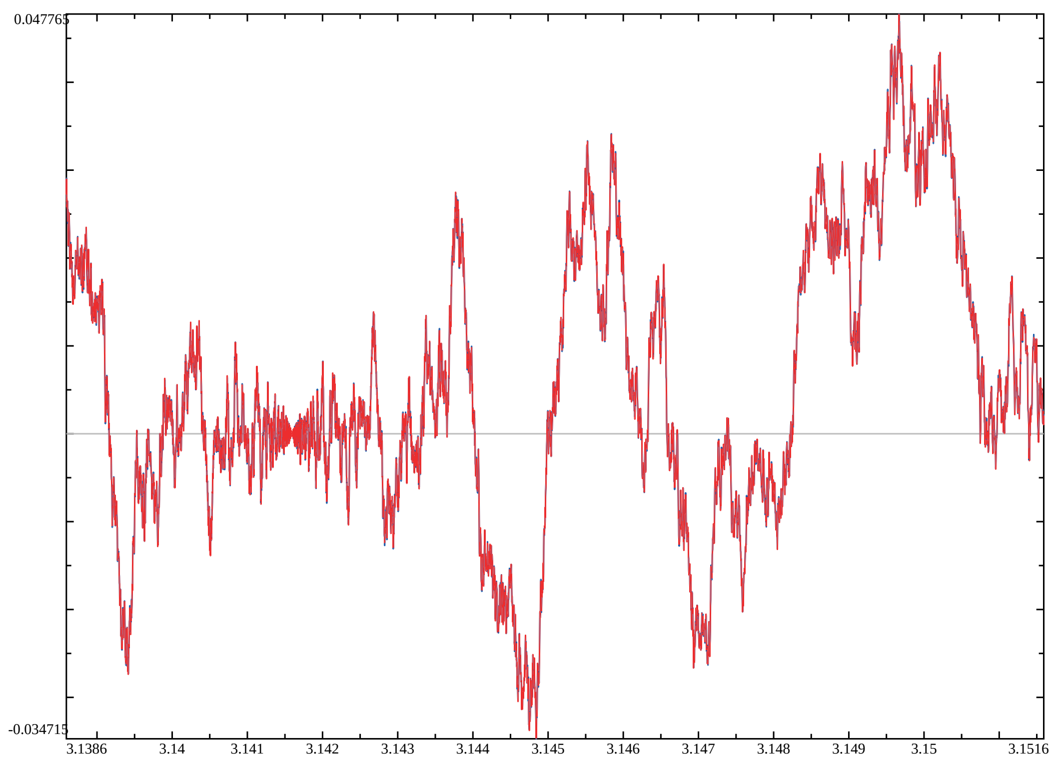
The minimal magnitude of a negative field discriminant with a fully distilled motive is $D = -283$ (which is a prime number). The polynomial is $x^4 - x - 1$ with two real roots and the Galois group S_4. The plots as above still show no visible discontinuities:



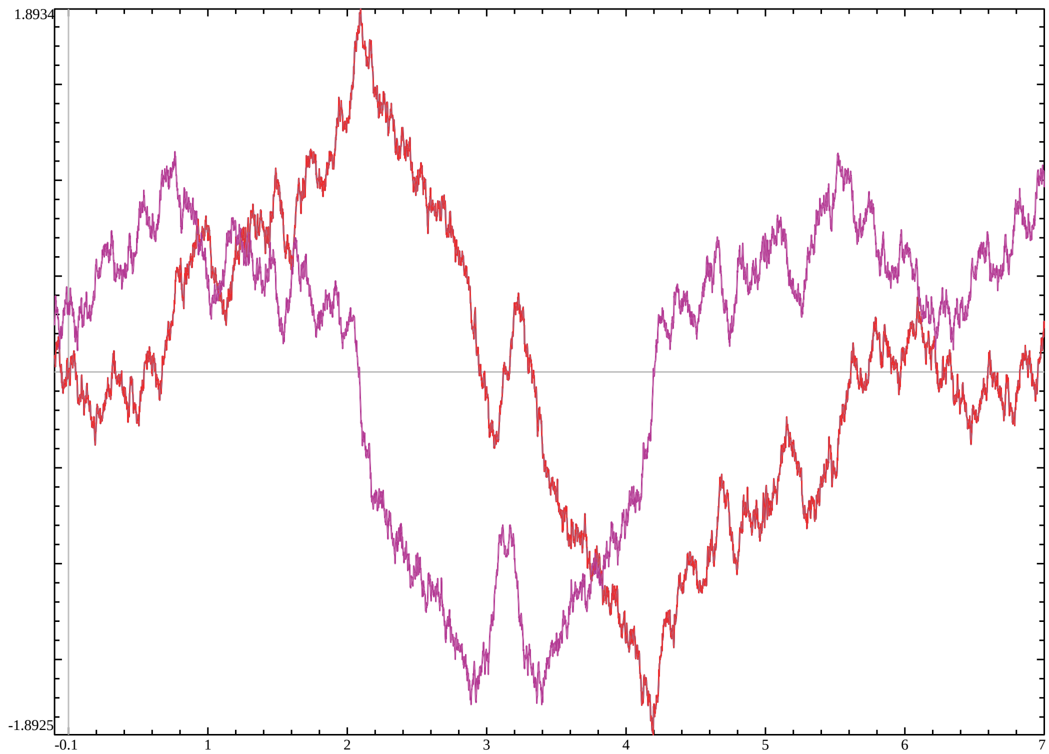
Here is how it looks near 0:



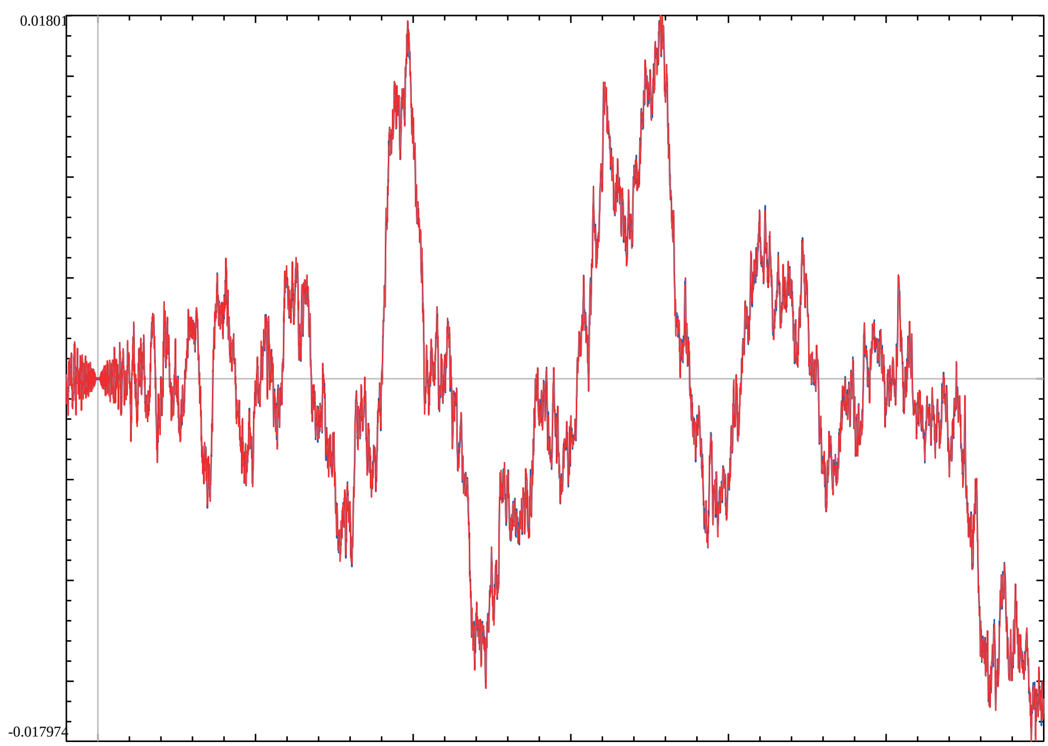
Near $t = \pi$ one can see yet another “irregular hourglass”:



Next, consider $x^4 - 4x^2 - x + 1$; it has 4 real roots, and the smallest field discriminant for such a case of a distilled motive[305]: $D = 19 \times 103 = 1{,}957$. Its Galois group is S_4. Here is one period of the real and imaginary parts:

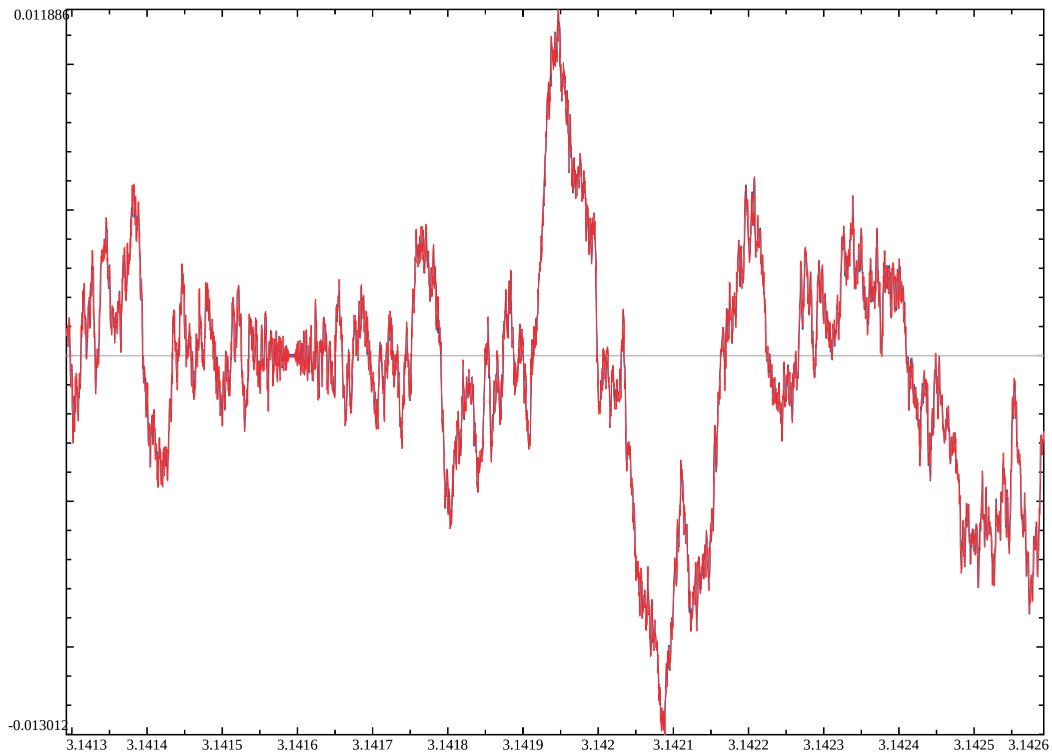


Near $t = 0$ it still shows no horizon-similarity:



[305] Moreover, another measure of complexity, the narrow class number (hence class number) turns out to be trivial: 1. So this example is “the simplest one” in all the possible senses.

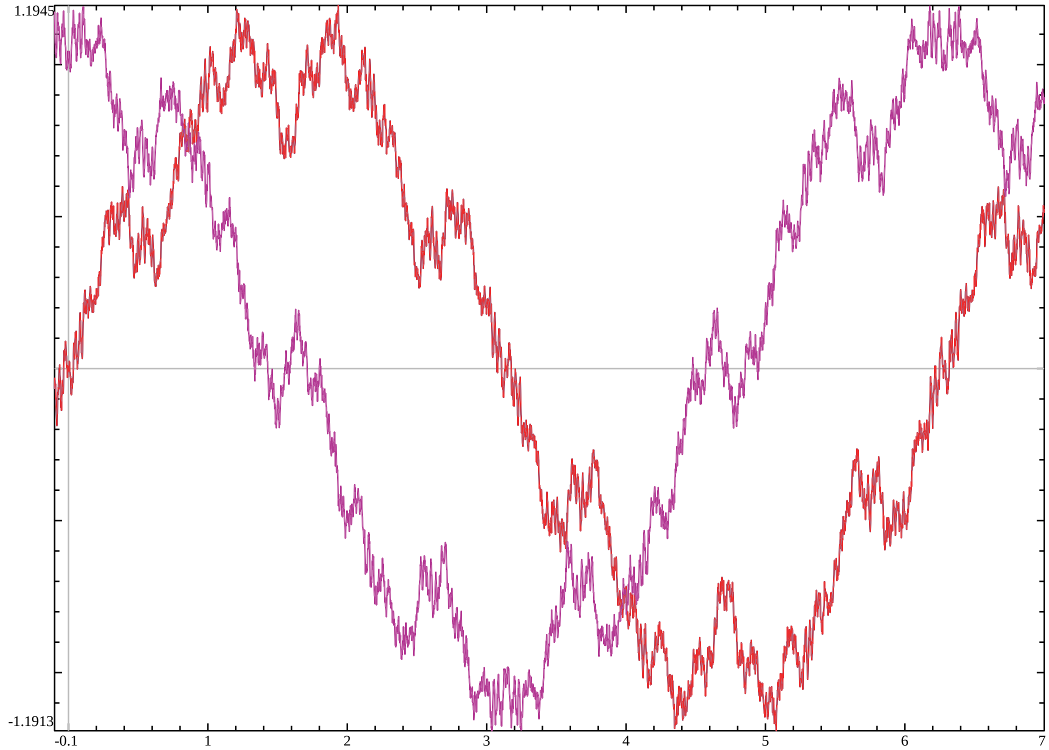
The behaviour near $t = \pi$ should not be surprising now:



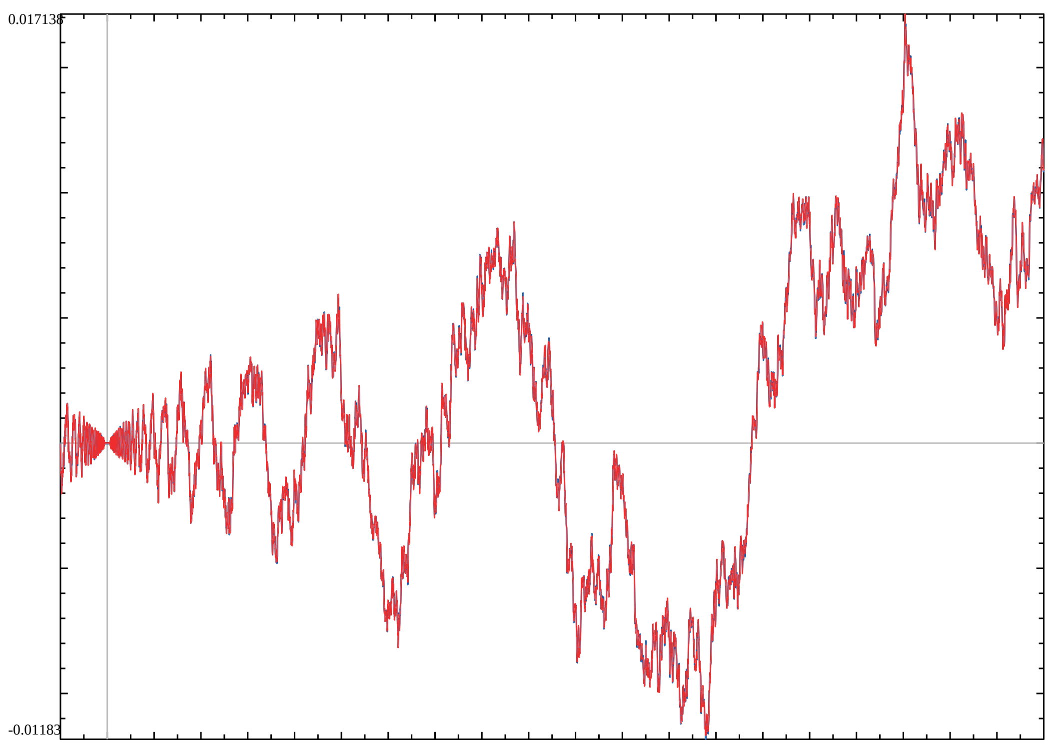
Finally, consider $x^4 - 2x^3 + 2x^2 + 2$, with the smallest field discriminant which is a square, and leads to a distilled motive: $D = 56^2 = 3{,}136$. It again has no real roots,[306] and (since the discriminant is a square) the Galois symmetries form the alternating group A_4. The graph of about one period of the real and imaginary parts shows that the plot has "extra symmetries" (as in the section on p. 55):

[306] The smallest square field discriminant $D = 163^2 = 26{,}569$ for the case with real roots $(x^4 - x^3 - 7x^2 + 2x + 9)$ is too large to hope to see patterns in the graphs.

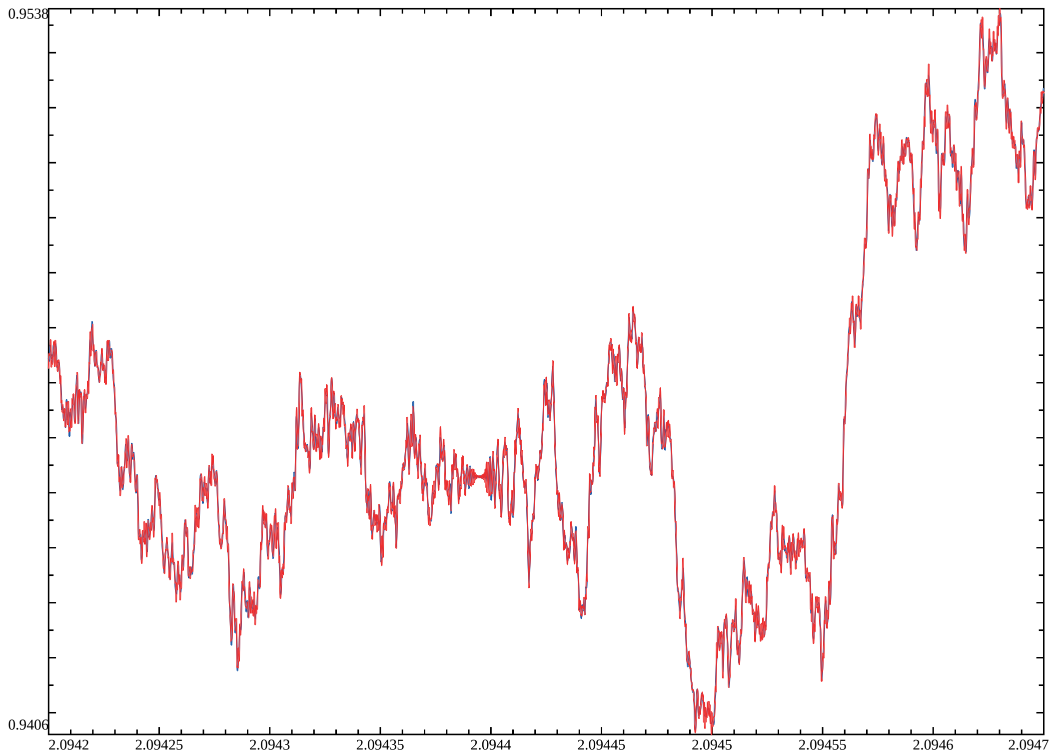
it is π-antiperiodic:



Here is how it looks near 0:



The graph near $t = \pi$ is the sign-flipped graph near $t = 0$. Instead, we plot the behaviour near $t = {}^{2\pi}/_{3}$, with a familiar “irregular hourglass”:



Distillation undoes “fusion”

Above (on p. 69),[307] we considered the “extra distillation” as “a black box”. While we gave an explicit recipe how to calculate the numbers N_n^{dist}, we did not explain either of

- What is the “reason” for using such an *ad hoc* recipe (except that “it works”: we got an exact fractal!).
- How could one “guess” that such a recipe may work.
- How these recipes are related to the rest of our discussion.

Before explaining these issues, recall the cases of distillation we have already dealt with. The idea is that in problems like “count the number $\widetilde{N}_{p^k}^{\mathrm{Gal}}$ of roots of a polynomial in modular arithmetic” the answer is often a sum of several “simpler parts”. However, eventually we want to translate these numbers to a fractal function $F(t)$; the process outlined in Footnote 179 on p. 74 goes in these steps:

$$\boxed{\widetilde{N}_{p^k}^{\mathrm{Gal}}} \xrightarrow{\text{“exp”}} \boxed{\overline{N}_n} \xrightarrow{\text{simplify}} \boxed{N_n} \xrightarrow{\text{Fourier}} \boxed{F(t)}.$$

The first arrow involves exp, so instead of $(\widetilde{N}_n^{\mathrm{Gal}})$ “breaking into a sum of simpler parts”, $(\overline{N}_n)$ is obtained from the corresponding “simpler parts” $(N_n^{[\alpha]})$ by a more complicated procedure. Below, we call this procedure *fusion*. For example, the *fusion* of sequences (A_n) and (B_n) calculated at[308] powers p^k, $k \geq 0$, with a prime p is the sequence

$$1, \quad A_p + B_p, \quad A_{p^2} + A_pB_p + B_{p^2}, \quad A_{p^3} + A_{p^2}B_p + A_pB_{p^2} + B_{p^3}, \ldots$$

[307] **N.B. (???) Initially, this was following the section on the $\mathbf{D}_4$-case. Now this needs to be restated.**

[308] From these powers, one extends to arbitrary n by multiplicativity (same as in Step (e) on p. 60).

To see the pattern, assuming $A_1 = B_1 = 1$, rewrite this using these "invisible factors" getting A_1B_1, $A_pB_1 + A_1B_p$, $A_{p^2}B_1 + A_pB_p + A_1B_{p^2}$, $A_{p^3}B_1 + A_{p^2}B_p + A_pB_{p^2} + A_1B_{p^3}, \dots$ (As expected — due to exp — this matches multiplication of power series[309] with coefficients A_{p^k} and B_{p^k}.)

In the beginning of this chapter we mentioned that if "the hidden symmetries" of these simpler parts are "unrelated", then after fusion these symmetries hide each other and become undetectable. So if we want to *find the "hidden symmetries"*, we need to identify "its simpler parts".

Assumption: Below, we always assume that these parts have "hidden symmetries" described by the Langlands program. It turns out that this completely identifies such parts.

In many situations considered in these notes one of the parts may be easily found, so the question becomes

$$\boxed{\text{Suppose we know one of these "simpler parts". Identify "what remains".}}$$

This follows the motto from Remark 45 on p. 75: "remove from the counts $\widetilde{N}^{\mathrm{Gal}}_n$ all traces of simpler polynomials".

We already met several cases where one of the parts of $\overline{N}^{[P]}_n$ matched the sequence $N^{[Q]}_n$ (or maybe $\overline{N}^{[Q]}_n$, see below) assigned to another polynomial sequence Q_m:

- $\boxed{\text{Naive}}$ For any polynomial sequence P_m, its $\overline{N}^{[P]}_n$ *always* "contains as a part" the sequence $\overline{N}^{\mathrm{point}}_n \equiv 1$ for the polynomial $Q_m \equiv P^{\mathrm{point}}_m \equiv m$.

  "Unfusing" this part is exactly the distillation process of Step (b) on p. 60 resulting in $N^{[P]}_n$. As in Footnote 180 on p. 74, this may be restated: $\overline{N}^{[P]}_n$ is a fusion of $\overline{N}^{\mathrm{point}}_n$ and $N^{[P]}_n$.
- $\boxed{\text{Reduc}}$ If the polynomial P is reducible $P_m = Q_mR_m$, then the corresponding sequence $\widetilde{N}^{\mathrm{Gal}}_n$ (obviously) breaks into a sum, hence $\overline{N}_n$ is a fusion of two parts $\overline{N}^{[Q]}_n$ and $\overline{N}^{[R]}_n$.[310]
- $\boxed{\mathsf{D}_4}$ If $\deg P = 4$ and P is of Galois type D_4, then there a quadratic polynomial $Q_m \equiv m^2 - \Xi$ such that $N^{[Q]}_n$ "works as a part of" $N^{[P]}_n$. (Hence $\overline{N}^{[Q]}_n$ "works as a part of" $\overline{N}^{[P]}_n$.)

The key fact is that the formulas for the fusion show that given a sequence (A_n) and the fusion of (A_n) and (B_n), it is easy to find (B_n) (if it is multiplicative[311]). Hence finding "what remains" in the framed question above is in no way complicated. **Conclusion:** if we know Ξ from the case $\boxed{\mathsf{D}_4}$ above, it is easy to find such "what remains" (since $N^{[Q]}_n$ and $\overline{N}^{[Q]}_n$ are given by Quadratic Reciprocity). Below, we call it (N^{dist}_n).

Essentially, this means that to explain the fractal symmetries of the plots in the preceding section one needs to show that:

- The case $\boxed{\mathsf{D}_4}$ above holds.
- For $P_m \equiv m^4 - m^3 - m^2 + m + 1$, the constant Ξ is -3.
- In this case "what remains" (N^{dist}_n) has "hidden symmetries 'of rank 2'" (i.e., the same fractal symmetries as in the non-abelian cases of degree 3).[312]

However, before we can explain these phenomena (in the section on p. 122), we need to introduce a unified approach to *all the cases* of sequences N_{p^k} we have met.

Grand Unification I: Denominators in Weil Conjectures

In (c), (d) on p. 60, in the section on p. 62, and the section on p. 68, we saw 7 different cases for the sequences N_{p^k} constructed for polynomial sequences P_m of degree 2, 3 and 4 leading to "fractal

---

[309] We have already mentioned this in Footnote 180 on p. 74.

In particular, one can boil down (a part of) the definition of N_n to $1 + \sum_k \overline{N}_{p^k}u^k = (1-u)(1 + \sum_k N_{p^k}u^k)$.

[310] Hence it is also a (repeated) fusion of $N^{[Q]}_n$ and $N^{[R]}_n$ with two copies of $\overline{N}^{\mathrm{point}}_n$.

Moreover, this may be restated as $N^{[P]}_n$ being a fusion of $N^{[Q]}_n$, $N^{[R]}_n$, and $\overline{N}^{\mathrm{point}}_n$. (We discuss this in detail in the section on p. 63.)

[311] Defined in Footnote 149 on p. 62.

[312] Recall that so far we did not show any *explanation* of existence of fractal symmetries even when $\deg P = 3$.

symmetries”. For particular values of p and particular polynomials P, we saw the sequences in the first column of this table:

Sequence	How to extend	Series of	Or	d
$-1, 0, 1, -1, 0, 1, \ldots$	3-periodic	$\frac{1}{1+u+u^2}$	$\frac{1-u}{1-u^3}$	3
$0, 1, 0, 1, 0, 1, \ldots$	2-periodic	$\frac{1}{1-u^2}$	$\frac{1-u}{(1-u)(1-u^2)}$	1, 2
$2, 3, 4, 5, 6, 7, \ldots$	a linear function	$\frac{1}{(1-u)^2}$	$\frac{1-u}{(1-u)^3}$	1, 1, 1
$0, -1, 0, 1, 0, -1, \ldots$	4-periodic	$\frac{1}{1+u^2}$	$\frac{(1-u)(1+u)}{1-u^4}$	4
$-1, 1, -1, 1, -1, \ldots$	2-periodic	$\frac{1}{1+u}$	$\frac{1-u}{1-u^2}$	2
$1, 1, 1, 1, 1, 1, 1, \ldots$	1-periodic	$\frac{1}{1-u}$	$\frac{1-u}{(1-u)^2}$	1, 1
$0, 0, 0, 0, 0, 0, 0, \ldots$	1-periodic	$\frac{1}{1}$	$\frac{1-u}{1-u}$	1

What is common between these sequences is that if one adds 1 in front, they are Taylor coefficients at 0 of very simple rational functions (indicated in the third column). In the fourth column, we rewrite them with the numerator $1-u$ (responsible for “distillation”),[313] and the denominator being a product of terms $1-u^d$, with the list of numbers d given in the right column.

We start our process of unification by reminding that for the cases of $\deg P = 2, 3$ and a “non-exceptional” primes p, we already described how to choose a row of this table (or, what is the same, choose the list of numbers d) in Steps (c)) and (d) on p. 60 and in the section on p. 62. It turns out that they may be restated in a uniform way:

The numbers d are degrees of the irreducible factors of the reduction $\bmod p$ of the polynomial P.[314]

Note that here we *must ignore* the multiplicity: if a certain factor of P appears many times, we *still* count it once.

---

[313] This numerator is always going to be eventually cancelled, since all the possible factors of denominators discussed below are divisible by $1-u$. (The fourth row appears only in D_4 case, so we write it with the “extra distillation” factor $1+u$ in the numerator.)

[314] For P of a small degree and an “exceptional” prime p, the same rule holds after a suitable “improvement” of the polynomial P consisting of a variable change $n \mapsto \alpha m + \beta$ and/or multiplying the polynomial by a constant in $\mathbb{Q}$. The aim is for the “improved” version to have as many distinct roots $\bmod p$ as possible.

For example, $P(n) := n(n-p) - p^3$ has a double root $n = 0 \bmod p$. Plugging in $n = pm$ and considering ${}^{P}\!/_{p^2} = m(m-1)$ “splits” this double root $\bmod p$ into two $m = 0, 1 \bmod p$. (So these roots contribute two numbers $d = 1, 1$.) Note that this would not work for $n(n-p) - p$: one cannot split *this* double root. (So this double root contributes one number $d = 2$.)

(To be honest, we must mention that for polynomials of higher degree the procedure for “improving” may be more involved than what is suggest above: for example, splitting different multiple roots may require different transformations: try to do the same for $P(n)P(n-1)$. I’m not even sure that nowadays it is known how to proceed in the case of general systems of polynomial equations![315])

[315] Judging by the answer of Matthew Emerton on 2010-07-29 in the discussion *Zeta Functions: Dedekind Versus Hasse-Weil* in **n-Cat Café**, with the approach of “point counting” we use in these notes is not known how to deal with “exceptional” primes in the general case; the only known cases are when the dimension of the set of solutions of our polynomial equations is 0 or 1 (and then the genus of a curve must be ≤ 1). (At least in absense of Hironaka resolution in positive characteristic.)

As a substitution for these missing definitions, the current approach needs to go through a certain “arithmetic theory of cohomology”. See also the answer of 2010-08-14, and the article posted by Minhyong Kim.

For example, a polynomial P of $\deg P = 3$ has no roots $\bmod p$ iff its reduction $\bmod p$ is irreducible; hence and p a red prime, the denominator is $1-u^3$, leading to the first sequence. In the same situation, having only one root $\bmod p$ (with multiplicity 1) leads to $(1-u)(1-u^2)$, giving the second case; the third case corresponds to having 3 distinct roots $\bmod p$ (leading to $(1-u)^3$). In the fifth case the denominator $1-u^2$ means (for $\deg P \leq 3$) that $P \mod p$ is an irreducible quadratic polynomial. The last but one case with $(1-u)^2$ may appear either for a polynomial of degree 2 with 2 roots, or a polynomial of degree 3 with a double root $\bmod p$ (it automatically has another simple root). The last case matches $1-u$, which means a single multiple root of multiplicity equal to the degree.

Finally, in degrees up to 7 the denominator $1-u^4$ of the fourth case may appear only when $P \mod p$ is an irreducible polynomial of degree 4. However, the "extra factor" $1+u$ in the numerator "makes sense"[316] only in the case D_4. (In the context of this section, we cannot explain this!)

For general polynomials of degree 4 which "do not allow an extra distillation" one would need to add more rows to our table — and one of them would have the same denominator $1-u^4$ in the fourth column, but no "extra factor" $1+u$ in the numerator. As we saw, for these polynomials our methods are not enough to expose their "hidden symmetries" (which are of "rank 3 which is too high" for our methods).

Remark 63: The denominators (denote them as $\Delta_p^{[P]}$) in the third columns are examples of *Weil denominators*[317] from Weil Conjectures, see Remark 43 on p. 74). The general case of Weil conjectures characterizes possible factors of the rational function associated to p and P in more complicated situations, when the polynomial equations $P=0$ depend on several variables, and there may be more than 1 equation. (Still, the corresponding numbers N_{p^k} are the Taylor coefficients of these rational functions.)[318]

Additionally, denote by $\overline{\Delta}_p^{[P]}$ the denominator in the fourth column (matching what happens before distillation). When P is clear from the context, we may omit the subscript.

Grand Unification II: Permutation matrices and Galois symmetries

Note that for a sequence to have a linear recurrent relation (with constant coefficients) is *equivalent* to being Taylor coefficients of a rational function. So what we described above is a significant refinement of Remark 43 on p. 74: *here* we claim that this rational function may be represented as a product of very simple terms.

To further demystify the denominators above, note that they are characteristic polynomials $\det(1-u\mathscr{M}_p)$ of certain permutation matrices $\mathscr{M}_p$.[319] The powers d appearing above form the *cycle decomposition* of this permutation. Moreover, one can recognize these cycles as orbits of this permutation.

Remark 64: Until the section on p. 122 all we care about $\mathscr{M}_p$ is that the sizes of cycles of this permutation coincide with the degrees of irreducible factors of $P \mod p$. Indeed, for non-exceptional primes the degree of the denominator is equal to the degree of the polynomial, hence the permutation with such cycle lengths may be thought of as a permutation of the roots of the polynomial![321]

---

[316] . . . meaning that it is a part of "the extra distillation" which leads to fractal symmetries of the corresponding function $F_{\text{dist}}(t)$.

[317] . . . or "Weil factors", since in the case of many unknowns such expressions may appear in the numerator too.

[318] **N.B. (???) Mention cohomology?**

[319] . . . which are matrices with exactly one non-zero entry (equal to 1) in every row and column. The importance of this description of numbers N_{p^k} lies in it being compatible[320] with the description in Footnote 179 on p. 74. These denominators, essentially, compress the sequence of numbers $\overline{N}_{p^k}$ of solutions into a very compact form!

[320] This immediately follows from Footnote 321 on p. 117 together with an appropriate generalization of Fermat's Little Theorem.

[321] Moreover, take into account that one can identify the complex roots of P with the roots modulo p (in an appropriate Galois extension). Factoring $P \mod p = Q_1Q_2\dots Q_r$, the modular roots of P separate into roots of Q_1, of

We postpone the more detailed discussion of Galois symmetries until the section on p. 131.[322]

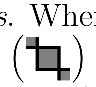
Grand Unification III: From reducible polynomials to “distillation”

The permutations of the preceding section are examples of a general construction: given a polynomial equation, one can assign to a prime p a particular permutation[324] of the roots of P. If $P = P^{[1]}P^{[2]}$, then this permutation of the roots of P coincides with “joined together” permutations assigned to $P^{[1]}$ and $P^{[2]}$.

In particular, this implies that the Weil denominator $\overline{\Delta}_p^{[P]}(u)$ assigned to P is the product of Weil denominators assigned to $P^{[1]}$ and $P^{[2]}$. Since $(\overline{N}_{p^k}^{[P]})$ are Taylor coefficients of $1/\overline{\Delta}_p^{[P]}(u)$, this implies

$$\boxed{\text{The sequence } (\overline{N}_{p^k}^{[P]}) \text{ is the fusion of the sequences } (N_{p^k}^{[P^{[1]}]}),\ (N_{p^k}^{[P^{[2]}]}).}$$

Likewise, if P is written as a product of L factors, this separates the roots of P into L flavors (“roots of $P^{[1]}$”, “roots of $P^{[2]}$”, etc.). The permutations considered above preserve the flavors, which describes $(\overline{N}_{p^k}^{[P]})$ as a fusion of several simpler sequences assigned to simpler polynomials.[325]

However, in 1920, Emil Artin understood that there is another way to look at the “flavoring” of the roots. In terms of the section on p. 117, instead of a collection of permutations, one should consider the corresponding permutation *matrices*. When the permutations preserve the flavors, the permutation matrices *are broken into blocks*, as in  (with 0s outside of the blocks; these 3 blocks match a flavoring into 3 flavors). Moreover, the sizes of these blocks do not depend on the choice of the permutation from our collection.

Q_2, etc. It turns out that the permutation in question has roots of Q_1 as one orbits, roots of Q_2 as another orbit, etc. (Compare with the motto from Remark 69 on p. 123. For details, see the construction of the Frobenious permutation in the section on p. 131.)

[322] Note that the non-real complex roots come in complex-conjugate pairs; this gives one (very trivial!) symmetry of the roots. What Galois discovered is that it makes sense to define other symmetries as well — nowadays we call this *Galois group*. Every such a symmetry permutes roots in a certain way.

(Formally speaking, a permutation σ of roots $x_1, \dots, x_d$ is a *Galois permutation* if any polynomial *relation* $\mathscr{P}(x_1, \dots, x_d) = 0$ *between roots* with integer coefficients continues to hold after the permutation; in other words, $\mathscr{P}(x_{\sigma_1}, \dots, x_{\sigma_d}) = 0$. *Galois theory* turns this clumsy definition into one of the most important tools for studying roots of polynomials equations.)

The relation to our denominators is that for every non-exceptional prime p, there is a particular Galois permutation of the roots whose permutation matrix is exactly as described above. This symmetry is named *the Frobenius element* for p (well, we cheated a bit: this symmetry is defined just “up to rotating it by other symmetries”; in other words, it is just a *conjugacy class* of a symmetry — but exceptional primes require more work — see Footnote 408 on p. 134).[323]

[323] The situation is similar (but not exactly the same) for many unknowns (and, maybe, many equations). The reason for the differences is that in our case two different complex roots “cannot be equal mod p” (whatever this means) for $p \gg 0$.

This can worked around by considering the action of Frobenius symmetries not on individual roots, but on “families of roots”. Moreover, while the geometry of families is very complicated, this complexity can be avoided by replacing families by their “algebro-topological footprints” in a suitable cohomology theory. While the intermediate steps are very complicated, the net result is as simple as above: a matrix is assigned to every prime number (but this time, it is not necessarily a permutation matrix). Instead of using sizes of orbits of the permutation as numbers d, one takes the characteristic polynomial of this matrix (for permutation matrices the latter holds the same information as the former) as the denominator we considered above.

For 1 equation with 1 unknown, only 0-dimensional families of roots may appear.

[324] One may need to put some fine print here — but it does not affect what we do in this section. Compare with Footnote 413 on p. 134.

[325] This is the $\boxed{\text{Reduc}}$ case of the section on p. 114. We illustrate it in the section on p. 63.

On the other hand, if we allow making a certain *change of basis* (what is critical is that this basis is the same for all the permutations we consider), such a "block decomposition" may be possible[326] even for certain irreducible polynomials P. This leads to another description of what we call "distillation":

Choose a block, and replace a permutation matrix by this block of this matrix (in the new basis).

This description is much more general than one in the section on p. 75, as well as complete and precise.

Now given a block $\mathscr{M}_p^{[l]}$ of the matrix $\mathscr{M}_p$ assigned to p, one can consider its characteristic polynomial $\Delta_p^{[l]}(u)$, and define $(N_n^{[l]})$ via the Taylor coefficients of $1/\Delta_p^{[l]}(u)$. The multiplicativity of characteristic polynomials of block-diagonal matrices implies

This presents $(\overline{N}_n)$ as a fusion of sequences $(N_n^{[l]})$.

For reducible polynomials, this works without any basis change, — but does not bring new insight: as before, this "fuses" the sequences $(N_n^{[P^{[l]}]})$ into the sequence $(N_n^{[P]})$ if $P = P^{[1]} \dots P^{[L]}$. However, the moment we allow a basis change, this gives us a new way to write $(N_n^{[P]})$ as a fusion — and (according to the Langlands program[327]) this is a fusion of *simpler* sequences. However, *this* is not very interesting[328]; the crucial part is that (at least conjecturally) these parts *also have "hidden symmetries"*!

Conclusion: What Artin discovered is that distillation allows splitting complicated problems of Number Theory into significantly simpler subproblems. According to the Langlands Program[329]

The distilled parts have "hidden symmetries" of simpler "types" than the initial problem.

In the beginning of this chapter we described two simplest types of "hidden symmetries": rank 1 and rank 2. Now we can enhance the framed rule claiming that these types corresponding to the 1×1 blocks and the 2×2 blocks in the rule. Using distillation, it is sometimes possible to split more complicated questions into "fusions" of problems with "hidden symmetries" of these "simple" types. We discuss what happens in higher ranks later (see the section on p. 169).[330] [331]

Notation: if the lth block of the matrix has size $d \times d$, we say that the corresponding hidden symmetry is *of rank d*. (It is the minimal possible length of the linear recurrence relation in k satisfied by the subsequences $N_{p^k}^{[l]}$.)

Remark 65: "A fully distilled block" is one which cannot be distilled into subblocks of smaller ranks. (In other words, no change of basis inside this block would subblock-diagonalize this block — simultaneously for all the permutation matrices we consider.) The Langlands program focuses on the behaviour of "fully distilled" cases.

---

[326] However, sometimes one may need to use a basis with complex coefficients. Compare with Footnote 363 on p. 126.

[327] This is *the Artin case* of the Langlands program (i.e., the simplest case, of 1 equation with 1 unknown).

[328] **N.B. (???) Any sequence can be written as a fusion — and we discussed it before! Ref!**

[329] We discuss the relevant to these notes part in the section on p. 171.

[330] In these notes, we do not discuss *how to use* the hidden symmetries. However, it is clear that knowing the symmetries of the problem may be a significant help for solving the problem.

One of the most striking examples is the Fermat's Last Theorem. It has been proven (after almost 400 years of trying!) by reducing it to existence of "hidden symmetries" of cubic equations with 2 unknowns. When this part of the Langlands Program was verified (it is of rank 2, so is "relatively simple"), this gave a proof of the FLT.

[331] **N.B. (???) Reference to Manin–Panchishkin?**

Going historically, the Class Field Theory claims that the sequences $(N_n^{[l]})$ of rank 1 are periodic and totally multiplicative.[332] The Langlands Program shows[333] that such sequences of rank 2 have a Fourier transform with the same kind of fractality as what we investigated for non-abelian cubic polynomials. The Langlands Program *conjectures* that such sequences of any rank have a certain very specific "hidden symmetry" — but for rank above 2, the description of the corresponding symmetric function is much harder than our explicit construction of $F(t)$ "by taking the Fourier transform" of N_n.[334]

Remark 66: For the purpose of these notes, "*motives*" are just "imagined geometric counterparts of the blocks of matrices $\mathscr{M}_p$". When we can "split" the roots of $P = 0$ into L "independent parts" (in other words, we can factor $P = P^{[1]} \dots P^{[L]}$), this leads to block-diagonalization of matrices $\mathscr{M}_p$ into L blocks (after a suitable shuffle of the basis). Likewise, *we imagine* that *any* simultaneous block-diagonalization of matrices $\mathscr{M}_p$ in a suitable basis "matches" a certain "imagined splitting into of the set of roots into 'motives'".

We allow this abuse of language since these "parts" present a huge mathematical interest, and (as we pointed in the section on p. 75) while there are very particular definitions of the notion of motive, it is not yet completely settled down.

Example: the naive distillation

The action of a permutation matrix on a vector can be thought of as permuting *real weights* assigned to the permuted points. Note that these permutations preserve the *total weight*,[335] as well as preserve the vector subspace corresponding to weight assignments with the total weight 0. **Conclusion:** there is a basis such that in this base any collection of permutation matrices acting on d points splits into a $1{\times}1$ block, and a $(d-1){\times}(d-1)$ block.

Given an arbitrary polynomial P, this is applicable to the considered above permutations of roots of P. This splitting into blocks corresponds to our "naive distillation", described in Step (b) on p. 60 and in Footnote 179 on p. 74. **Conclusion:** the statement of the preceding section (combined with the arguments from the other "Grand Unification" sections) implies the presence of "hidden symmetries" of rank $d-1$ for the sequence N_n assigned to P.

For the convenience of the reader, we summarize these arguments here. Since we gave several independent recipes of the sequence N_n, we start with the recipe going via $\overline{N}_n$ from Footnote 179 on p. 74. Recall that by this definition $\overline{N}_n$ is the fusion of N_n with $\overline{N}_n^{[Q]}$ for $Q(x) := x$, which is $\overline{N}_n^{[Q]} \equiv 1$. (Indeed, the equation $Q(x) = 0$ has exactly ${}^{[Q]}\widetilde{N}_{p^k}^{\mathrm{Gal}} \equiv 1$ root in any Galois arithmetic; the Taylor series for logarithm shows that $\overline{N}_{p^k}^{[Q]} \equiv 1$. The rest follows from multiplicativity of $\overline{N}_n$.)

First, the $1{\times}1$ block above is (1), hence its characteristic polynomial is $\det(1 - u(1)) = 1 - u$. (So this indeed matches $\overline{N}_{p^k}^{[Q]} \equiv 1$ from the preceding paragraph!) Since fusion corresponds to multiplication of denominators $\Delta_p^{[l]}(u)$, *removing* the $1{\times}1$ block with characteristic polynomial $1 - u$ corresponds to putting $1 - u$ into the *numerator*.

---

[332] Note that total multiplicativity (together with T-periodicity — needed if the sequence is defined only for positive n) implies even-or-oddness. Indeed, $(-1 \mod T)^2 = 1 \mod T$ (together with $(1 \mod T)^2 = 1 \mod T$) show that $N_{T-1} = \pm 1 =: \varepsilon$, hence $N_{T-n} = \varepsilon N_n$.

[333] See the section on p. 171.

[334] For these descriptions to work, our collection of matrices should be related to a certain system of polynomial equations. In our examples, the matrices are permutation matrices of Galois symmetries of complex roots of a polynomial in one variable — or common blocks of such matrices after an appropriate change of basis.

Yet more generally, instead of considering blocks of permutation matrices, we may assign any invertible matrix $\mathscr{M}_g$ to any Galois symmetry g provided that this assignment is compatible with the composition of symmetries: $\mathscr{M}_{gg'} = \mathscr{M}_g \mathscr{M}_{g'}$. Such an assignment is called an *Artin representation*, and the claims above work for the corresponding sequences (N_n) as well.

[335] In other words, they have a common eigenvector (with the common eigenvalue 1).

Recall that according to the section on p. 117 we assign to p the permutation $\mathscr{M}_p$ of roots with cycles of lengths $d_1, \ldots, d_R$ where d_r are degrees of the irreducible components of $P \mod p$. What remains to be shown is the Taylor coefficients of $1/\det(1 - u\mathscr{M}_p)$ coinciding with $\overline{N}^{[P]}_{p^k}$.

Since $\det(1 - u\mathscr{M}_p)$ does not change when we reorder the columns and rows of $\mathscr{M}_p$ in the same way, we can assume that points in any cycle of the corresponding permutation are next to each other, hence $\mathscr{M}_p$ is a block matrix. Then $1/\det(1 - u\mathscr{M}_p)$ is a product with each factor corresponding to one cycle.

On the other hand, ${}^{[P]}\widetilde{N}^{\mathrm{Gal}}_{p^k}$ is a sum of ${}^{[P^{[l]}]}\widetilde{N}^{\mathrm{Gal}}_{p^k}$ if $P \mod p = P^{[1]} \ldots P^{[L]}$. Since going from $\widetilde{N}^{\mathrm{Gal}}$ to $\overline{N}$ involves exp, this implies that $\overline{N}^{[P]}_{p^k}$ is a product of $\overline{N}^{[P^{[l]}]}_{p^k}$. Comparing with the preceding paragraph, it is enough to consider the case when $P \mod p$ is irreducible (hence $\mathscr{M}_p$ corresponds to a cyclic permutation).

If $\deg P = d$, then $\mathscr{M}_p$ corresponds to a cyclic permutation of d points, so[336] $\det(1 - u\mathscr{M}_p) = 1 - u^d$. This reduces our proof to showing that in this case $\overline{N}^{[P]}_{p^k}$ is 1 when d divides k and is 0 otherwise—which is equivalent to ${}^{[P]}\widetilde{N}^{\mathrm{Gal}}_{p^k}$ being d when d divides k and being 0 otherwise. This means that P has no roots in the Galois field with p^k elements unless $d|k$, otherwise P has all its d roots in this field.

Since it is out of scope of these notes, all we will say about the last statement is that it is the most elementary fact in the theory of Galois fields. (Note that it implies that for a fixed p^k the counts ${}^{[P]}\widetilde{N}^{\mathrm{Gal}}_{p^k}$ of solutions are the same for all irreducible polynomials P of degree d with coefficients $\mod p$.)

Finally, we want to cover other recipes of ours for finding the sequence N_n.

Recall that in the discussion after the table in the section on p. 115 we already matched the sequences of Step (b) on p. 60 and of the section on p. 62 to suitable collections of numbers d_l. Moreover, we have shown that such a collection coincides with the degrees of irreducible components of $P \mod p$, and the numbers N_{p^k} coincide with Taylor coefficients of $(1-u)/\prod_l (1 - u^{d_l})$.

Comparing to the preceding discussion, one can see that our recipes for the cases $\deg P \leq 3$ give the same result as the recipe from Footnote 179 on p. 74. (However, recall that the former recipes are less specific than the latter in the case of exceptional primes.)

Remark 67: The 1×1 and $(d-1)\times(d-1)$ blocks considered above "work" with *every* permutation of the roots. Moreover, it is easy to see that this is *the only* decomposition into blocks[337] which is compatible *with every* permutation matrix.

In particular, for polynomials P of Galois type S_d (with $\deg P = d$)—for which by definition[338] every permutation appears as one of the permutations $\mathscr{M}_p$—no further splitting into blocks is

[336] This follows since $\mathscr{M}_p$ diagonalizes in the basis $(1, \zeta, \zeta^2, \zeta^3, \ldots)$ with $\zeta^d = 1$ with eigenvalues ζ.

[337] Recall that to specify blocks, we need to specify a basis and say which elements of the basis "go into" which block. However, here we mean not the uniqueness of the basis, but the uniqueness of the *span* of the basis vectors going into a particular block.

[338] … together with Chebotaryev's Density Theorem.

possible. For these polynomials,[339] [340] no “extra distillation” is possible, so the complexity=rank of the corresponding “hidden symmetries” cannot be decreased below $d-1$.

Remark 68: In fact, the permutations of the roots which appear as $\mathscr{M}_p$ always act transitively on the roots. Moreover, an “extra” distillation is possible iff they do not act *2-transitively*.

For example, for $n=4$ only the abelian polynomials and polynomials of type D_4 allow “extra distillations”.

The “extra” distillation: the case D_4

The preceding section shows what happens when every permutation of roots may appear as the permutations $\mathscr{M}_p$ for some (non-exceptional) prime p. While this is “the general case” (see Footnote 340 on p. 122), some polynomials P are exceptions. This is due to “asymmetries” between *pairs* of roots: some pairs (x_k, x_l) of roots are “in special position” comparing to the other pairs (see also Remark 68 on p. 122). Since the permutations in question are *Galois symmetries*, they must send any “special pair” to a “special pair”.

For example, for $P_m = m^4 - m^3 - m^2 + m + 1$ from the section on p. 68, given a root x define $x' := -1/x = (x-1)(x^2-1)$ — which turns out to also be a root. Moreover, $(x')' = x$. Call x' *the opposite root* to x; this breaks 4 roots into 2 pairs of “opposite roots”.[341]

One can immediately see that any permutation which sends “an opposite pair” to an opposite pair acts on 4 roots in exactly the same way as the symmetries of a square act on its vertices. This means that possible “symmetries of the roots” form either an abelian group, or the group D_4 (as P above does[342]). (In the other direction, in any abelian/D_4 case one can write a polynomial $x'(x)$ with rational coefficients which works as above.) Moreover:

All such permutation matrices taken together may be split into blocks of sizes 1, 1, and 2.

Indeed, the preceding section shows how to split into blocks of sizes 1 and 3. To further split the “distributions of weights” with the total weight 0, consider assignments of weights to 4 vertices of a square such that the total weight of every diagonal is 0. This is complemented by the vector $\boldsymbol{v}$ assigning weights 1 to “one diagonal”, and weights -1 to the other.

Conclusion: to get the sequence N_n^{dist} for P one needs to *cancel* the contribution of $\boldsymbol{v}$ into the 3×3 block considered above. Note that a permutation matrix which “preserves oppositeness” either multiplies $\boldsymbol{v}$ by -1 (when the permutation exchanges the diagonals) or preserves $\boldsymbol{v}$ (when the permutation “leaves every diagonal “in place”). Hence the corresponding $\Delta_p^{[l]}(u)$ is $1 \pm u$ in these cases.[343] [344]

---

[339] As van den Warden proved in 1931, a random polynomial (with integer coefficients) is of this type with probability 100%. In other words, if one randomly chooses a polynomial P with integer coefficients less than N in magnitude, the probability gets closer and closer to 100% as N grows.

(However, there are other “very reasonable” ways to choose a random polynomial, and these ways break the result above. For example, instead of restricting the magnitude of coefficients, one can restrict the magnitude of the discriminant of P; then this probability above is less than 100%.)

[340] **N.B. (???) Ref? Do not we need field discriminant? Malle’s conjectures?**

(Heuristically, to “explain” why in degree 3 the abelian case is much more rare than the non-abelian one, note that it happens when the discriminant is a square. — And the larger “a random number” is, the less probable that it is a square.)

For more detailed info, see Melanie Wood’s notes of 2014: (§11 for the results in the cubic case; it also contains many conjectures).

[341] Moreover, if $x_3 = x_1'$, then $x_{2,4}$ satisfy $X^2 + x_1^2(x_1-1)X - 1 = 0$.

[342] This happens when the roots of P can be expressed in terms of $\sqrt{a + b\sqrt{L} + c\sqrt{M} + d\sqrt{LM}}$ when at least 2 of b, c, d are $\neq 0$ and the external square root cannot be avoided.

[343] It turns out that for exceptional p one should add a possibility $\Delta_p^{[l]}(u) \equiv 1$.

[344] Note that this is the first place where we need more info about $\mathscr{M}_p$ than the lengths of its cycles.

When one "cancels" $\Delta_p^{[l]}(u)$, it is the same as to put it into the corresponding numerator. Hence, identifying every Galois symmetries by the corresponding symmetry of a square, one enhances[345] the table from the section on p. 115:

Symmetry of a square	Denominator	First distillation	"Extra" distillation	What remains	$N_{p^k}^{\mathrm{dist}}$, $k=1,2,\ldots$
Rotation by 0°	$(1-u)^4$	$1-u$	$1-u$	$(1-u)^2$	$2,3,4,5,\ldots$
Rotations by ±90°	$1-u^4$	$1-u$	$1+u$	$1+u^2$	$0,-1,0,1,0,-1,\ldots$
Rotation by 180°	$(1-u^2)^2$	$1-u$	$1-u$	$(1+u)^2$	$-2,3,-4,5,\ldots$
Mirror-through-a-midpoint	$(1-u^2)^2$	$1-u$	$1+u$	$1-u^2$	$0,1,0,1,0,1,\ldots$
Mirror-through-a-vertex	$(1-u)^2(1-u^2)$	$1-u$	$1-u$	$1-u^2$	$0,1,0,1,0,1,\ldots$

Here the first two columns follow the recipe of the section on p. 117 literally: ds in factors $1-u^d$ of the "denominator" are the sizes of the orbits of the symmetry acting on 4 vertices of the square[346] (as in the section on p. 117). Likewise for the last two columns: the sequence $N_{p^k}^{\mathrm{dist}}$ is the Taylor coefficients at $u=0$ of $1/(\text{What remains})$.[347] (In other words, the coefficients of the polynomial in "What remains" give the recursion relation for $N_{p^k}^{\mathrm{dist}}$.) The only difference with what we did in the section on p. 115 is that "What remains" is the result of dividing the "denominator" by *both* the "distillation" factors (while before, getting N_{p^k}, we only divided by the first distillation factor).

As in the section on p. 115, given a non-exceptional prime p one consider the corresponding permutation of the roots. We did not explain it before, but this "Frobenius element" is actually "a Galois symmetry"—and for us this means that it must preserve "the oppositeness". Hence it matches one of the rows of the table above[348] —hence the last column gives the corresponding sequence $N_{p^k}^{\mathrm{dist}}$.

Knowing $N_{p^k}^{\mathrm{dist}}$ for every p, multiplicativity describes[349] the sequence N_n^{dist}, hence its Fourier transform $F_{\mathrm{dist}}(t)$.

Remark 69: To compare with what we did in the section on p. 68, we need a more explicit description of the "extra distillation factor[350]" $1\pm u$. "Combine" the roots using the weights in $\boldsymbol{v}$; this gives $v := x_1 - x_2 + x_3 - x_4$ (assuming $x_1' = x_3$ and $x_2' = x_4$). Since any Galois symmetry of P sends $\boldsymbol{v}$ to $\pm\boldsymbol{v}$, the same holds for v. Hence any such symmetry preserves v^2. **Conclusion:** by Galois theory, $\Xi := v^2$ is a rational number.

Putting $Q_m := m^2 - \Xi$, any Galois symmetry of P either exchanges the roots $\pm v$ of Q, or keeps them in place. Moreover, the sign in the extra distillation factor $1\pm u$ for a particular prime p matches these two possibilies for $\mathscr{M}_p$. What allows to use this observation to find the sign in $\pm\boldsymbol{v}$ is

[345] Although the new table covers only the cases of non-exceptional p.

[346] … in other words, one looks where the powers of the symmetry send a particular vertex.

For example, if the mirror passes through a vertex, then the reflection keeps two vertices in place (giving two orbits of length 1, each contributing the factor $1-u$), and join the other two vertices into an orbit of length 2 (this orbit contributes the factor $1-u^2$). In terms of permutation matrices (as in the section on p. 117) these denominators are the characteristic polynomials of the matrix (by the same reason as in the preceding section).

[347] Since we consider $k\geq 1$, the leading coefficient 1 is omitted.

[348] Moreover, for an exceptional p, instead of one of the entries in the "What remains" column, one should choose a certain divisor of one of these entries. This describes "exceptional" sequences $N_{p^k}^{\mathrm{dist}}$ up to a finite number of choices: they may be as in the table, or $1,1,1,1,\ldots$ (for the divisor $1-u$), or $-1,1,-1,1,\ldots$ (for $1+u$), or $0,0,0,\ldots$ (for the divisor 1).

Note that the explicit recipe which divisor of which entry to choose is quite convoluted. Compare with Footnote 408 on p. 134.

[349] … up to a finite number of choices due to our lack of precise description of what to do for exceptional primes. Note that this is exactly the same ambiguity as in Step d on p. 60.

[350] This polynomial is always a factor of what is in the "Denominator" column.

This desciption is *really* needed because of the 3rd and 4th rows of the table. They show that this factor is not determined by the "Denominator" column.

the astonishing property of the Frobenius permutations; we illustrate this property by the "motto":

$$\boxed{\text{"The permutation of the roots of } P \text{ associated to a prime number } p \text{ does not depend on } P.\text{"}}$$

This motto needs to be "decoded"; below, we apply it to two polynomials P and Q. For example, this statement is void if the roots of P and Q are "independent"; however, if the roots of Q can be expressed as polynomial expressions in terms of roots of P, then the above property ensures that the roots of Q are permuted in exactly the same way[351] as these expressions in terms of the (permutated) roots of P.

Conclusion: the permutation matrix $\mathscr{M}_p$ sends $\boldsymbol{v} \mapsto -\boldsymbol{v}$ iff the permutation of two roots of Q associated with p is non-trivial. On the other hand, we know the cycle structure of the latter permutation; hence it is non-trivial iff $Q \mod p$ is irreducible. Hence

$$\boxed{\mathscr{M}_p \boldsymbol{v} = \left(\tfrac{\Xi}{p}\right)\boldsymbol{v}} \qquad \text{and} \qquad \boxed{\text{The "extra distillation numerator" is } 1 - \left(\tfrac{\Xi}{p}\right)u.}$$

(In fact, the second statement works[352] even for exceptional p.) For $P_m = m^4 - m^3 - m^2 + m + 1$ we already know that $x_2 + x_4 = -x_1^2(x_1 - 1) =: -t$ and $x_1 + x_3 = t + 1$, hence $\Xi = (2t+1)^2 = -3$. Since $\left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right)$ which coincides with p modulo 3, one reconstructs the rules of the section on p. 68.

Conclusion: the double-framed statement on p. 119 implies exact fractality of the plots in the section on p. 68.

Remark 70: The arguments of the preceding remark work with any polynomial of type D_4. Moreover, recall that the motto from Remark 45 on p. 75 "explained" distillation as "remove from the counts $\widetilde{N}_n^{\mathrm{Gal}}$ all traces of simpler polynomials". The context of that remark was about "the naive distillation", when we remove "the traces of the equation $x = 0$". However, the argument above also follows this motto: essentially, the preceding remark shows that two 1×1 blocks of $\mathscr{M}_p^{[P]}$ taken together coincide with two blocks forming the matrix $\mathscr{M}_p^{[Q]}$ — hence what we did was "removing the traces of the polynomial Q".

Moreover, one can immediately see that "blocks of $\mathscr{M}_p^{[Q]}$ appear as blocks of $\mathscr{M}_p^{[P]}$" happens when roots of Q can be written as[353] *linear combinations* with rational coefficients of 1 and of roots of P.

$F_{\mathbf{dist}}(t)$ and how to recover the sequence of colors

Recall that the initial motivation of these notes was the question: Given a polynomial sequence P_m, which primes p are "green", and which "red":

$$\boxed{\text{Which primes appear as factors of numbers } P_m\text{, and which not?}}$$

In the chapter on p. 33, to describe the "hidden symmetries" of this question when P is non-abelian of degree 3, we encoded these answers into numbers N_n and visualized the Fourier transform $F(t)$ of this sequence. Moreover, by definition the "color" can be recovered given N_p: it is red if $N_p < 0$, and green otherwise.[354] [355]

As the table above and Remark 69 on p. 123 shows, the case $\deg P = 4$ of type D_4 behaves very similarly: knowing N_p^{dist} and $p \mod 3$, we can find whether p divides one of numbers $P_m := m^4 - m^3 - m^2 + m + 1$. (For general P of $\deg P = 4$ and type D_4, instead of $p \mod 3$ one should

[351] We already used this argument in Footnote 321 on p. 117.

[352] The case $\left(\frac{\Xi}{p}\right) = 0$ is not in the table above, since it appears only for exceptional primes.

[353] **N.B. (???) Can one use Chebotarev's to show that if one characteristic polynomial divides the other, then such an inclusion holds?**

[354] Indeed, recall that "the naive" distillation leads to $\widetilde{N}_p^{\mathrm{res}} = \widetilde{N}_p^{\mathrm{Gal}} = N_p + 1$.

[355] Moreover, in the section on p. 72 we saw that when $\deg P = 3$, one can go in the opposite direction as well: the numbers N_p may be found if we know the "color" of p, and one more bit of information $\left(\frac{c}{p}\right)$ for a certain number c (the *conductor*). (The latter bit is a periodic function of p: it depends only on $p \mod 4|c|$.)

know $p \mod 4|\Xi|$ — which, by Quadratic Reciprocity, determines whether Ξ is a quadratic residue mod p. Here Ξ is defined in Remark 69 on p. 123.)

Indeed, for a prime p it is easy to see that $N_p = N_p^{\text{dist}} + \left(\frac{\Xi}{p}\right)$; hence $\widetilde{N}_p^{\text{res}} = N_p^{\text{dist}} + \left(\frac{\Xi}{p}\right) + 1$ are given by this table:

$\downarrow p \mod 3$ $\quad N_p^{\text{dist}}\rightarrow$	-2	-1	0	1	2
-1	$\times$	$\times$	**0**	**1**	**2**
0	$\times$	**0**	**1**	**2**	**3**
1	**0**	**1**	**2**	**3**	**4**

(Here $p \mod 3$ should be replaced by $\left(\frac{\Xi}{p}\right)$ for a general P of degree 4 and type D_4). Recall that each colored number $\widetilde{N}_p^{\text{res}} = N_p + 1$ counts elements divisible by p among any p consecutive elements of the sequence (hence determines the color!); moreover, $\times$ marks impossible combinations.[356]

In particular, the analogue of Remark 15 on p. 35 holds in the case D_4 as well: we can find whether a given prime number p may be a divisor of numbers P_m by inspecting $p \mod 4|\Xi|$ and a certain definite integral involving p and the function $F_{\text{dist}}(t)$. (Moreover, one can replace this generalized function by its antiderivative $F_{\text{dist}}^{(-1)}(t)$ which "makes perfect sense" as a function.)

Cubic reciprocity: Class Field Theory in degree 3

When we have been handling the non-abelian case of degree 3 and the case D_4 (pp. 59,68), we introduced "black box" treatments which encapsilated powerful theories (which we — more or less — explained in this chapter) into an easy-to-formulate format. Moreover, in the quadratic case (p. 62) we could explain all the details (except proofs!) of Quadratic Reciprocity — which allows a complete description of numbers N_n.

One can introduce a similar "black box" in any abelian cubic case (as in the preceding section[357]) as well. In fact, it is going to be much closer to the quadratic case: it wouild give *a formula* for N_p for a prime p. Due to this similarity, this is sometimes called *Cubic Reciprocity.*[358]

It starts with the *anzatz*[359] $N_p = \xi_p + \overline{\xi}_p$ with number ξ_p for prime p satisfying $\xi(\xi^3 - 1) = 0$. This formula determines each ξ_p uniquely up to complex conjugation. The principal result of Class Field Theory is

$$\boxed{\text{There is a periodic totally multiplicative sequence } \zeta_n \text{ which works as } \xi_n.}$$

In other words: *there is a choice* of the sign of the imaginary part of numbers ξ_p which makes this sequence periodic and totally multiplicative. (Note that the sequence ζ_n is uniquely defined up to *simultaneous* complex conjugation of all numbers ζ_n.)

---

[356] This assumes that P has integer coefficients. The general case works too if one excludes ps dividing the denominators of the coefficients (and the leading coefficient).

[357] **N.B. (???) Check "preceding section" throughout this one: should be "the section on p. 67" instead.**

[358] ... although this seems like a misnomer, since it focuses only on "hidden symmetries of rank 1". The "real" cubic counterpart of the Quadratic Reciprocity should better cover both the case of rank 1 and rank 2. — (And this *is what we are doing in these notes*!)

[359] ... in other words, with a trick which is going to be explained *at the end, by the result of calculations.*

Since $0 \le \widetilde{N}_p^{\text{res}} \le 3$ and $N_p = \widetilde{N}_p^{\text{res}} - 1$, the possibility of this anzatz is equivalent to the number $\widetilde{N}_p^{\text{res}} = \widetilde{N}_p^{\text{Gal}}$ of roots of $P \mod p$ cannot be[360] 2. (Indeed, $\xi = 0, 1, (\pm\sqrt{-3} - 1)/2$ give $N_p = 0, 2, -1$.)

[360] In turn, this follows from the most elementary facts of number field theory: if P gives an abelian (or normal) field extension over $\mathbb{Q}$, then for $P \bmod p$ the multiplicities of all the roots are the same. This implies that $\widetilde{N}_{p^k}^{\text{Gal}}$ divides the degree or is 0.

First, observe how close this statement is to Euler's formulation of quadratic reciprocity (p. 16); the only differences are:

- In the Euler case, one uses ξ_p instead of[361] $\xi_p + \bar{\xi}_p$.
- In the Euler case, the equation is[362] $\xi(\xi^2 - 1) = 0$ instead of $\xi(\xi^3 - 1) = 0$.
- What Euler required was multiplicativity w.r.t. multiplication by -1 (but it implies the general case, see p. 210).

(In fact, we show in Remarks 71, 73 on p. 126 that these conditions completely determine the sequence N_n from the preceding section.)

Consider the same issues from the "Artinian point of view": the permutation matrices $\mathscr{M}_p$ in the abelian case of degree 3 are necessarily *cyclic*: they are either identity, or matrices of the cyclic permutations (123) or (132). Since these 3 matrices are powers of the matrix for (123), they may be *simultaneously diagonalized* in the eigenvector basis of this matrix. This is the "discrete Fourier transform" basis $(1, \zeta^m, \zeta^{2m})$ with a fixed non-trivial cubic root[363] ζ of 1 and $m = 0, 1, 2$.

Conclusion:

- The motive of solutions of an abelian cubic equation splits into a motive of a point, and two complexly-conjugate motives of rank 1.
- These motives match decomposition of matrices $\mathscr{M}_p$ into three 1×1 blocks.
- (By Remark 73 on p. 127,) the sequences $N_n^{[l]}$ (with l indexing the non-trivial blocks) identify with the periodic totally-multiplicative sequences ζ_n, $\bar{\zeta}_n$.

This demonstrates that this splitting uncovers the "hidden symmetries of rank 1"—indeed, these blocks behave exactly as required in the beginning of this chapter.

Remark 71: Moreover, same as what we saw for Quadratic Reciprocity (p. 209), if we know the conductor c (which is the shortest period of ζ_n) *at least sometimes* the rest follows:

$$\boxed{\text{The framed condition above and } c \text{ uniquely determines the numbers } N_p \text{ if } c \text{ is a power of prime.}}$$ [364]

Let us show how this works when $c = 9$ (as it happens for P from the preceding section).[365]

Indeed, let ζ be any non-trivial cubic root of 1. Assume that $p \nmid c$ (one can show that otherwise $N_p = 0$). Then $N_p = \zeta^m + \zeta^{-m}$ if $\pm p \equiv_c 2^m$, with $m = 0, 1, 2$. Indeed, 2 is the primitive root mod 9 (since $2^3 \equiv_9 -1$ and $2 \not\equiv_9 -1$). The total multiplicativity and 9-periodicity imply that $N_p = \zeta_2^m + \zeta_2^{-m}$, hence all we need to show is that $\zeta_2 \not\equiv_9 0, 1$.

On the other hand, if $\zeta_2 \equiv_9 0, 1$, then N_p would not depend on $p \nmid c$. Finally, this would contradict Chebotarev's Density Theorem.

---

361 For general abelian P, one needs to replace $\xi_p + \bar{\xi}_p$ by a longer sum over $\deg P - 1$ algebraic conjugates of ξ_p. (In particular, the largest possible value of this sum is $\deg P - 1$—matching the maximal possible value of N_p.)

362 In general, the conditions on ξ_p can be replaced by "ξ is 0 or a root of 1". The sequences ζ_n which appear in Quadratic Reciprocity can be characterized by $\zeta_n \in \mathbb{R}$.

363 This gives an example where one *must* use complex coefficients to block-subdivide a collection of matrices $\mathscr{M}_p$.

364 In general, varying Q among abelian polynomials of degree 3, there are 2^{m-1} possible sequences $(N_n^{[Q]})$ with the given c; here m is the number of distinct prime factors of c. This is equivalent to the 3-part of $(\mathbb{Z}/c)^\times$ being a product of m cyclic groups. (And this follows from c being a product of distinct factors which are either 9 or a prime $p \equiv_3 1$. See Hasse's *Arithmetische Theorie der kubischen Zahlkorper auf klassenkorpertheoretischer Grundlage*, Mathematische Zeitschrift 31 (1930) pp. 565-582.)

(Indeed, it is easy to show that total multiplicativity implies $\zeta_n = 0$ if $(n, c) > 1$; this reduces the question to mappings $(\mathbb{Z}/c)^\times \to \mathbb{Z}/3$.)

365 The discriminant of this polynomial is 81, hence the preceding footnote implies that c is 1 or 9. In fact, the conductor cannot be 1 since due to the "Conductor-Discriminant Formula", the field discriminant is c^2, hence $c^2 > 1$ due to the Minkowski's theorem.

Remark 72: By multiplicativity of N_n, to find this sentence all we need to know is N_{p^k}. In this section we already covered the case $k = 1$. However, for abelian cubic P the general case follows immediately from the description of embeddings of Galois fields[366]:

$$\boxed{\text{If } P \mod p \text{ has } \zeta + \overline{\zeta} + 1 \text{ roots in } \mathbb{Z}/p\text{, then } \widetilde{N}^{\mathrm{Gal}}_{p^k} = \zeta^k + \overline{\zeta}^k + 1.}$$

Indeed, this is clear if all 3 roots (counting multiplicity) are in $\mathbb{Z}/p$ (i.e., $\zeta = 0, 1$). Otherwise all 3 roots are in the cubic extension of $\mathbb{Z}/p$, hence $\widetilde{N}^{\mathrm{Gal}}_{p^k} = 0, 3$ depending on $3|k$ — which coincides with the formula above.

Remark 73: Together with what we know about ζ_p, the formula of the preceding remark represents $\widetilde{N}^{\mathrm{Gal}}_{p^k}$ as a sum of 3 *periodic totally multiplicative* sequences. (Indeed, $\zeta_p^k = \zeta_{p^k}$.)

However, $\exp \sum_{k\geq 1} a^k u^k/k = \sum_{k\geq 0} a^k u^k$ (compare with Footnote 179 on p. 74). **Conclusion:** $\overline{N}_n$ is completely determined by the numbers ζ_p and is a fusion of 3 periodic totally multiplicative[367] sequences (ζ_n^m) for $m = 0, 1, 2$. (Here $0^0 =: 1$.)

[366] We already discussed this before Remark 44 on p. 75.

[367] The *total multiplicativity* of (ζ_n) does not extend to (N_n): it is ruined by the addition "inside" the fusion rules.

Appendix: Getting closer to the Langlands Program

In construction! (Lousy — and quite incomplete — exposition.)

More on the fractality laws in a reducible case

This continues[368] the discussion of the plots shown in the section on p. 63.[369]

In the language of the section on p. 75 and of Remark 65 on p. 119 "the corresponding motive is not fully distilled" — and the patterns corresponding to the factors are "overlayed on top of each other", contaminating these patterns.

In short: for a reducible polynomial the sequence of red/green colors *is a "mix"* of colors for the factors of this polynomial. Likewise for numbers N_k from the section on p. 59: they are determined by the corresponding numbers N_k^{quadr} for $4m^2 + 2m - 3 = 0$.

Recall that in the sections on p. 114 and p. 118, we introduced the reducible case $\boxed{\text{Reduc}}$ as one of the motivations of the notion of distillation. We *claimed* that distillation (or, in this particular case, factorization) simplifies the "hidden symmetries" a lot — but we did not provide the examples. Now we can give the example: factoring out $2m - 1$ changes the plots above to the plots for $4m^2 + 2m - 3$, — which we considered in the section on p. 62.

Conclusion: to see the results of factorization on the plots in the case of degrees of the factors $1 + 2$, compare the plots here to the plot in that section: fractality of Fourier transform is replaced by the periodicity of coefficients.

Remark 74: In fact, it can be shown that the values of $F^{(-1)}(t)$ "change *only* due to jumps". (In other words, $F(t)$ is a sum of δ-functions; or one can say that $F(t)$ is "an Eisenstein series".)[370] More precisely, $F^{(-1)}(t - 0) - F^{(-1)}(t_0 + 0)$ equals the sum of jumps of $F^{(-1)}$ between t and t_0 (if $t > t_0$).

However (as we said in Footnote 159 on p. 65), while the statement above is true, it is true in a very non-expected way: the sum should be taken not over *all the jumps*, but over *any one* of "two halves" of the set of jumps. Together with our description of the positions and heights of jumps, this leads to a very explicit formula[371] for $F^{(-1)}(t)$.

Remark 75: While reducible polynomials are (usually?) not covered by Langlands' approach,[373] it looks like the graph above is still an exact fractal. And in fact, the transformation $T \mapsto -1/13T$ (here $T := t/2\pi$) exchanges jumps of the first and the second type (two "halves" of the preceding remark); moreover, after multiplication by $\sqrt{13}$ (and taking into account the law for how δ-functions change

---

[368] **N.B. (???) Fix "the plots above" etc. to proper references.**

[369] **N.B. (???) The next four paragraphs duplicate what is at the end of that section.**

[370] As we said, the graphs *suggest* this. On the other hand, it is probably too naive to rely on visual appearance in detection of Eisenstein series. Observe that adding a term with a continuous $F^{(-1)}(t)$ would not influence "the general visual appearence" of the graph: the contribution of this term would be lost in all the "fractal noise" of the jumps in the graph.

[371] To do this, one needs to rearrange this sum *smartly*, since it is obviously not absolutely convergent; we discuss this in the section on p. 135. After this, we can describe $F^{(-1)}(t)$ as a certain infinite summation over jumps which:

- Converges "as a generalized function".
- Converges absolutely for all t except for "very rare pathological values" of t.[372]

Mathematically, such objects are described using *modular symbols*.

[372] We do not know what happens in these "pathological values".

[373] The functions $F(t)$ predicted by the Langlands program are "cusp form" — which are, in a certain very precise sense, "functions 'opposite' to Eisenstein series".

under coordinate transform) one can see that the formulas for jumps at the points of the first and the second type are also exchanged by this transformation.[374]

Together with Remark 74 on p. 128 (and 77 on p. 129), this explains the fractality law "on any one particular side" of $t = 0$. Moreover, the transformations $T \mapsto {}^{(aT+b)}/_{(13cT+d)}$ with $ad - 13bc = 1$ send points $2\pi\frac{R}{S}$ to points of the same type, and again, they are "compatible" (in the same sense as above) with the jumps of $F^{(-1)}$ (up to the sign $\left(\frac{a}{13}\right)$). This shows that the corresponding fractal transformations do not change the graph!

Conclusion: there are two descriptions of $F^{(-1)}(t)$ as a sum over jumps. The transformations $T \mapsto {}^{(aT+b)}/_{(13cT+d)}$ with $ad - 13bc = 1$ are compatible with each one of these descriptions. The transformation $T \mapsto -{}^{1}/_{13T}$ *exchanges* these two descriptions. Together, this shows that all points are "one-sided horizon-similar".

Remark 76: Note that Eisenstein series are *direct generalizations* of Euler's formulation of Quadratic Reciprocity. Indeed, essentially Euler's formulation claims that the corresponding numbers N_k^{quadr} form a periodic sequence.[375] In terms of $F(t)$, this means that[376] it is a sum of δ-functions, and in terms of $F^{(-1)}(t)$, this means that it is a locally-constant function. In other words: the variation of $F^{(-1)}(t)$ is described as a sum of (a finite number of) jumps (at points $2\pi{}^{R}/_{S}$ with certain denominators S — and, in fact, the magnitude of the jumps is proportional to the Legendre symbol $\left(\frac{R}{c}\right)$. Compare with the graph on p. 62).

For the Eisenstein series for $(2m-1)(4m^2+2m-3)$ above, the only thing which changes is that we allow jumps with *any* denominator S with $13|S$ (instead of $S = 13$ for $4m^2+2m-3$).

Remark 77: To have an honest exact fractality we need $F^{(-1)}(t)$ to match near 0 what "$F^{(-1)}(t)$ is near infinity" — but $F^{(-1)}(t)$ has a jump at 0. In other words, $F(t)$ has a δ-function singularity at $t = 0$. One can see that to preserve "the spirit and letter of the fractality law", we must ensure that $F(t)$ also has "a δ-function singularity at $t = \infty$". Since a Fourier series $\sum_n N_n \cos nt$ is a periodic function, and periodic functions do not behave like this, we need to add another term into our definition of $F(t)$:

$$F(t) = N_\infty \delta_\infty(t) + \sum_n N_n \cos nt$$

for a certain value of N_∞. Unfortunately, $\delta_\infty(t)$ makes no immediate sense in math.

To explain what δ_∞ may mean, recall the pictures of the absolute of Lobachevsky geometry from Remark 23 on p. 38. There the t-axis "bends" around the disk so that $t = -\infty$ comes next to $t = +\infty$. This way, the t-axis becomes a circle with 1 point removed from it (essentially, an arc of 360°). On one side of the removed point is the $t = -\infty$ end of the arc, on the other side is $t = +\infty$.

In other words:

$$\boxed{\text{the } t\text{-axis with the added point } t = \infty \text{ becomes a circle (usually named } \mathbb{RP}^1).}$$

Moreover, it makes sense to restrict a (generalized) function on the circle $\mathbb{RP}^1$ to a function on $\mathbb{R}$ — but the δ-function with support at the added point vanishes after such a restriction. Hence while "extending" a (generalized) function from $\mathbb{R}$ to the circle above, we may add an arbitrary multiple of $\delta_\infty(t)$ — as we needed to do above.

[374] However, in the transformation, one should take *the absolute value* of the factor T (or $1/T$) of the fractality law, same as we did in Remark 27 on p. 42. (This is the *Maass* case!)

[375] Compare with the section on p. 62.

[376] … with the fine print from Footnote 150 on p. 63.

With thus modified function $F(t)$, the fractality law we established to work *separately* "on *each of two sides* of every point $2\pi^{R}/s$" now works also "in a certain interval *containing* every given point $2\pi^{R}/s$". In particular, this includes the jumps at rational multiples of 2π.[377 378]

Remark 78: The problem with the plots above is that (usually?) the Langlands program focuses on the behaviour of "distilled" motives. *This* is why we needed to "distill" our sequence N_m for its Fourier transform to have[379] the "expected" fractal properties. For an irreducible polynomial P, the motive for $P(m) = 0$ is a mixture of a motive[380] of a point (recall that a point is a solution to[381] $m = \text{const}$; the corresponding N_m are all 1) with "what remains"; however, if $P = P_1P_2$, then this motive is a mixture of motives for $P_{1,2}$; if $\deg P_1 = 1$ and $\deg P_2 = 2$, then it is a mix of two copies of a motive of a point, and the "what remains" motive for P_2.

Our procedure of "distillation" would remove one copy of the point-motive; what remains is "a point" fused with "what remains" for P_2 — which is exactly the "undistilled motive for P_2"! So in addition to showing what happens for reducible P, the pictures above also show the result of our procedure applied to a quadratic polynomial $4m^2 + 2m - 3$, but without the step of "distillation".

Remark 79: Compare with the plot of the abelian case in degree 3 (on p. 67).

Above we mixed a motive of a point (zero equations with zero unknowns!) with a motive for a quadratic equation. They corresponded to two functions $\widetilde{N}_p^{\text{res}} \equiv 1$ and $F_p^{(-1)} \equiv N_p^{\text{periodic}}$ (for prime p). In the abelian case we were mixing two motives which are both periodic.[382]

Remark 80: As explained above, for $P_m = (2m-1)(4m^2+2m-3)$ we saw in the section on p. 135,[383] the corresponding "undistilled" denominator $\overline{\Delta}(u)$ is a product $\overline{\Delta}^{[2m-1]}(u)\overline{\Delta}^{[4m^2+2m-3]}(u)$; the former factor is (as for any linear polynomial) $1 - u$, and the latter is $(1-u)\Delta^{[4m^2+2m-3]}(u)$. (This follows immediately from the definition of our "naive" distillation; compare with the section on p. 120. Recall that $1 - u$ appears here as the characteristic polynomial of $1{\times}1$ matrix 1.)

Conclusion: $\Delta^{[P]}(u)$ coincides with $\overline{\Delta}^{[4m^2+2m-3]}(u)$, hence the plots above coincide with the plots of the antiderivative of the Fourier transform $\overline{F}_{4m^2+2m-3}(t)$ of the "undistilled" sequence $\overline{N}_n$ for the quadratic polynomial $4m^2 + 2m - 3$. So in addition to showing what happens for a reducible P, these

---

377 Due to the need for this modification, the fractality law for this function is different from what we considered before. *This* is why we needed to consider first the example with larger conductor.

Moreover, strictly speaking, the Langlands program does not cover reducible cases.

378 We needed to cheat in this discussion. As stated, the term $\delta_\infty(-1/t)$ would be killed by the factor $|t|$, and/or one won't be able to divide it by $|t|$.

We are saved by the fact that in 1-dimensional case covariant k-tensors have 1 component for any k — so in this regard they do not differ from scalars. What *does differ* is how they change under coordinate transform. With the transform $t \mapsto -1/t$ they are divided by $|t|^{2k}$. So if we assume that $k = \frac{1}{2}$ for $F(t)$, then the factor in our transform is not needed anymore — it is "absorbed into the geometric nature" of $F(t)$ — which becomes a ½-density. In this context, the discussion of δ-functions above makes perfect sense.

(Note that the fact that k may be fractional is *also* a special feature of the 1-dimensional case.)

379 … with the fine print from Footnote 150 on p. 63.

380 Here "a mixture of motives" is a geometric counterpart of a fusion of sequences $N_n^{[l]}$. Compare with Remark 66 on p. 120.

381 It also can be thought of as corresponding to 0 equations with 0 unknowns.

382 Compare with our discussion of a quartic polynomial with field discriminant 117 on p. 104. The corresponding motive is also a mix of two distilled motives. That time it is a "periodic" motive mixed with a "modular form" motive.

A similar "mixing" may happen for cubic equations with *two* unknowns (*elliptic curves*). Some of these curves have "extra" symmetries called "complex multiplication" — and then their motive splits in way very similar to examples above. (This is a very classical topic in math.)

383 **N.B. (???) Check the order of sections!**

plots also show[384] the result of our procedure applied to a quadratic polynomial $4m^2 + 2m - 3$, but without "the step of distillation".

Recall that since N_n for a quadratic polynomial is periodic (even or odd), its Fourier transform is a sum of δ-functions, hence $F^{(-1)}$ is a locally constant function.[385] This illustrates what kind of a major simplification is obtained by our step of ("naive") distillation: it replaces very complicated fractality pattern for $\overline{F}_{4m^2+2m-3}(t)$ (so for $\overline{N}_n$) to "just periodicity" for N_n.

In other words: without our "naive distillation", for $P_m := 4m^2 + 2m - 3$, the complexity would jump from the "rank=1" case to the much more involved "rank=2" case. Likewise, for the generic cubic polynomials, the complexity would jump from "easy-to-visualize" explicit fractality of the "rank=2" case to the (enormously more complicated) "rank=3" case.

Frobenius

In the section on p. 115 we gave a (partial) description of how the different variants of the sequences N_{p^k} correspond to "Galois" symmetries of the complex roots of the polynomial P. In short, the assignment above can be broken into two steps:

$$\boxed{\text{A prime number } p} \mapsto \boxed{\text{“Its Frobenius symmetries”}} \text{ and } \boxed{\text{“A symmetry”}} \mapsto \boxed{\text{A recursion relation}}.$$

We are not going to discuss the first step, except for noting that for a non-exceptional p, all the corresponding Frobenius symmetries lead to the same recursion relation. (The Footnote 408 on p. 134 discusses exceptional primes.) The *significance* of this step is that instead of infinitely many (non-exceptional) primes, we can consider the finite set of symmetries (or: certain permutations) of the complex roots.

For "abelian" polynomials P, these symmetries commute, and there are as many of them as the degree of the polynomial. For our sequences N_{p^k}, this means that, after suitable distillations (which decreases the "rank" from $\deg P$ to 1) $N_{p^k} = N_p^k$ (compare with Footnote 405 on p. 133). Every corresponding — generalized — function $F(t)$ is a finite sum of shifts of δ-functions.[386] [387]

Artin representations

[388]

People who have heard of Artin L-function can immediately recognize[389] that our numbers N_m are exactly the coefficients of this function (for our assignment of 2×2 matrices).[390]

Finally, recall that *in the simplest cases* this part of Langlands program is *already known*:

$$\boxed{F(t) \text{ has required fractal properties when } N_m \text{ are coefficients of an “uncomplicated” Artin } L\text{-function.}}$$

---

[384] **N.B. (???) Rewrite in terms of flavors of "hidden symmetries". Similar to what we do after plotting deg=2.**

[385] In fact, this is correct only for one of the real/imaginary part of $F_{\mathbb{C}}(t)$. See Footnote 151 on p. 63.

[386] Or at least this holds for either real or imaginary part of $F_{\mathbb{C}}(t)$. (See Footnote 151 on p. 63.)

[387] We discussed such distillations in the section on p. 67.

[388] **N.B. (???) This is a duplicate of the notes on p. 171.**

[389] In addition to what we did in the section on p. 115, one needs to check that the standard definition of *the Frobenius permutation* gives a 3-cycle if there are no roots mod p (the "red" primes), a transposition in the case of 1 root, and the trivial permutation in the case of 3 roots.

[390] Since our language is not good enough for a *general* description of what happens in *exceptional* primes, this does not verify the match if m is divisible by an exceptional prime. Still, in our particular case one can check such matches as well.

According to Langlands–Tunnell results (of $\approx$ 1980)[391] a case is "uncomplicated" if the matrices are 2×2, and it is not the "icosahedral" case: products of these 2×2 matrices do not match the composition laws of the symmetries of an icosahedron.[392]

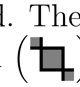
In terms of matrices, this means that in a certain basis these 4×4 matrices not only split into 1×1 and 3×3 blocks,[393] but also the 3×3 block may be split again into another 1×1 block and a 2×2 block.[394]

It turns out that the proper numerator is $(1-u)^2$ for odd rows of this table, and $1-u^2$ for even rows.

Random yet-unincorporated bits and pieces

Remark 81: As usual with distilled components, for prime indices "mixing the components" is just addition: $N_p = \zeta_p + \overline{\zeta}_p$. For composite indices the relations become more involved: for powers of primes one requires $N_{p^k} = \zeta_p^k + \zeta_p^{k-1}\overline{\zeta}_p + \ldots + \overline{\zeta}_p^k$ (unless $\zeta_p = 0$, this may be rewritten as $\zeta_p^k + \zeta_p^{k-2} + \ldots + \zeta_p^{-k}$), and for more complicated indices this quickly goes out of hand.[395]

Remark 82: This is very similar to many problems of Linear Algebra simplifying a lot by diagonalization of a matrix (when applicable). Indeed, the "permutation matrices" we associated to every Galois symmetry in the section on p. 115 can be all diagonalized in the same basis iff the polynomial P is abelian. Selecting a particular vector of the basis chooses a particular linear factor $1 - Z_p u$ of the characteristic polynomial $\Delta_p(u)$ of the permutation associated to any non-exceptional[396] [397] prime p. It is easy to see that $Z_p = \zeta_p$ or $Z_p = \overline{\zeta}_p$ (hence Z_p can be used as ζ_p described above) — and Class Field Theory shows that any sequence Z_p obtained this way is periodic![398]

Remark 83: When dealing with several matrices, it is very rare that one can diagonalize them all in the same basis — and this is why abelian polynomials are so rare. However, even if the full diagonalization is not possible, sometimes a "partial" diagonalization may be achived. Then there is a basis in which all the matrices allow the same "block-diagonal" decomposition, as in  (with 0s outside of the blocks).

Recall that for the "non-trivial part" of Class Field Theory[399] to hold, one needs a particular way to choose one of ζ_p and $\overline{\zeta}_p$ simultaneously for all p. It turns out that after splitting into 1×1-blocks, each such way to choose may be identified with a particular block (=eigenvector) via considering the characteristic determinant $1 - Z_p u$ of this block. In general, choosing the lth block (out of L blocks) chooses a particular factor $\Delta_p^{[l]}(u)$ of the characteristic polynomial $\Delta_p(u)$ for every non-exceptional p.

[391] In Knapp's notes in the Edinburgh Proceedings *Representation Theory and Automorphic Forms*, 1997, this is Theorem 8.9 (together with the paragraph after Theorem 8.7).

[392] The icosahedral case is also known, but only in the "even" case (for polynomials of degree 5, this means that they have 1 real root) since 2009. See Khare–Wintenberger paper *Serre's modularity conjecture. I*.

[393] These blocks correspond to the "total weight" of a weight distribution on roots, and the distributions with total weight 0.

[394] The second 1×1 block consists of weight distributions proportional to $1, -1, 1, -1$, and the last one to the *odd* weight distributions, for which the weights at opposite vertices are opposite.

[395] To avoid these complications, mathematicians use "a black box" encapsulation of the Dirichlet series. This way, the formulas for N_n given above are just particular cases of the statement that one Dirichlet series is a product of two others.

[396] If P is abelian, one can just force Z_p=0 for any exceptional prime.

[397] **N.B. (???) For degree 3 (or prime) only???**

[398] This is an amplification of the statement above about a possibility to choose a periodic ζ_p.

[399] **N.B. (???) The framed statement of periodicity.**

For prime p, the relation between these is immediate:[400] $N_p = \sum_{l=1}^{L} N_p^{[l]}$.

Remark 84: For rank above 2, the description of the corresponding symmetric function is much harder than our explicit construction of $F(t)$ "by taking the Fourier transform" of N_n.[401]

Remark 85: (In terms of numbers N_n, a prime p appears as a divisor iff[402] [403] $N_p \geq 0$.)

Remark 86: (In fact, according to Class Field Theory abelian polynomials exist in any degree, and the description above still works if we replace the conditions on ζ_n to be roots of 1 of degree $\deg P$, and instead of complex conjugation, consider all Galois symmetries of the roots. Note that with this formulation Quadratic Reciprocity becomes the case $n = 2$.)

Remark 87: The symmetry is the "periodicity".[404]

Remark 88: This "denominator" is a polynomial in an "ancillary" variable u.

Remark 89:
- The sequence N_{p^k} (for $k \geq 0$) is encoded as a Taylor series at $u = 0$ of a certain rational function.
- The denominator of this function is a product of the terms $1 - u^d$ with certain numbers d.
- The numbers d which appear are the degrees of the factors of the polynomial P reduced $\bmod p$. (Here one should ignore the multiplicity of every factor: $(m-1)^3(m^4 + 3m - 5)^2$ gives two numbers $d = 1, 4$. This gives $(1-u)(1-u^4)$.)
- The numerator is $1 - u$. (In fact, it is a common factor of all factors $1 - u^d$.)

Remark 90: On the other hand, this "extra distillation" was not following the motto "a contribution of simpler polynomials" above. The distilled "parts" are very simple, but they do not "match" any simpler polynomial.

Remark 91: (According to the Class Field Theory) the resulting sequence N_n (of rank 2) can be further distilled into two components (of rank 1 each).[405]

Remark 92: In the section on p. 102 we covered the flavors of non-abelian polynomials of degree 4; the principal difference is how large is the group of Galois symmetries: size 8 for D_4; size 12 for A_4; and size 24 for S_4.

---

[400] In general, one can express N_n in terms of $N_m^{[l]}$ with $m|n$ and $1 \leq l \leq L$. However, if n has many prime factors, this description becomes more and more cumbersome. In practice, it is much easier to work with *Dirichlet series* — see Footnote 395 on p. 132.

[401] For these descriptions to work, our collection of matrices should be related to a certain system of polynomial equations. In our examples, the matrices are permutation matrices of Galois symmetries of complex roots of a polynomial in one variable — or common blocks of such matrices after an appropriate change of basis.

Yet more generally, instead of considering blocks of permutation matrices, we may assign any invertible matrix $\mathscr{M}_g$ to any Galois symmetry g provided that this assignment is compatible with the composition of symmetries: $\mathscr{M}_{gg'} = \mathscr{M}_g \mathscr{M}_{g'}$. Such an assignment is called an *Artin representation*, and the claims above work for the corresponding sequences (N_n) as well.

[402] ... if P_m is a polynomial with integer coefficients and the leading coefficient 1. In general, to get from "integer values" to "integer coefficients" one needs to consider αP with $\alpha \in \mathbb{N}$ and $\alpha | (\deg P)!$. This means that the equivalence works only when $p \nmid \alpha$.

Likewise, if the leading coefficient is not 1, then one gets another finite set of exceptional p␣s.

[403] Moreover, with a finite number of exceptions (when $N_p = 0$), N_p takes only values ± 1. Hence up to these exceptions, finding N_p is equivalent to answering the framed question.

[404] **N.B. (???) Multiplicativity as rescaling?**

[405] Each of these $\deg P - 1$ components has an extremely simple form: the corresponding rational function is $1/(1 - C(p)u)$ with $C(m)$ a certain periodic function of m. Moreover, the sequence $(N_n^{[l]})$ corresponding to this component coincides with $C(n)$ for all n.

For $\deg P = 2$, there is only such component, hence the "whole" N_n is of this form ("Quadratic Reciprocity").[406] The corresponding C takes values in $0, \pm 1$. (In general, $C(n)$ is 0 or a root of 1.)

[406] **N.B. (???) Ref!**

Remark 93: The numbers d which appear are the lengths of cycles of a certain permutation ("Frobenius symmetry") of the roots "associated with a prime number p".[407]

Remark 94: Up to ("conjugation by") a symmetry of a square, all the symmetries of the square break into 5 flavors.

The Frobenius symmetry associated to p is in fact *only a flavor*: the built-in ambiguity of the construction of this symmetry allows *every* symmetry in a conjugacy class "to work as" Frobenius symmetry. Fortunately, in these notes we care only about properties of these symmetries which are shared by all the symmetries of the same flavor.[408]

Remark 95: To distill further, notice that out of the expected $4! = 24$ permutations only $|\mathsf{D}_4| = 8$ are Galois permutations; what decreases the number of "symmetries" may be thought of as "a certain structure" on the set of 4 roots.[410]

Remark 96:

The matrix assigned to the composition of permutations is a product of matrices assigned to two factors.

Summarizing:[412]

- To every non-exceptional prime number p we assign a particular Frobenious permutation.[413]
- The coefficients of this characteristic polynomial can be considered as coefficients of the recurrence relation.[414]
- Our numbers[415] $N_{p^k} =: a_k$ are defined by this recurrence relation. (We start with $a_0 = 1$, and $a_k = 0$ for $k < 0$.)

---

[407] This is literally true for non-exceptional primes p only (up to remarks in Footnote 322 on p. 118). The Footnote 408 on p. 134 discusses exceptional primes.

[408] As explained, this is applicable only to non-exceptional primes! In general, one should *"average this property"* over the ambiguity in the definition of the Frobenius symmetry.[409] For non-exceptional primes the ambiguity allows this symmetry to be any element of *one* particular flavor (in other words: of a certain conjugacy class). Then averaging "does not do anything", since the properties we consider are constant over a flavor. However, in exceptional cases this ambiguity gives *a union* of several flavors.

Surprisingly, an average as above is always the same kind of product as $\Delta(u)$. For example, in the case of a "most exceptional" p, when $P \mod p$ is a power of a linear function, *any* symmetry can be Frobenius. Then averaging "our property" $\Delta(u)$ (described above), and taking into account which flavors contain one symmetry, and which two, one gets ${}^1\!/\!{}_8((1-u)^4 + 2(1-u^4) + (1-u^2)^2 + 2(1-u)^2(1-u^2)) + 2(1-u^2)^2 = 1-u$. So the denominator associated to such p is $1-u$. (In turn, this denominator is going to be cancelled by the numerator $1-u$ of the first distillation.)

[409] The standard description is much more complicated. Nevertheless, it is equivalent to our description.

The major advantage of "the standard description" is that it explains the mystery in the last paragraph of the preceding footnote.

[410] **N.B. (???) Need to add: the permutation (12)(3)(4) of vertices is not induced by a symmetry of a square. Which are?**

Given any root x_1 of a polynomial $P(x)$, the long division algorithm shows that one can factor P as $P(x) = (x - x_1)Q(x)$ with the coefficients of the polynomial Q being rational expressions of x_1. Likewise, one can distinguish different *Galois types* of irreducible polynomials (with rational coefficients) by investigating their factors whose coefficients are allowed "to depend on" x_1.

For example, P is *abelian* iff Q (hence P) can be factored into linear factors with such coefficients. (This trivially holds if $\deg P = 2$.) Moreover, for polynomials of degree 4, one can distinguish all the non-abelian types[411] using similar recipes. For example, Q factors as "linear×quadratic" iff the type is D_4. To tell the remaining types A_4 and S_4 apart, one should allow rational expressions involving two different roots x_1 and x_2. The type is A_4 iff $x_{3,4}$ may be written as such expressions.

[411] To distinguish 2 different abelian types, one needs another approach. Look how many square-free numbers $d \in \mathbb{Z}$ can be represented as $d = R(x_1)^2$; here $R(x_1)$ is a rational expression of x_1. For the type $\mathbb{Z}/(4)$ there is one such d (the square-free part of the square root of discriminant); for the type $\mathbb{Z}/(2) \times \mathbb{Z}/(2)$ there are three.

[412] **N.B. (???) We keep only the items which mention the relevant footnotes.**

[413] Well, only a conjugacy class — but all permutations in a class have the same characteristic polynomial.

[414] For example, a polynomial $1 - 3u + 2u^2$ gives a recursion relation $a_k - 3a_{k-1} + 2a_{k-2} = 0$.

[415] ... from the section on p. 59.

Supplementary Musings: "Ghost jumps" in Eisenstein series

In construction! (Lousy — but complete — exposition.)

This part of the notes has very little to do with our principal aims. However, the situation uncovered in Remark 74 on p. 128 is so mind-boggling that I could not leave it alone, and was forced to write explanations which are somewhat more detailed (and way more complicated) than what is done in the rest of these notes.

Note that this part is in extremely preliminary stage, and was not optimized for reading in any way. So unless one is *really* interested in why sums of δ-functions behave in the way we claim in Remark 74 on p. 128, this "appendix to appendix" should better be skipped in the first few readings.

However, we want to stress that as calculations in Analytic Number Theory go, what we do below is completely pedestrian, and is very close to 0 on the "0 to 10" difficulty scale. What is really surprising is *the result*, and not the calculations themselves.

(This appendix is in a very early stage. It was not massaged yet in any way to simplify reading!)

Examples of dealing with Eisenstein series

Example of Eisenstein calculation:

The case $M=16$ of $M\times$ Tetrahedral number $+\,1$ is proportional to the polynomial $(2n-1)(4n^2+2n-3)$, which can be rewritten after the substitution $k:=2n-1$ as $k(k^2+3k-1)$. Applying our recipes above for the numbers N_m (see the section on p. 115) *literally* to this decomposable polynomial, one gets numbers N_{p^k}, $k\geq 1$, which are 2,3,4,5,… when $\left(\frac{p}{13}\right)=1$, or 0,1,0,1,… when $\left(\frac{p}{13}\right)=-1$, or 1,1,1,1,… when $p=13$ (here we use Legendre symbol from p. 208). (The latter case may requires the rule from the section on p. 115.) One can immediately see that $N_m=\sum_{d|m}\left(\frac{d}{13}\right)$ when $m=p^k$; since both sides are "multiplicative"[416] the same identity holds for arbitrary m. Consider a more general sequence $\sigma_m:=\sigma_m(s):=\sum_{d|m}\left(\frac{d}{13}\right)d^s$; here s is a real (or complex) parameter. With $s=0$, one gets N_m; we are going to consider negative s, then take $\lim_{s\to -0}$.

Our aim is[417] to calculate the Fourier transform of the sequence σ_m. Start with rewriting the condition $d|m$ as $\sum_{r \bmod d}\mathbf{e}(m\cdot {}^r\!/\!_d)=d$; otherwise the sum is 0. Here $\mathbf{e}(t):=\exp 2\pi i t$. Hence one can rewrite

$$\sigma_m=\sum_{d>0}\left(\frac{d}{13}\right)d^{s-1}\sum_{r \bmod d}\mathbf{e}(m\cdot {}^r\!/\!_d)$$

(note that the combined summation is absolutely convergent iff $s<-1$).

Grouping together terms with $\pm r \bmod d$, the Fourier transform is

$$\tfrac12\sum_{d>0}\left(\frac{d}{13}\right)d^{s-1}\sum_{r \bmod d}\sum_{m>0}\left(\mathrm{e}^{\mathrm{i}m(t+2\pi {}^r\!/\!_d)}+\mathrm{e}^{\mathrm{i}m(t-2\pi {}^r\!/\!_d)}\right).$$

In this form, the complex conjugation replaces summation over $m>0$ by $m<0$; hence taking the real part gives (on $[0,2\pi]$)

$$\tfrac14\sum_{d>0}\left(\frac{d}{13}\right)d^{s-1}\sum_{r \bmod d}\left(\sum_{m}\left(\mathrm{e}^{\mathrm{i}m(t+2\pi {}^r\!/\!_d)}+\mathrm{e}^{\mathrm{i}m(t-2\pi {}^r\!/\!_d)}\right)-2\right)=\sum_{d>0}\left(\frac{d}{13}\right)d^{s-1}\sum_{0\leq r<d}\left(\pi\delta(t-2\pi {}^r\!/\!_d)-\tfrac12\right).$$

---

[416] As in Step (e) on p. 60. See Footnote 149 on p. 62.

[417] In fact, we came to this calculation "going backwards": we took the conjectured formula for jumps in the function $F^{(-1)}(t)$ from p. 65, then calculated the Fourier transform of the derivative of a periodic function with such jumps, then (after we saw a match with N_m) inverted this process.

or, putting $\ell_s := \sum_u \left(\frac{u}{13}\right) u^s$ (absolutely convergent for $s < -1$)

$$\pi \sum_{d>0} \left(\frac{d}{13}\right) d^{s-1} \sum_{0\le r<d} \delta(t - 2\pi r/d) - \tfrac{1}{2}\ell_s.$$

Writing $d = Du$, $r = Ru$ with $u = (d, r)$, this becomes

$$\pi\ell_{s-1} \sum_{D>0} \left(\frac{D}{13}\right) D^{s-1} \sum_{0\le R<D,\ (R,D)=1} \delta(t - 2\pi R/D) - \tfrac{1}{2}\ell_s.$$

Periodic extention from $[0, 2\pi]$ gives

$$\pi\ell_{s-1} \sum_{D>0,\ R,\ (R,D)=1} \left(\frac{D}{13}\right) D^{s-1}\delta(t - 2\pi R/D) - \tfrac{1}{2}\ell_s.$$

For $s < -1$, everything was absolutely convergent, hence our calculations make perfect sense: the latter sum is the real part of the Fourier transform of the sequence σ_n.[418] Moreover, later we are going to show that ℓ_s and $\sum_{D>0,\ R,\ (R,D)=1} \left(\frac{D}{13}\right) D^{s-1}\delta(t - 2\pi R/D)$ extend as analytic functions to $\operatorname{Re} s < 1$, and that this implies that our formula for Fourier transform is valid for such values of s (if one reorders the summation above as described below). Moreover, since $\ell_0 = 0$, for $s = 0$ the last term disappears.

Conclusion: the real part of the Fourier transform of σ_m (in other words, the sum of the Fourier series) is the sum of δ-functions with non-0 coefficients concentrated in all rational multiples of π with denominators prime to 13. From this, it is very natural to expect that the antiderivative has jumps at these numbers, and the height of such a jump is equal to the coefficient at the corresponding δ-function. (Note that this predicts the correct jump $\pi\ell_{-1} \approx 2.08$ at 0.) In particular, the jump at $2\pi/13$ would be 0.

However, on the graph on p. 65 we saw that at $2\pi/13$ *there is* a jump!

Note that for $s < -1$ our series converge absolutely, hence manipulations make perfect sense. *One might have assumed* that since σ_m and the coefficients at δ-functions in the final answer depend analytically on s, they should match for any s. However, this is not how analysis works; fine print[419] in theorems on analytic dependence on paramaters breaks the match.

Indeed, if one believes the calculation above gives a correct answer for $s = 0$, then the coefficient at δ-function at $t = 2\pi/13$ should be 0; however, the graph of antiderivative on p. 65 has a non-trivial jump there.

There is a lot to say about $s \ge -1$.

To show that the heuristic argument above *must break*, consider a different approach to the same Fourier transform. The explicit description of N_{p^k} for this polynomial implies that $N_{p^k} = 0$ if $p \ne 13$ and $\left(\frac{p^k}{13}\right) = \left(\frac{p}{13}\right)^k$ is not 1. Therefore $N_m = \left(\frac{m'}{13}\right) N_m$; here we write $m = m'13^k$ with $(m', 13) = 1$. This leads to a different sequence $\widetilde{\sigma}_m := \widetilde{\sigma}_m(s) := \left(\frac{m'}{13}\right) \sum_{d|m} \left(\frac{d}{13}\right) d^s$ with the same limit[420] N_m when $s \to -0$.

To calculate the Fourier transform, rewrite the factor $\left(\frac{m'}{13}\right)$. Let $\rho_l(m) := \sum_{v \bmod 13^l} \left(\frac{v}{13}\right) \mathbf{e}(m \cdot v/13^l)$ for $l \ge 1$. We claim that $\left(\frac{m'}{13}\right) = \frac{\sqrt{13}}{13^l}\rho_l(m)$ when $l = k + 1$, and that the RHS is 0 otherwise.

---

[418] (In fact, one can consider our summation of δ-functions even in the space of measures, and not generalized functions. — Recall that summation — or taking limits — in generalized functions is much "more forgiving" than in measures. (For example, consider $\lim n(\delta(t - 1/n) - \delta(t))$.)

[419] Which one???!!!

[420] Another way to see this is to note that $\sum_{d|m} \left(\frac{d}{13}\right) = \sum_{d|m'} \left(\frac{d}{13}\right) = \sum_{d|m'} \left(\frac{m'/d}{13}\right) = \left(\frac{m'}{13}\right) \sum_{d|m} \left(\frac{d}{13}\right)$. Compare this with $\sum_{dd'=m'} \left(\frac{d}{13}\right) d^b = m'^b \left(\frac{m'}{13}\right) \sum_{dd'=m'} \left(\frac{d'}{13}\right) d'^{-b}$. Hence $\sigma_m(s) = m'^s \widetilde{\sigma}_m(-s)$.

Indeed, write residues $\bmod 13^l$ with $l \geq 1$ as $v = v' + 13v''$; here v' runs through a particular collection of 13 lifts of 13 residues mod 13, and v'' runs through residues $\bmod 13^{l-1}$. Then ρ_l may be rewritten as $\sum_{v' \bmod 13} \left(\frac{v'}{13}\right)\mathbf{e}(m \cdot {}^{v'}\!/_{13^l}) \sum_{v'' \bmod 13^{l-1}} \mathbf{e}(m \cdot {}^{v''}\!/_{13^{l-1}})$. Note that the latter sum vanishes if $l-1 > k$, and the former is $\sum_{v' \bmod 13} \left(\frac{v'}{13}\right)\mathbf{e}(m' 13^{k-l} v')$, hence vanishes if $l \leq k$. Otherwise, if $l = k+1$, this leads to

$$13^{l-1} \sum_{v' \bmod 13} \left(\frac{v'}{13}\right)\mathbf{e}({}^{m'v'}\!/_{13}) = 13^{l-1}\left(\frac{m'}{13}\right) \sum_{v' \bmod 13} \left(\frac{m'v'}{13}\right)\mathbf{e}({}^{m'v'}\!/_{13}),$$

and $m'v'$ runs through all residue mod 13. Therefore the latter sum does not depend on m' (for $13 \nmid m'$). By properties of quadratic Gauss sums, it is $\sqrt{13}$, proving the claim above. This leads to

$$\delta_{l\,k+1}\widetilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{v \bmod 13^l} \left(\frac{v}{13}\right)\mathbf{e}(m \cdot {}^{v}\!/_{13^l}) \sum_{d|m} \left(\frac{d}{13}\right) d^s;$$

moreover, one may assume that $d|m'$. As above, we may rewrite the condition $d|m'$, getting

$$\delta_{l\,k+1}\widetilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{v \bmod 13^l} \left(\frac{v}{13}\right)\mathbf{e}(m \cdot {}^{v}\!/_{13^l}) \sum_{d} \left(\frac{d}{13}\right) d^{s-1} \sum_{r \bmod d} \mathbf{e}(m' \cdot {}^{r}\!/_{d});$$

additionally, $\mathbf{e}(m' \cdot {}^{r}\!/_{d}) = \mathbf{e}(m \cdot {}^{r}\!/_{13^k d})$. Hence one can rewrite this as

$$\delta_{l\,k+1}\widetilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{13 \nmid d} d^{s-1} \sum_{v \bmod 13^l} \left(\frac{vd}{13}\right)\mathbf{e}(m \cdot {}^{vd}\!/_{13^l d}) \sum_{r \bmod d} \mathbf{e}(m \cdot {}^{13r}\!/_{13^{k+1} d}).$$

Additionally, the residues of $13^k r \bmod d$ are all distinct, so one can replace $13r$ by $13^{k+1} r$. Hence,

$$\delta_{l\,k+1}\widetilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{13 \nmid d} d^{s-1} \sum_{v \bmod 13^l} \left(\frac{vd}{13}\right) \sum_{r \bmod d} \mathbf{e}(m \cdot {}^{(vd + 13^l r)}\!/_{13^l d}).$$

Obviously, using $R = vd + 13^l r$ the last two sums may be replaced by $\sum_{R \bmod D} \left(\frac{R}{13}\right)\mathbf{e}(m \cdot {}^{R}\!/_{D})$; here $D := 13^l d$; denote $D_{13} := 13^l$. Hence, summing over $l \geq 1$:

$$\widetilde{\sigma}_m = \sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{R \bmod D} \left(\frac{R}{13}\right)\mathbf{e}(m \cdot {}^{R}\!/_{D}).$$

Calculating the real part of Fourier transform as above, we get (on $[0, 2\pi]$)

$$\sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{0 \leq R < D} \left(\frac{R}{13}\right)(\pi\delta(t - 2\pi {}^{R}\!/_{D}) - {}^{1}\!/_{2}) = \pi\sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{0 \leq R < D} \left(\frac{R}{13}\right)\delta(t - 2\pi {}^{R}\!/_{D}).$$

Extending to $t \in \mathbb{R}$, and collecting the terms with the same ${}^{R}\!/_{D}$ (as above), this becomes

$$\pi\ell_{s-1}\sqrt{13} \sum_{13|D,\ D>0} D_{13}^{-s} D^{s-1} \sum_{R,\ (R,D)=1} \left(\frac{R}{13}\right)\delta(t - 2\pi {}^{R}\!/_{D}).$$

here we used the equality $\ell_{s-1} = \sum_u \left(\frac{u}{13}\right) u_{13}^{-s} u^{s-1}$.

One concludes that the real part of the Fourier transform of $\widetilde{\sigma}_m$ is the sum of δ-functions (with non-0 coefficients) concentrated in rational numbers with denominators divisible by 13. (At least for $s < -1$, when our series converge absolutely, hence manipulations make perfect sense.)

Note that this set of points where δ-functions are concentrated is exactly complementary to what we got in the previous calculation (for σ_m). Moreover, for $s = 0$ this answer predicts coefficient 0 at $t = 0$ — but the graph of antiderivative on p. 65 has a non-trivial jump there.

Summation of homogeneous functions on a lattice:

Consider homogeneous functions $\Psi(\tau)$ of degree d on $\mathbb{R}^n$; in other words, $\Psi(a\tau) = a^d\Psi(\tau)$ for $a > 0$. Restriction identifies these functions with functions ψ on the sphere $|\tau| = 1$. Assume that ψ, its derivatives and second derivatives[421] are bounded by M. Then the second derivatives of Ψ are bounded as $CM|\tau|^{d-2}$ with a certain constant C. Therefore $2\Psi(\tau) - \Psi(\tau - \tau_0) - \Psi(\tau + \tau_0)$ is bounded as $CM|\tau|^{d-2}|\tau_0|^2$. Hence the same estimate holds for $\Psi(\tau) - \int_\square \Psi(\tau + \tau')\,d\tau'/|\square|$; here $\square$ is a paralellipiped centered at 0 with the largest diagonal $|\tau_0|$, and $|\square|$ is its volume; we require $2|\tau| > (1 + \varepsilon)|\tau_0|$ with $\varepsilon > 0$.

Conclusion: given a lattice $\mathscr{L}$ in $\mathbb{R}^n$ and a bounded function α on $\mathscr{L}$, the sum $\sum^\circ_{\tau\in\mathscr{L}} \alpha(\tau)(\Psi(\tau) - \int_\square \Psi(\tau + \tau')\,d\tau'/|\square|)$ converges absolutely for $d-2 < -n$; here $\sum^\circ$ means skipping[422] τ with $|\tau| \le |\tau_0|$. Since in this context it is much easier to estimate integrals than sums, this observation is the principal tool in summation of values of homogeneous functions — however, we need a slightly different approach.

Assume that α is even $\mathscr{L}'$-periodic (here $\mathscr{L}'$ is a sublattice of $\mathscr{L}$) with average 0. Take a centrally-symmetric collection U of representatives of all translations $\tau + \mathscr{L}'$ of $\mathscr{L}'$ inside $\mathscr{L}$; assume that U is finite and contains 0. Then $\sum_{\tau_0\in U} \alpha(\tau + \tau_0)\Psi(\tau + \tau_0)$ can be also bounded as above, $CM|\tau|^{d-2}|\tau_0|^2$, with τ_0 the “diameter” of U. Hence the external sum in $\sum'_{\tau\in\mathscr{L}'} \sum_{\tau_0\in U} \alpha(\tau + \tau_0)\Psi(\tau + \tau_0)$ is absolutely convergent for $d - 2 < -n$; here prime means that we omit $\tau = 0$. (Note that the internal sum is finite.) If $d < -n$, then this sum coincides with $\sum_{\tau\in\mathscr{L}\smallsetminus U} \alpha(\tau)\Psi(\tau)$ (which converges absolutely). **Conclusion:** the former summation method (with added $\sum'_{\tau\in U} \alpha(\tau)\Psi(\tau)$) gives a generalization of summing $\sum'_{\tau\in\mathscr{L}} \alpha(\tau)\Psi(\tau)$: it gives correct answers for $d < -n$, and makes sense on a larger set $d - 2 < -n$ of degrees d.[423] [424]

Remark 97: An important related question is the possibility of analytic continuation when the sum is restricted to points $\tau \in \mathscr{L} \cap C$; here C is a cone with the vertex at 0.[425] When one restricts summation to shifts $\tau + U$ of U with $\tau \in \mathscr{L}'$ which are completely contained inside C, the same arguments as above show that the corresponding sum over τ absolutely converges for $d < 2 - n$, and for a fixed $d < 2 - n$ the obtained function of C is uniformly $O(\alpha)$; here α is the “solid angle” of the cone.

The points of C not involved in the summation above are in a narrow strip near ∂C; one can immediately see that for cones with boundary of dimension $n - 1$ this summation absolutely converges

[421] In fact, the arguments below work also when the first derivative is Lipschitz with the constant M, and the function is bounded by M.

[422] We need to skip such values since Ψ is not defined at the origin (at least when d is negative). In what follows, skipping $0 \in \mathscr{L}$ leads to a lot of clumsiness in the formulas below.

One could avoid this clumsiness completely if we would require that Ψ is homogeneous only for $|\tau| > 1$, and is sufficiently smooth near 0.

[423] One can also consider complex d. Then the conditions are $\operatorname{Re} d < -n$ and $\operatorname{Re} d - 2 < -n$.

[424] Likewise, if α is odd, a similar argument (with the first derivatives instead of the second ones) shows convergence for $d - 1 < -n$. In fact, in both cases more cancellations are possible, and it turns out that one can analytically continue to *any* d.

[425] The property of analytic continuation in s is important since it, in a certain sense, cancels “being only conditionally convergent”. Note that the latter condition shows than one needs some additional information to define the sum of numbers in a set (one needs to specify the “order” of summation: the method to rearrange the given infinite sum into a sum of sums of sums etc.). On the other hand, the depence on this “method” disappeares if we require that the “method” satisfies these additional conditions:

- Every “intermediate” infinite sums of the method is absolutely convergent when s is in a certain set Σ;
- The set Σ is connected and contains the given value s_0;
- On an open part of Σ the series converges absolutely;
- We are intersted in the sum when $s = s_0$;

provided that the sum (well defined for values of s where it converges absolutely) has an analytic extension.[426]

[426] Indeed, absolutely convergent sums preserve analyticity.

Because of this if Σ is open, then the analyticity condition above follows from other conditions: two methods with the same open connected set Σ must give the same results for every $s \in \Sigma$ if they coincide on an open part of Σ.

for $d < 1-n$. Moreover, if the function ψ vanishes on ∂C, it absolutely converges for $d < 2-n$. This restricts the question of the behaviour of analytic continuation to calculations on ∂C.

The situation becomes particularly simple for $n = 2$, when C is an angle (so ∂C is automatically of dimension 1), and C is controlled by two numbers: the angles b', b of bounding rays. The "completely contained" part of the sum gives a function $\varphi(b', b)$ which such that $|\varphi(b', b_2) - \varphi(b', b_1)| \leq O(b_2 - b_1) + O(D^{d-1})$, here D is the minimal denominator of rational numbers between b_1 and b_2.[427]

However, consideration of the "remaining" terms is more delicate.

Remark 98: Here we examine only the case when the cone C is polyhedral. One can immediately see that the summation over parts near "edges" of this cone absolutely converges for $d < 2-n$; hence the question of analytic continuation is reduced to what happens near highest-dimensional faces of ∂C. Essentially, we need to investigate what happens near a (part of a) hyperplane.

Consider values of $\tau \in \mathscr{L}'$ such that the region $\tau + U$ considered above is bisected by the given hyperplane Π. Assume that these values "form a staircase": for a certain projection to Π there is at most one such τ with the given projection.[428] Assume that the kernel of Π is spanned by a vector l_0 in the lattice $\mathscr{L}'$. Then the image $\mathscr{L}^\circ$ of $\mathscr{L}'$ under this projection may be lifted back to $\mathscr{L}'$, consider a linear functional β which vanishes on this lifting, and takes value 1 on l_0.

It is not hard to see that the way $\tau + U$ is bisected by Π depends only on $\beta(\Pi(\tau)) \bmod \mathbb{Z}$, and that this value does not depend on the choice of β (for fixed l_0). Denote by U_τ the part of such "bisected" $\tau + U$ which is "above Π". Hence if we want to sum values of a certain function over points in all sets U_τ, the essential component is the behaviour of the fractional part of a linear function β on the lattice $\mathscr{L}^\circ$.

In particular, if we sum $\alpha\Psi$ where Ψ is changing slowly for $|\tau| \gg 0$, then one can rewrite $\alpha(\tau+\tau_0)\Psi(\tau+\tau_0) = \alpha(\tau+\tau_0)\Psi(\tau) + \alpha(\tau+\tau_0)(\Psi(\tau+\tau_0) - \Psi(\tau))$. Under our assumptions on Ψ, the difference $\Psi(\tau+\tau_0) - \Psi(\tau)$ is $O(|\tau|^{d-1})$ for large τ. This immediately implies that these difference terms contribute an absolutely convergent part into summation over U_τ. On the other hand, the first term contributes $\Psi(\tau) \sum_{\tau_0 \in U_\tau} \alpha(\tau_0)$, and the sum depends only on $\beta(\Pi(\tau))$. In fact, the sum may be written as $\xi(\beta(\Pi(\tau)))$ with a 1-periodic locally constant function ξ.

We conclude that the question of analytic continuation of a sum over $C \subset \mathbb{R}^n$ is reduced to investigating $\sum_{\tau \in \mathscr{L}^\circ} \xi(\beta(\tau))\Psi(\tau)$ (or similar sums over polyhedral cones in $\mathscr{L}^\circ \subset \mathbb{R}^{n-1}$); here β is a linear function on $\mathscr{L}^\circ$, and ξ is a 1-periodic piecewise-constant function. Under our assumptions ξ is odd. Note that for questions above, we are interested in cases when the degree d of homogeneity of Ψ and the dimension $n' = n-1$ of $\mathscr{L}^\circ$ satisfy $d < 1 - n'$.

Note that when $\xi \circ \beta$ is periodic (and automatically odd) on $\mathscr{L}^\circ$, the argument in Footnote 424 on p. 138 implies the required analytic dependence. This happens when the hyperplane has a normal in the dual lattice to $\mathscr{L}$. (For $n = 2$ this happens when the slopes of ∂C are rational.)

Apply this to the functions $\Psi(R, D) = \Psi_0(R, D)|D|^{s-1}$; here $n = 2$, $\tau = (R, D)$, and Ψ_0 is of homogeneity degree 0. Let $\mathscr{L}$ be the integer lattice. Suppose that Ψ_0 is smooth away from 0, and $\Psi_0(R, D)|D|^{s-1}$ has bounded second derivatives on $|\tau| = 1$ for a certain range of s␣s with $\operatorname{Re} s < 1$. (For example, this happens when Ψ_0 has a zero of sufficiently large order on $D = 0$.) Under the above assumptions on α, one concludes that using the summation method above, one can extend the function $\sum'_{\tau \in \mathscr{L}} \alpha(\tau)\Psi(\tau)$ of s from the region $\operatorname{Re} s < -1$ to $\operatorname{Re} s < 1$ as an analytic function of s.

Moreover, one can see that the conditions on Ψ_0 hold if the function $\Psi_0(R, 1)$ and its first two derivatives vanish suffiently quickly when $R \to \infty$. This immediately implies that $\sum'_{(R,D)} \alpha(R, D) D^{s-1} \delta(t -$

---

[427] If one of $b_{1,2}$ is rational with denominator D', then $1/D$ is bounded by $D'|b_2 - b_1|$. So in this case the second term is also similar to the Lipschitz estimate if $d \leq 0$. In the case we are most interested in, when $d = -1$, the Lipschitz estimate holds for "not badly approximable" numbers (which can be approximated by rationals with more than quadratic precision).

[428] In general, one can assume that the number of such preimages is bounded. The method below work with this general case as well, so the assumption above is needed only to simplify notations.

${}^{R}/_{D}$) (which is a well-defined generalized function if $s < -1$) extends as an analytic function of s (with values in generalized functions!) to the region $s < 1$.

In particular, every Fourier coefficient of the latter generalized function depends analytically on s when $s < 1$. In particular, if we know Fourier coefficients for $s < -1$, we can extend them analytically to $s < 1$, and the extended value is the Fourier coefficient of the extended generalized function.

Moreover, one can go in different direction: start with a sequence depending on parameter s; suppose that for $s < -1$ the corresponding Fourier series converges to $\sum'_{(R,D)} \alpha(R,D) D^{s-1}\delta(t - {}^{R}/_{D})$, with α satisfying the conditions above. Then we know that for $s < 1$ the Fourier series converges (in the sense of generalized functions) to

$$\sum_{(R,D)\in U}{}' \alpha(R,D)D^{s-1}\delta(t-{}^{R}/_{D}) + \sum_{(R',D')\in\mathscr{L}'}{}' \sum_{(R,D)\in U} \alpha(R'+R, D'+D)(D'+D)^{s-1}\delta\left(t - \frac{R'+R}{D'+D}\right).$$

Moreover, the estimates above (with the second derivatives) show that the second antiderivative of this generalized function is a function on $\mathbb{R}$ of class $\mathscr{L}_1$. In particular, the third antiderivative is a well-defined absolutely continuous function.

The case $n = 2$

For $n = 2$ the situation of the preceding remark is reduced to analytic continuation of the sum $\sum_m \Xi(m\gamma)/m^s$; here Ξ is an odd 1-periodic piecewise-constant function. Instead of such Ξ, it is enough to consider the case when $\Xi = \xi_\beta$, here $\xi = \xi_\beta$ is 1-periodic and is $1-\beta$ on $[0,\beta]$ and $-\beta$ on $[\beta,1]$ (so that the average value[429] of ξ is 0). One way to estimate such sums is the Abel's summation formula: if we can show that $\sum_m \xi(m\gamma)$ grows sufficiently slow, then $\sum_m \xi(m\gamma)/m^s$ converges. For example, if $\left|\sum_{m\le M}\xi(m\gamma)\right|$ grows not quickier than $M/\log^2 M$, then $\sum_m \xi(m\gamma)/m^s$ converges for $s \ge 1$. In fact, for the aim of analytic continuation, it is enough if the “M-summation” $\sum_M {}^{1}/_{M^2}\sum_{m\le M}\xi(m\gamma)$ converges absolutely, and that ${}^{1}/_{M}\sum_{m\le M}\xi(m\gamma) \to 0$. Since the second property is much easier to show, and does not need any new method, we cover only the first one.

To estimate such sums, assume[431] $|\gamma - {}^{A}/_{B}| < {}^{1}/_{B^2}$, take any B consecutive numbers $m\gamma$ and consider the set of their fractional parts. One can see that the elements of this set differ no more than by ${}^{1}/_{B}$ from numbers ${}^{k}/_{B}$ with $k = 0,\dots,B-1$ (when considered $\operatorname{mod}\mathbb{Z}$, so we glue 0 and 1 together). In particular, the count of these fractional parts which are in $[0,\beta]$ may be estimated, and one can see that $|\sum_m \xi(m\gamma)|$ over this range of m is bounded by 3. Hence one can estimate[432] $\left|\sum_{m\le M}\xi(m\gamma)\right| \le 3K_M$ if M may be represented as a sum of K_M denominators of continued fractions for γ. (Indeed, each of these denominators works as the number B above.)

For example, if M is between such denominators Q_l and Q_{l+1}, then $K_M \le \sum_{k\le l+1} a_k$; here a_k are coefficients of the continued fraction of γ. This may be improved to $K_M \le \sum_{k\le l} a_k + M/Q_l$.

Typically, the sequence Q_l grows much quickier than a_l. Let $A_l = \sum_{k\le l} a_k$; note that $Q_{l+1} = a_{l+1}Q_l + Q_{l-1}$. Running the M-summation above for numbers between Q_l and Q_{l+1} gives an estimate $\sum_{Q_l\le M<Q_{l+1}} {}^{(A_l + M/Q_l)}/_{M^2}$. The term A_l contributes at most ${}^{A_l}/_{Q_l} = \sum_{k\le l} {}^{a_k}/_{Q_l}$; summing such terms over l, the number a_k comes with a coefficient $\sum_{l\ge k} {}^{1}/_{Q_l}$; since $Q_{k+2} > 2Q_k$, this coefficient is bounded

[429] For irrational γ the value of Ξ or ξ at a point of jump does not matter for analytic continuation. For rational γ one approach is to follow the standard convention: average values below/above the jump. This gives the average of two sums: for a closed cone and for an open cone.[430]

These two sums correspond to taking value either above or below the jump (depending on whether the cone is closed or open, and on whether the hyperplane is “top” or “bottom” boundary). *These* sums may have a pole in analitic continuation; to avoid a pole, we need an additional condition: the average value of α on the hyperplane should be 0.

[430] **N.B. Is it???**

[431] As in Dirichlet's approximation theorem.

[432] For rational γ the estimate may be replaced by the “last” A_k (defined below), which in turn is bounded by the denominator of γ.

as ${}^4/_{Q_k}$. Now summing over k gives $4\sum_k {}^{a_k}/_{Q_k}$; however, ${}^{a_l}/_{Q_l} \sim {}^1/_{Q_{l-1}}$, hence this part of summation converges absolutely.

The remaining term M/Q_l contributes at most $\int_1^{a_{l+1}+1} {}^1/_{Q_l^2 \varkappa}\, Q_l \,d\varkappa = {}^{\log(a_{l+1}+1)}/_{Q_l}$. **Conclusion:** if $\sum_l {}^{\log a_{l+1}}/_{Q_l}$ converges, then M-summation converges absolutely, hence the analytic continuation works for $s \leq 0$, and coincides with the sum of the series for $s = 0$. Since one can replace $\log a_{l+1}$ by $\log Q_{l+1}$ without changing convergence, this condition is the Bruno condition; in a certain precise sense, only "extremely pathological" numbers fail this condition.

This shows that if an angle C has non-pathological directions of the bounding rays, the sum over $\tau \in \mathscr{L} \cap C$ "makes sense" for $s = 0$. Moreover, one can find it by

- For every $\tau' \in \mathscr{L}'$ add together the terms corresponding to points of $\mathscr{L} \cap C$ inside the translation $\tau' + U$ of U;
- Sum up (in any order) the obtained totals for translations $\tau' + U$ fully contained inside C;
- Add the sum of the totals for the remaining translations $\tau' + U$ (in the order of the distance from the origin).

Additionally, it shows that this method of summation is compatible with subdivision of an angle into several smaller angles: when it is compatible[433] for $s < -d$, the analytic continuation must also be compatible.

On the other hand, we already saw that for directions of ∂C with rational slope ${}^p/_q$ we can do much more: the M-summation converges quickly enough (the remainder is bounded by ${}^q/_{M_0^{1-d}}$ with M_0 being the cut-off) iff the average of α on the boundary ray is 0. In fact, for "typical" numbers, coefficients of the continued fraction satisfy $a_k < \lambda(k)$ for all but a finite number of k provided $\sum_k {}^1/_{\lambda(k)}$ converges (Khinchin's estimate in "Th. 30"). Hence $a_k = o(k \log^2 k)$ and $A_k = o(k^2 \log^2 k)$. Compare this with Q_k, which grow at least as a geometric progression. This shows that for such numbers $K_M = O(\log^3 M)$. Essentially, this adds "only logarithmic terms" to our estimate of the remainder of M-summation valid for rational slopes. (Moreover, analytic continuation works up to $s < 1$.)

The last considerations become important when we consider how the sum changes when one replaces the cone C with another one C' which differs by a small rotation of one of the boundary rays. By compatibility with subdivision, we can consider the case of an open and very sharp angle C instead (of magnitude $|C|$). First[434] restrict attention to the rational slopes of the boundary rays. To avoid poles of analytic continuation (see Footnote 429 on p. 140), assume that the average value of α on any hyperplane in $\mathscr{L}$ is 0.[435]

One can break the sum into 4 parts (below U^τ is the translation $\tau + U$ of U with $\tau \in \mathscr{L}'$) running over:

- The U^τs fully contained inside C.
- The U^τs which are bisected by the second ray, but not the first one.[436]
- The $U^\tau \smallsetminus C$s with U^τ bisected by the second ray. (Sum taken with opposite sign.)
- The $U^\tau \cap C$s with U^τ bisected by the first ray.

Above, we estimated the first sum as $O(|C| + N_C^{d-1})$ with N_C the smallest magnitude of a point of a lattice strictly inside C (note that $|C| = O(1/N_C)$). Note that in the remaining parts we can omit U^τ if it has no points with magnitude $< N_C$. Then (similarly to the first one) the second term is bounded as $O(N_C^{d-1})$. For rational slopes one gets an estimate $O(q \cdot N_C^{d-1})$ for the other two terms; here q is the maximum of denominators of slopes of boundary rays.

---

433 Of course, the ray separating the angles should be included in one of the angles only.

434 **N.B. Check???**

435 Note that this is not very restrictive. For example, if $\mathscr{L}'$ is of a prime index in $\mathscr{L}$, this adds the condition $\alpha(0) = 0$.

436 Here we sum over the "whole" U^τ — as opposed to $U^\tau \cap C$.

However, the latter estimate does not ensure continuty, since (in the case of rational slopes) N_C and q are of the same order of magnitude. So to examine discontinuities for $d = -1$, we need to investigate the behaviour of $\max_M \left|\sum_{m\leq M} \Xi(m\gamma)\right| \big/ Q$, here Q is the length of the period (the denominator of γ)[437]. To simplify bookkeeping, assume that $|\Xi|$ and jumps of Ξ are bounded by 1.

Proceed as above: assume that $|\gamma - {}^r\!/_q| < {}^1\!/_{Nq}$; then the sum $\mathbf{s_n} := \sum_m \Xi(m\gamma)$ over $n \leq N$ consecutive values of m differs from the corresponding sum $\sum_m \Xi(m\gamma_0 + \delta)$ (with $\gamma_0 := {}^r\!/_q$, and an approriate δ) by no more than the total variation v of Ξ. Note that the latter sum vanishes for $n = q$ and $\delta = 0$; when it vanishes for every δ we get an estimate $v + 2\max_M \left|\sum_{m\leq M} \Xi(m \cdot {}^r\!/_q)\right|$ for $\mathbf{s_N}$, and the estimate $({}^1\!/_N + {}^1\!/_Q) \cdot \left(v + 2\max_M \left|\sum_{m\leq M} \Xi(m \cdot {}^r\!/_q)\right|\right)$ for $\max_M \left|\sum_{m\leq M} \Xi(m\gamma)\right| \big/ Q$. (Indeed, we need about ${}^Q\!/_N + 1$ such runs to cover the whole period.)

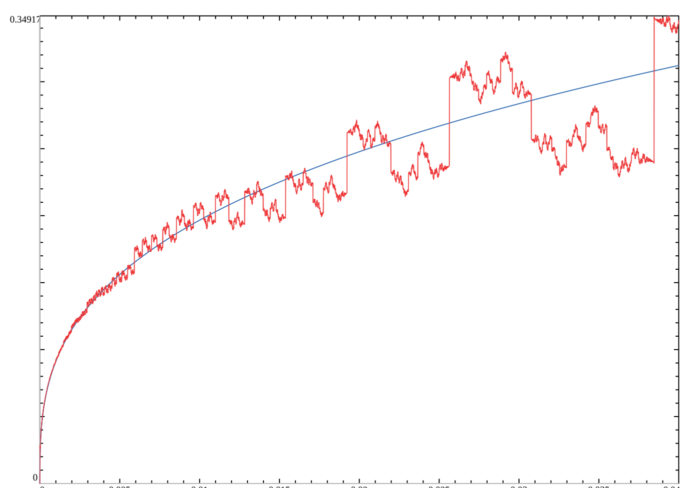
Moreover, when $\gamma \neq \gamma_0$, one has ${}^1\!/_N + {}^1\!/_Q < 2q\Delta$ with $\Delta := |\gamma - {}^r\!/_q|$. Hence when there is no dependence on δ, and one ray of the angle (with rational slopes!) is fixed, the sum is bounded by a multiple of Δ.

Conclusion: consider our regularized sum $\sum_\tau \alpha(\tau)\Psi(\tau)$ (with Ψ of homogeneity degree -1) taken over $\tau \in \mathscr{L}$ in an open angle C. Fix α and Ψ; then the sum can be bounded (in magnitude) by $\text{const} \cdot q_1^2|C|$ (with q_1 being the denominator of the slope γ_1 of one of the boundary rays of C in a particular basis of $\mathscr{L}$) provided:

- the boundary rays of C go in “rational” directions w.r.t. C;
- the function Ψ is smooth away from 0;
- the function α is double-periodic with average 0;
- the function α has average 0 on both boundary rays;
- the function α has average 0 on any line in $\mathscr{L}$ going in the direction γ_1.

Together with additivity, this gives a partial description of the behaviour of the sum in an open angle C when one varies one of the rays $\mathscr{R}$ of the angle. Call an $\mathscr{L}$-rational direction *admissible* if the average of α on the line of this direction through 0 vanishes; call it *strongly admissible* if the same holds for all translations of this line. Restrict attention to angles with admissible direction of $\mathscr{R}$; then near a strongly admissible direction, there is a jump of $\sum_{\tau\in\mathscr{R}} \alpha(\tau)\Psi(\tau)$, and there are one-sided Lipschitz estimates on any side of the jump.

For $\alpha(p,q) = \left(\frac{q}{13}\right)$ on $\mathbb{Z}^2$ any rational directions is admissible; it is strongly admissible iff the slope $\gamma_1 = {}^{p_1}\!/_{q_1}$ has denominator prime to 13. Likewise, for $\alpha(p,q) = \left(\frac{p}{13}\right)$ on the sublattice $\mathscr{L} \subset \mathbb{Z}^2$ given by $13|q$, strongly admissible directions have $13|q_1$.



As we already saw, these two cases lead to the same sums over angles; since any rational direction is strongly admissible for one of these cases, this explains the observed properties of the graph on p. 63. (Note that this explanation works for rational directions only; to include — typical? — irrational values would require additional arguments.)

What we proved above is that any one of two sums above has the “expected” jumps at the points where the terms we sum have jumps, however, it *also* has *“spurious”* jumps; they happen where the *other* sum has “expected jumps“, and

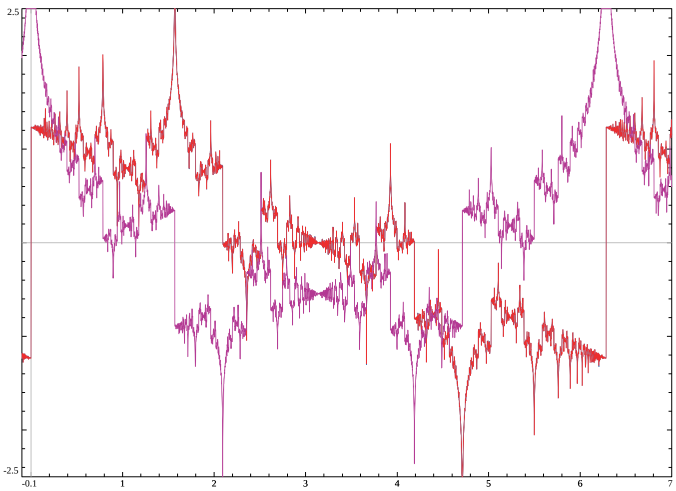
[437] Note that the position of jumps of Ξ depends on γ. Moreover, at some values of γ the jumps may “collide”. However, this dependence turns out to contribute only negligible terms into the estimates we need, so we are going to ignore it.

are of the same height as these jumps. However, having a proof does not make *an explanation.* How come these spurious jumps appear?

As a partial explanation, observe what happens when we "deform" the sum, changing s from 0 to negative values. The plot above shows what happens near $t = 0$ when $s = -1/3$, together with the graph of $y = C \cdot t^{1/3}$. This plot has only the "expected" jumps at points p/q with $13 \nmid p$, of magnitude $\sqrt{13}/q_{13}^s q^{1+s}\left(\frac{p}{13}\right)$. In general, the curve to fit when $t \to +0$ is[438] $C_s \cdot t^{-s}$. When $s \to 0$, the coefficient C_s goes to ½, and y goes to ½ for positive t. Since the function is odd, with $s = 0$ we get a jump of 1 at $t = 0$.

Conclusion: the spurious jumps appear only when $s = 0$. For other values of s, one gets "less confusing" power-law singularities at the positions of spurious jumps.

Remark 99: For the case of negative discriminant, the corresponding Eisenstein series do not behave as nice, so the surprise factor of "spurious" jumps coming from clear blue sky disappears. Indeed, in this case the toy transform mixes together the real and the imaginary parts. As the graphs on p. 63 show, when the real part has jumps, the imaginary part must have a log-singularity; hence if one expects jumps, then the real part has *both* jumps and log-singularities. This breaks the symmetry between "expected" and "spurious" features.



Indeed, this is what happens in reality. On the right is the example plot for a decomposable polynomial $x^3 + x$ with discriminant -4. The only special prime is 2, and one could use `LST = [[2,[1,0,3]]]` for `PN_nINIT()` (see Footnote 704 on p. 217). Moreover, it looks like there is no jump at π — so the situation is not as clear-cut as in the case of positive discriminant.

A final (and probably most important) remark: the presence of log-singularities on a dense subset shows that both "real" graphs[439] are going to fill the plane. This suggests that the approach we used would probably make little sense for *odd* functions α (compare with even/odd cases of Euler formulation on p. 16).

In other words: in this context, the Maass case is much more interesting than the "case of modular forms".

[438] Heuristically, this may be explained by the last identity of Footnote 420. If we could replace m'^s by m^s in this identity, then the Fourier transforms of $\sigma(s)$ would be "the fractional derivative of order s" of the Fourier transform of $\widetilde{\sigma}(-s)$. Since "taking a derivative" is "a convolution with the function δ'", and "taking derivative of order s" is a convolution with $t^{-1-s}/\Gamma(-s)$, under "the heuristic assumption above" if one expects $\delta(t - t_0)$ to appear in the Fourier transform of $\widetilde{\sigma}(-s)$, one should also expect a singularity of type $(t - t_0)^{-1-s}/\Gamma(-s)$ to appear in the Fourier transform of $\sigma(s)$ (and the same with σ and $\widetilde{\sigma}$ exchanged).

Taking antiderivative of this produces a singularity of type $(t - t_0)^{-s}/\Gamma(1 - s)$. This is *almost* exactly what we saw above. However, this heuristic does not work ideally: we needed an extra factor 0.815 to make a match in the plot above. (Naive approach taking into account only the jump of σ at 0 would lead to the coefficient $12/(13 - 13^s) \approx 0.9543$.)

[439] As opposed to simulations made by interpolating from a small collection of values of t.

Supplementary Musings: closing the gaping holes

In construction! (Lousy — but more or less complete — exposition (except for ζ-function).)

The overwhelming consideration governing the design of these notes is to simplify the exposition as much as I could. For this, a lot of interconnections were ignored and (sometimes) a *deus ex machina* has been invoked — without sufficient explanations.

Here we treat such dust which we put under the carpet before. It is quite probable that these omissions have very little to do with the Langlands program, but are only related to the particular shortcuts we use in these notes.

More details on the M-family

On p. 18, we introduced the M-family of polynomials "$M \cdot$ tetrahedral numbers $+ N$" (considering rational coefficients allows us to fix $N = 1$) and announced that it contains a very large pool of "interesting" cases. Here we explain why the "simpler" a cubic polynomial is, the better is the chance that it produces the same function $F(t)$ as some polynomial from the M-family.[440]

First, note that changing variable by substituting $x = an + b$ with suitable a, b into a polynomial $P(x)$ would change the list of prime divisors of the values of the polynomial in only a finite number of positions. Since we may ignore "exceptional" primes anyway, we can use this substitution to consider only the polynomial $x^3 + N'x + N''$ with *two* suitable parameters $N', N'' \in \mathbb{Z}$. (The discriminant of this polynomial is $D = -4N'^3 - 27N''^2$.)

In fact, a more involved analysis shows that the same happens not only for linear substitutions, but for quadratic "*Tschirnhausen transforms*"[441] as well. Essentially, this means that the function $F(t)$ depends not on the polynomial, but on the cubic extension of $\mathbb{Q}$ defined by this polynomial.

Our family corresponds to $N' = -1$, $N'' = N/M \in \mathbb{Q}$, and $D = 4 - 27N''^2$. To find N'' matching the given square-free part d of D, one needs to solve $x^2 - 3y^2 = d$. Proceeding as in Footnote 280 on p. 100 leads to the conditions

- $d \not\equiv_8 2, 3, 7$ — automatically satisfied for cubic discriminants, and
- $d \equiv_9 0, 1, 4, 6, 7$, and
- 3 must be a quadratic residue mod prime divisors p of d — equivalent to $p \equiv_{12} \pm 1$.

Essentially, if d has K prime divisors larger than 3, the fraction of such numbers d such that the equation above has solutions is about ${}^2/_{3 \cdot 2^K}$. Since K is usually very small unless d is very large, this explains why a lot of cases of small discriminants can be represented by our family. **Conclusion:**

> While most cubic polynomials do not give $F(t)$ from an M-family, many "simple" ones do.

(In fact, the M-family contains 30% of the possible field discriminants below 1,500 in magnitude, and 25% for the cut-off at 25,000. On the other hand, when one considers the suitable modular forms with small conductors, it seems that M-family may miss a significant number of them; see Footnote 445 on p. 145.)

Remark 100: Above, we ignored existence of different cubic extensions whose discriminants coincide. One can check that for small $|D|$, such coincidences happen rarely: the smallest positive/negative cases are $3^4 \times 7^2 = 3{,}969$ (cyclic), $2^2 \times 3^5 \times 23 = 22{,}356$ (non-cyclic) and $4 \times 3 \times 2351 = 28{,}212$ (a fundamental discriminant), and $-1{,}228$.

---

440 Here one can use the magnitude of the coefficients as a measure of simplicity. (However, the more precise measure is the magnitude of "the field discriminant"; we use a similar measure below.)

441 Given a quadratic polynomial $\Pi(x)$, this transform is another cubic polynomial $P_\Pi(x)$ such that $P_\Pi(\Pi(x_0)) = 0$ for every root x_0 of $P(x)$. (The paper of Buhler and Reichstein is a very good introduction.)

To analyse this, note that cyclicity is determined by $d = 1$, hence there may be no coincidence of discriminants between cyclic and non-cyclic cases. In cyclic cases such coincidences happen when D is a products of more than one number from the list 9^2, 7^2, 13^2, 19^2 etc.[442]

The simplest coincidences of non-cyclic extensions are related,[443] by Class Field Theory, to having more than one subgroup of index 3 in the Class Group $\mathsf{Cl}(\mathbb{Q}[\sqrt{d}])$. (This is the same as having more than 2 elements of order 3 in this group.) So the rarity of this situation is related to the class number being typically not very large. (Unfortunately, there is very little proven about the related statistical properties of these groups…)

Remark 101: One can check that although the examples above produce the same discriminant (hence conductor), still they result in different functions $F(t)$. Moreover, this is a general situation: if two polynomials result in the same function $F(t)$, then they are related by a Tschirnhausen transforms.[444]

3 smallest conductors

Remark 102: Using `mfeigensearch` as described on p. 218, one can see that our function F investigated in the section on p. 68 matches a modular form with the conductor[445] $c = 39$. One can find the corresponding L-function by `lf=lfunartin(NF=nfinit(nfsplitting(P)),gal=galoisinit(NF),[[0,1;-1,0],[` to check the coincidence with the initial coeffients calculated by our rules, use `lfunan(lf,#N_n)==N_n`.

Compare this with a direct match of the graphs: a peak on the graph of $F^{(-1)}(t)$ near $t = 0.228$ matches a peak near $t = 4.44$ on the purple graph in the beginning of that section (for $\operatorname{Im} F_{\mathbb{C}}^{(-1)}(t)$). This leads to $c \approx {}^{4\pi^2}\!/_{4.44\cdot 0.228} \approx 38.998$.

The flattened parts of the graphs

It is not hard to explain why such parts appear on graphs of partial sums of Fourier series for functions $H(t) := tG(-{}^1\!/_t)$ (or $|t|G(-{}^1\!/_t)$) with a periodic function G. **Answer:** When the average of

[442] Indeed, by Class Field Theory one should start with subgroups of index 3 in $(\mathbb{Z}/m)^\times$; exclude subgroups induced by surjections $(\mathbb{Z}/m')^\times \to \mathbb{Z}/3$ with $m'|m$ and $m' < m$. The remaining subgroups match cyclic cubic extensions of discriminant m^2 (by the "Conductor-Discriminant Formula"). Obviously, the number of such subgroups is the number of points in $\mathbb{P}^{k-1}(\mathbb{Z}/3)$ which are not in the coordinate cross, here k is the number of divisors of m which are either 9, or prime p with $3|p-1$. So what is needed to allow several choices is $k > 1$ — leading to the answer above.

[443] Unfortunately, this relation does not lead to a complete answer. As the example above with a non-fundamental discriminant $D = 22{,}356$ shows, it is not possible to avoid consideration of (more complicated) "*ray class groups*". (One can recognize $(\mathbb{Z}/m)^\times$ from the preceding footnote as the simplest example of a ray class group.)

[444] Indeed, coincidence of functions $F(t)$ is analysed in Exercise 6.4 of the collection edited by Cassels and Fröhlich. It may happen non-trivially only in the non-abelian case, and the corresponding field extensions should have the same discriminant (=conductor). The exercise concludes that the corresponding Galois subgroups of the compositum field must be "conjugation numerically-equivalent": any conjugacy class should intersect these two subgroups in the same number of elements. **Conclusion:** this situation is not possible for cubic extensions: the subgroups are going to be conjugate (hence the fields are isomorphic)!

(Indeed, since discriminants coincide, we get two abelian cubic extensions of the same quadratic extension. Hence the combined compositum of these cubic fields and the quadratic field is Galois. Moreover, the Galois group must be the external product of $\mathbb{Z}/2$ acting on $\mathbb{Z}/3 \times \mathbb{Z}/3$ as multiplication by -1. We need to show that any two conjugation numerically-equivalent subgroups of index 3 are conjugate.

However, this group is isomorphic to the group of translations and central reflections on the plane over the field $\mathbb{Z}/3$. Hence any subgroups of index 3 must consist of reflections in points on a line, and translations along this line. Looking at conjugacy classes of translations shows that two lines corresponding to two subgroups must be parallel. But then they are conjugated by a reflection in any point not on these lines.)

[445] This is the second smallest conductor of a modular form appearing in these notes. We already saw the smallest possible conductor 23 in "the case $M = 6$ in the M-family" considered in the section on p. 51.

In fact, as `mf=mfeigensearch([[1..40],1], [])` in `GP/PARI` shows, 39 is the third smallest possible conductor for a modular form (of weight 1). The second one is 31 — and it corresponds to the cubic polynomial $P = x^3 + x - 1$. (This conductor does not seem to appear in the M-family — at least with $M = {}^R\!/_S$ and $|M|, |M| \leq 3{,}000$. It appears in Exercise D7 on p. 28.)

G is 0, and the antiderivative of G decays cubically near 0, the partial sum of the Fourier series for $H(t)$ is going to have a flattened zone near 0. To explain this, we use the saddle-point method.

First of all, to have a Fourier series to sum, we would need H to be periodic. Note that the conditions of periodicity of G and H are too severe, and restrict our flexibility too much. On the other hand, if we do not require that H is periodic, we can just cut off its Fourier *transform*[446] at some particular frequency — this would give the same result as cutting off the Fourier series. (In other words, instead of taking the *inverse* Fourier transform, we restrict integration to a particular interval.)

The crucial observation is that the combination of

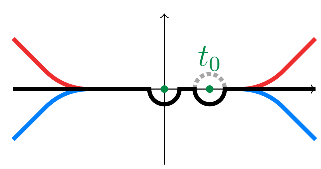
- Take Fourier transform $\widehat{H}(\tau)$ of H;
- Cut it off to a certain interval of τ;
- Take the inverse Fourier transform;

is equivalent to taking a *convolution* with a certain function.[447] One can immediately see that this function is proportional to $\operatorname{sinc} t := {}^{\sin t}/_{t}$ (with appropriately rescaled t).

Above, we were considering the Fourier series for H; now consider the Fourier series for G. Now, when we do not require that H is periodic, one can take G to be a Fourier monomial (essentially, a trigonometric function!). If we can show that a flat region appears for any such G, then summing up the Fourier series for a *general* G would prove[449] the general fact about appearence of flat regions. This leads to

$\boxed{\text{The convolution } t\cos{}^1/_t \star \operatorname{sinc} kt \text{ has a flattened region near } t=0. \text{ Same for sin instead of cos.}}$

In fact, in calculations it is easier to replace $\cos x$ by $\exp \mathrm{i}x$ (and then take the real part, if needed).[450]



Now the problem boils down to giving a ballpark estimate for ${}^1/_\pi \int {}^t/_{(t-t_0)} \exp {}^{\mathrm{i}}/_t \sin k(t-t_0)\, \mathrm{d}t$ for large $k > 0$ and small t_0. Note that the integral is over $\mathbb{R} \setminus 0$, but since $\exp {}^{\mathrm{i}}/_t$ is bounded on $\operatorname{Im} t \leq 0$, we can "add a tiny half-circle below $t = 0$" (as on the right) and replace the path of integration by any deformation of $\mathbb{R}$ passing below $t = 0$; otherwise, we can deform the contour of integration arbitrarily in the plane $\mathbb{C}$.

Moreover, we can replace sin by two exponents — and for $\exp \mathrm{i}({}^1/_t \pm k(t-t_0))$ the regularization of Footnote 450 on p. 146 is equivalent to deforming the contour so that it goes into upper half-plane when $|t| \to \infty$ for sign "+", likewise into lower half-plane for sign "−" (the red and blue contours above).

---

[446] … which, for a periodic function H, is going to be a sum of δ-functions on $\mathbb{Z}$ with coefficients equal to the Fourier coefficients. In other words, with this approach we are *forced* to consider generalized functions.

[447] Indeed, it is enough to check that changing $H(t)$ to its shift $H(t+t_0)$ would shift the result of this operation. This happens because (up to questions of convergence[448]) any linear operator on functions which *commutes* with shifts is a convolution with a certain function C. (To find this function, apply the operator to the δ-function.)

[448] Note that above we *already skipped* certain questions of convergence. Indeed, when cutting-off a δ-function to an interval $[\tau_0, \tau_1]$, what if the δ-function *is at* τ_1?

While switching to the language of generalized functions avoids *a lot* of questions of convergence, some of these questions remain ("they are inavoidable"). To fight these, the standard approach is to apply suitable *"mollifications"* or *"regularizations"* — which *change* the results of the operations.

In this notes, we are consistently ignoring these questions — and we continue to do it here. Just note that *in the final result*, when H *is periodic*, and $\widehat{H}(\tau)$ a sum of δ-functions on $\mathbb{Z}$, a cut-off does not cause any problem as far as the ends of the interval $[\tau_0, \tau_1]$ are not integer. This essentially shows that for such an interval the "changes" mentioned above are going to cancel each other!

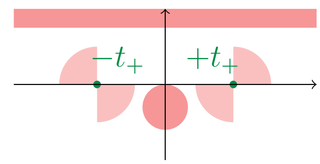
[449] Here we again are going to ignore questions of convergence…

[450] Note that when computing the convolution, the integral is improper, behaving as $\int t(1+{}^{\mathrm{i}}/_t+\dots)\, {}^{\sin(t-t_0)}/_{(t-t_0)}\, \mathrm{d}t$ (after replacing $\cos t$ by $\exp \mathrm{i}t$, as we do below). Essentially, we integrate $\sin(t-t_0) + {}^B/_t \sin(t-t_0) + \dots$; all the terms except the first one converge (though not absolutely), and the first term has an average 0 over period — so can be dealt with any sane "regularization". (For example, one can subtract $\sin(t-t_0)$ from the expression we integrate. This is equivalent to the particular way to bend the paths of integration we use below.)

However, splitting sin into exponents leads to both terms having a pole at $t = t_0$, so *before deformation*, we need to choose on which side of this pole our contour is going to pass. Choosing circling t_0 from below (as on the picture), the integral with the $-$-sign vanishes (since the expression under the integral decreases when we deform the blue contour down[451]).

Conclusion: We need to estimate $^1\!/_{2\pi\mathrm{i}} \exp -\mathrm{i}kt_0 \int {}^t\!/_{(t-t_0)} \exp \mathrm{i}({}^1\!/_t + kt)\, \mathrm{d}t$ over the red deformation of the black contour above. Alternatively, we can pass *above* t_0 (the dotted gray line) — and then the integral gives the difference between the remainder term of the Fourier series (with opposite sign).

The saddle points (the critical points of ${}^1\!/_t + kt$) are at $t = \pm t_+ := \pm {}^1\!/\sqrt{k}$; so the geometry of the magnitude of the function we integrate along the contour depends significantly at whether ${}^1\!/\sqrt{k} < |t_0|$ or ${}^1\!/\sqrt{k} \geq |t_0|$. *This* is the reason for the change of behavior for small $|t_0|$.



On the right, we show (in red) the regions where the magnitude of $\exp \mathrm{i}({}^1\!/_t + kt)$ is small, as well as the sectors near $\pm t_+$ where the magnitude is smaller than in the other two sectors. (To see how these sectors are positioned, it is enough to note that the second derivatives of $\mathrm{i}({}^1\!/_t + kt)$ at $\pm t_+$ are $\pm {}^{2i}\!/_{t_+^3}$.)

According to the saddle-point method,[452] to approximate the integral one should consider the contour passing through the red zones.[453] [454] To get the principal term of the approximation, one should replace the argument of $\exp \mathrm{i}({}^1\!/_t + kt)$ by its quadratic approximations at $\pm t_+$, and replace the contour by straight lines passing through $\pm t_+$ (inside the red sectors). So the main term is mostly contributed by the zones of size $t_+^{3/2}$ about $\pm t_+$.

Note that for $|t_0| < t_+$ the contour of integration for the saddle point method is compatible with the red contour above (hence the integral is the partial sum of the Fourier series), otherwise it passes *above* t_0 (hence the integral is the remainder of the Fourier series — with opposite sign).

It is not very hard to see that for $k \gg 1$ the remainder term of the approximation is negligible. Likewise, the contribution of the non-constant part of the factor ${}^t\!/_{(t-t_0)}$ is neglible for $k \gg 1$ unless $|t_0 \mp t_+| = O(t_+^{3/2})$. In particular, for $k \gg 1$ this factor gives a practically constant contribution in at least one of two integrals about $\pm t_+$.

Conclusion: the principal term in the approximation of the integral $\int_{\mathbb{R}} {}^t\!/_{(t-t_0)} \exp \mathrm{i}({}^1\!/_t + kt)\, \mathrm{d}t$ is

$$\exp(2\mathrm{i}\sqrt{k}) \int_{\mathbb{R}} \frac{t}{t-t_0} \exp \mathrm{i}\frac{(t-t_+)^2}{t_+^3}\, \mathrm{d}t + \exp(-2\mathrm{i}\sqrt{k}) \int_{\mathbb{R}} \frac{t}{t-t_0} \exp -\mathrm{i}\frac{(t+t_+)^2}{t_+^3}\, \mathrm{d}t$$

(above, this integral had an extra factor $^1\!/_{2\pi\mathrm{i}} \exp -\mathrm{i}kt_0$; in this formula, $2\sqrt{k}$ is ${}^1\!/_{t_+} + kt_+$). If the contour of integrations is the contour of the saddle-point method (essentially, this means that it goes through the red zones), one can use the formula

$$\int_{\mathbb{R}} \frac{1}{(t-t_1\mathrm{i})} \exp -\frac{t^2}{2}\, \mathrm{d}t = 2\pi\mathrm{i}(1 - \operatorname{Erf} t_1) \exp \frac{t_1^2}{2};$$

---

[451] For the Maass case, we need to consider the corresponding integral over $\mathbb{R}_+$ instead of $\mathbb{R}$, so this argument does not work. Instead, we can proceed as below: the saddle-point method would deform this contour into $-\mathrm{i}\mathbb{R}_+$ with the saddle point $-\mathrm{i}t_+$. Hence the integral has the order of magnitude of $t_+^{3/2} \exp -2\sqrt{k}$.

For our plots, $\sqrt{k}$ is typically about 100, so in the Maass case this term — while non-0 — is still completely negligible.

[452] Recall that this method has two faces: first, it gives a certain approximation to the integral in question. The second part comes in two flavors: on one hand, one can use some ready-to-use estimates of the error of the approximation above; alternatively, one may look for an "ad hoc" estimate of this error — and it is usually even simpler than finding a pre-cooked estimate.

Here we ignore the question of estimating the error term.

[453] **N.B. (???) $-1/t$ and lower half-plane!**

[454] The function is very rapidly decreasing near the top of the red circle. This means that the considerations above are applicable not only to the integral over $\mathbb{R}$, but also to the integrals over $\mathbb{R}_+$ and $\mathbb{R}_-$ (for them, one needs to restrict attention to one saddle point).

This is what makes the conclusion of this section applicable to the Maass case (when the factor at $G({}^1\!/_t)$ is $|t|$, and not t) as well.

here $\sqrt{2\pi}\,\mathrm{Erf}\,t$ is the antiderivative of $\exp -t^2/2$ vanishing at $t=-\infty$. (This formula holds as far as $t_1\mathrm{i}$ is above the contour of integration. It is easy to establish using the Fourier transform.) Therefore

$$\int_{\mathbb{R}} \frac{1}{(t-t_1)} \exp -\frac{t^2}{\beta}\,\mathrm{d}t = 2\pi\mathrm{i}\left(1-\mathrm{Erf}\left(-\mathrm{i}\frac{t_1}{\sqrt{\beta/2}}\right)\right)\exp -\frac{t_1^2}{\beta};$$

as far as t_1 is above the contour of integration (and $\operatorname{Re}\beta \geq 0$). If t_1 is below the contour of integration, then one needs to subtract $2\pi\mathrm{i}\exp -{t_1^2}/{\beta}$. (In other words: omit 1 in the formula.)

Introducing $E_m(t) := 2\pi\mathrm{i}\left(m - \mathrm{Erf}\left(-\sqrt{\mathrm{i}}t\right)\right)\exp \mathrm{i}{t^2}/{2}$ and $E_+ := E_1$, $E_- := E_0$ leads to

$$\int_{\mathbb{R}} \frac{1}{t-t_1}\exp \mathrm{i}\frac{t^2}{2}\,\mathrm{d}t = E_\pm(t_1)$$

with $+$ or $-$ chosen depending on whether t_1 is above or below the contour of integration. When we integrate over $(1+\varepsilon\mathrm{i})\mathbb{R}$ (with $\varepsilon > 0$, to make the integral quickly converging) and $t\in\mathbb{R}$, then the sign is $-\operatorname{sign} t$.

Note that $E(t) := E_{-\operatorname{sign} t}t$ is odd, behaves asymptotically as $\sim -{\sqrt{2\pi\mathrm{i}}}/{t}$ for large $|t|$, and jumps from $\pi\mathrm{i}$ to $-\pi\mathrm{i}$ when t crosses from -0 to $+0$. Let $\widetilde{E}_\bullet(t) := tE_\bullet(t) + \sqrt{2\pi\mathrm{i}}$. Then $\widetilde{E}$ is a bounded even continuous function behaving as $O({1}/{t^2})$ for large t. (Both E and $\widetilde{E}$ are smooth when the argument is positive.) Then

$$\int_{\sigma\mathbb{R}} \frac{1}{t-t_1}\exp \mathrm{i}\frac{t^2}{\beta}\,\mathrm{d}t = E_\pm\left(\frac{t_1}{\sqrt{\beta/2}}\right) \qquad\qquad \text{if } 0 \leq \operatorname{Arg}{\sigma^2}/{\beta} \leq \pi.$$

(Here the $\pm$ sign must match $\operatorname{Im}{t_1}/{\sigma}$.) Moreover,

$$\int_{\mathbb{R}} \frac{t-t_2}{t-t_1}\exp \mathrm{i}\frac{t^2}{\beta}\,\mathrm{d}t = (t_1-t_2)E_\pm\left(\frac{t_1}{\sqrt{\beta/2}}\right) + \sqrt{\beta\pi\mathrm{i}} = \sqrt{\beta/2}\widetilde{E}_\pm\left(\frac{t_1}{\sqrt{\beta/2}}\right) - t_2E_\pm\left(\frac{t_1}{\sqrt{\beta/2}}\right).$$

For $\beta > 0$ the contour of integration may be $\mathbb{R}$ (or $(1+\varepsilon\mathrm{i})\mathbb{R}$ with $\varepsilon \geq 0$); for $\beta < 0$ it may be $-\mathbb{R}$ (or $(-1+\varepsilon\mathrm{i})\mathbb{R}$ with $\varepsilon \geq 0$). With $\varepsilon > 0$ and $t_1\in\mathbb{R}$ this sign is $-\operatorname{sign} t$ in both cases.

Combining all this together, the saddle-point method replaces the integral $\int_{\mathbb{R}} {t}/{(t-t_0)}\exp \mathrm{i}({1}/{t}+kt)\,\mathrm{d}t$ by

$$\exp(2\mathrm{i}\sqrt{k})\left(t_+E\left(\frac{t_0-t_+}{\sqrt{t_+^3/2}}\right)+\sqrt{t_+^3/2}\widetilde{E}\left(\frac{t_0-t_+}{\sqrt{t_+^3/2}}\right)\right)+\exp(-2\mathrm{i}\sqrt{k})\left(t_+E\left(\frac{t_0+t_+}{\sqrt{t_+^3/2}}\right)-\sqrt{t_+^3/2}\widetilde{E}\left(\frac{t_0+t_+}{\sqrt{t_+^3/2}}\right)\right).$$

Recall that for $|t_0| < t_+$ this approximates the partial sum of the Fourier series, otherwise it approximates the remainder of the Fourier series (with the opposite sign).

The next term in the saddle-point approximation would involve the next term of Taylor series of ${\mathrm{i}}/{t} + \mathrm{i}kt$ at $\pm t_+$. For this term, we can use

$$\int_{\mathbb{R}} \frac{1}{(t-t_1\mathrm{i})}t^3\exp -\frac{t^2}{2}\,\mathrm{d}t = \sqrt{2\pi}(1-t_1^2) + 2\pi t_1^3(1-\mathrm{Erf}\,t_1)\exp\frac{t_1^2}{2}$$

which decreases as $O({1}/{t_1^2})$ for $t_1 \gg 0$. Hence, as above

$$\int_{\mathbb{R}} \frac{1}{t-t_1}t^3\exp -\frac{t^2}{\beta}\,\mathrm{d}t = \sqrt{\pi\beta}\left(\frac{\beta}{2}+t_1^2\right) + 2\pi\mathrm{i}t_1^3\left(1-\mathrm{Erf}\left(-\mathrm{i}\frac{t_1}{\sqrt{\beta/2}}\right)\right)\exp -\frac{t_1^2}{\beta},$$

$$\int_{\mathbb{R}} \frac{1}{t-t_1}t^3\exp \mathrm{i}\frac{t^2}{\beta}\,\mathrm{d}t = \sqrt{\mathrm{i}\pi\beta}\left(\frac{\mathrm{i}\beta}{2}+t_1^2\right) + 2\pi\mathrm{i}t_1^3\left(1-\mathrm{Erf}\left(-\mathrm{i}\frac{t_1}{\sqrt{\mathrm{i}\beta/2}}\right)\right)\exp \mathrm{i}\frac{t_1^2}{\beta}.$$

Introducing $Y_m(t) := \sqrt{2\pi\mathrm{i}}\,(\mathrm{i}+t^2) + 2\pi\mathrm{i}t^3\left(m - \operatorname{Erf}\left(-\sqrt{\mathrm{i}}t\right)\right)\exp\mathrm{i}{}^{t^2}\!/2$ and $Y_\pm$, Y the same way as for E_m, the last integral is[455] $({}^{\beta}\!/2)^{3/2}Y\left({}^{t_1}\!/\sqrt{\beta/2}\right)$ when $t_1\in\mathbb{R}$ and the integral is over $(1+\varepsilon\mathrm{i})\mathbb{R}$ with $\varepsilon>0$. Moreover, $Y(t)$ is even continuous and decreases as $O({}^1\!/t^2)$ for large $t\in\mathbb{R}$. Likewise

$$\int_{\mathbb{R}}\frac{t-t_2}{t-t_1}t^3\exp\mathrm{i}\frac{t^2}{\beta}\,\mathrm{d}t = (t_1-t_2)\left(\frac{\beta}{2}\right)^{3/2}Y\left(\frac{t_1}{\sqrt{\beta/2}}\right).$$

Therefore the cubic terms in Taylor series for ${}^{\mathrm{i}}\!/t$ at $t=\pm t_+$ contribute

$$-\mathrm{i}t_0\sqrt{t_+/8}\left(\exp(2\mathrm{i}\sqrt{k})Y\left(\frac{t_0-t_+}{\sqrt{t_+^3/2}}\right) - \exp(-2\mathrm{i}\sqrt{k})Y\left(\frac{t_0+t_+}{\sqrt{t_+^3/2}}\right)\right),$$

to the remainder, and one should expect[456] that the whole remainder has the same order of magnitude. Note that for $|t_0|\lesssim t_+$ these terms have similar magnitude to the terms with $\widetilde{E}$ in the formula for the main saddle-point expression. Moreover, all these terms are majorated by the term with E provided $|t_0|\lesssim t_+\ll 1$.

Conclusion: the principal term in the saddle-point integral when $|t_+|, |t_+t_0^2|\ll 1$ is

$$\frac{t_+}{2\pi\mathrm{i}}\exp(-\mathrm{i}kt_0)\left(\exp(2\mathrm{i}\sqrt{k})E\left(\frac{t_0-t_+}{\sqrt{t_+^3/2}}\right) + \exp(-2\mathrm{i}\sqrt{k})E\left(\frac{t_0+t_+}{\sqrt{t_+^3/2}}\right)\right),$$

and for $|t_0|<t_+$ this approximates the partial sum of the Fourier series, otherwise it approximates the remainder of the Fourier series (with opposite sign). The error has relative magnitude $\sqrt{t_+}$.

The jump in E matches the total sum of the Fourier series for $t_0\approx t_\pm$; replacing E by E_+ (which has no jump) provides a good estimate on the right of t_+ if $t_0\approx t_+$. For larger values of t_0, one should take into account that the contour going across t_0 subtracts the value of $t_0\mathrm{e}^{\mathrm{i}/t_0}$ in the integral for the partial sum of the Fourier series, but subtracts a similar value of with the quadratic approximation to the exponent above in the saddle point integral. This quadratic approximation[457] differs by ${}^{(t_+-t_0)^3}\!/t_0t_+^3$.

To compensate for this, one needs to add t_0e^{i/t_0} and subtract $t_0\mathrm{e}^{\mathrm{i}/t_0+\mathrm{i}(t_+-t_0)^3/t_0t_+^3}$ if $t_0>t_+$. Moreover, taking an asymptotic expression for the second E-term above does not worsen the order of magnitude of the error:

$$\frac{t_+}{2\pi\mathrm{i}}\exp(\mathrm{i}k(2t_+-t_0))\left(E_+\left(\frac{t_0-t_+}{\sqrt{t_+^3/2}}\right) - \exp(-4\mathrm{i}\sqrt{k})\frac{\sqrt{\pi\mathrm{i}t_+^3}}{t_0+t_+}\right), \qquad -t_+/2\lesssim t_0\lesssim\left(1+O\left(\sqrt{t_+}\right)\right)t_+.$$

The error is of the relative magnitude $\sqrt{t_+}$ of the terms above. For larger t_0, to keep such precision one needs to add $t_0\mathrm{e}^{\mathrm{i}/t_0}(1-\mathrm{e}^{\mathrm{i}(t_+-t_0)^3/t_0t_+^3})$.)

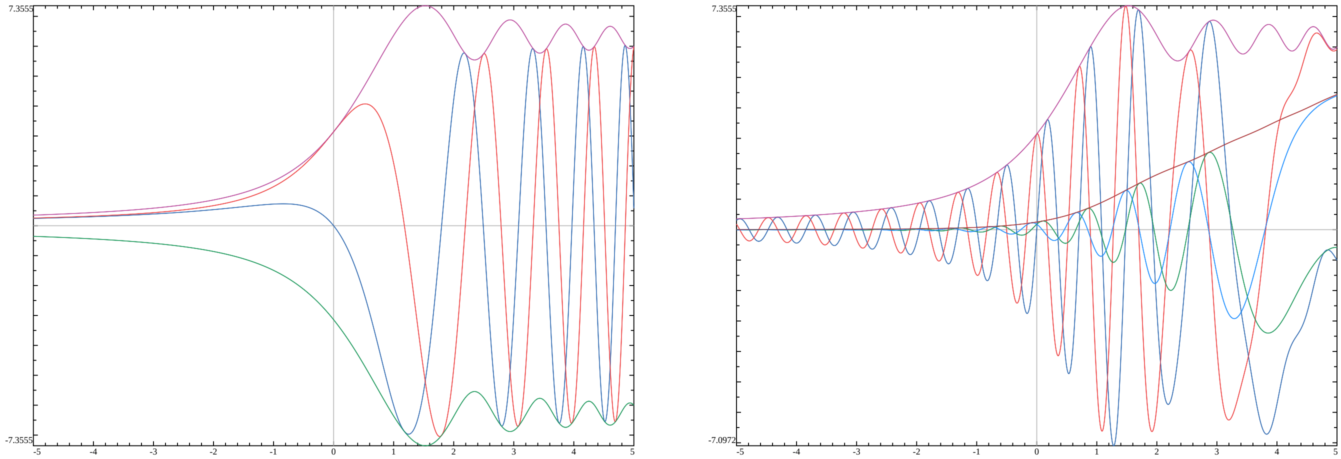
Let $\tau := (t-t_+)/\sqrt{t_+^3}$. When $|\tau|\lesssim 1$, the principal part of the expression above is one with $E_+(\tau\sqrt{2})$. Therefore, up to a phase factor, the partial sums of the Fourier series of $\mathrm{e}^{\mathrm{i}/t}$ have a "universal

[455] **N.B. (???) It looks like this should be easily generalizable?!**

[456] **N.B. (???) Check!**

[457] If $|t_0|$ is large, then for the saddle-point to work, the contour should go *above* t_0. In this case we need the formula *without* $\llcorner$s, and it approximates not the integral over the red contour, but its modifiction passing *above* t_0. This means, essentially, that the formula above should be corrected by a difference in value between $t_0\exp\mathrm{i}({}^1\!/t_0+kt_0)$ and the corresponding "saddle-point approximation", where we replace $\mathrm{i}({}^1\!/t+kt)$ by its quadratic approximation at $\pm t_+$ (one which is closer to t_0).

behaviour” near the point t_+ where the phase transform happens:



On the left plot, we show the real/imaginary parts of this universal behaviour (ignoring the phase factor), as well as its $\pm$magnitude.[458] The contribution of the phase factor depends[459] on k. On the right plot, the graphs of the same colors show (using the same colors) how the universal behavior combines with the phase factor when $k = 10{,}000$ (note that our graphs of fractally-symmetric functions correspond to $k \lesssim 10{,}000$).

It is easy to identify the red or blue graphs on the right picture with what one can see after zooming (a lot!) into the flattened zones of our plots of fractally-symmetric functions. However, there is a small, but clearly visible mismatch: in our plots, there is a definite growth of the amplitude of oscillations on the right of $\tau = 0$ (in other words, on the right of $t_0 = t_+$). The reason for this is that on the graph above $\tau \in [-5, 5]$, or $t_0 \in [t_+(1 - 5\sqrt{t_+}), t_+(1 + 5\sqrt{t_+})]$; recall that $t_+ = {}^1\!/\!\sqrt{k}$. With (relatively!) small values of k we have on our graphs, $\sqrt[4]{k} \lesssim 10$ which is very small (and comparable with the width of the domain of the plots above). Because of this, the magnitude of t_0 changes a lot on the interval in question — and the oscillations have a magnitude close to $|t_0|$.

Since this relative change of magnitude of oscillation is about ${}^1\!/\!\sqrt[4]{k}$, it is “negligible”, and is described by the term $\widetilde{E}$ we discarded above. This term is plotted above in light blue and green[460] [461] (for the same value $\sqrt[4]{k} = 10$).

Alternatively to using E_+, we can use E and add the residue

$$t_0 \exp \frac{\mathrm{i}}{t_0} = t_+ \exp \frac{\mathrm{i}}{t_+} \exp\left(-\frac{\mathrm{i}\tau}{\sqrt{t_+}}\right) \cdot \left(1 + \sqrt{t_+}\tau\right) \exp \frac{\mathrm{i}\tau^2}{1 + \sqrt{t_+}\tau} \qquad \text{when } t_0 = t_+(1 + \sqrt{t_+}\tau).$$

The factors before $\cdot$ match the factors in our formulas above. So define $E_{\mathrm{full}}(\tau, \alpha)$ as ${}^1\!/\!2\pi\mathrm{i} E_+(\tau\sqrt{2})$ if $\tau \le 0$, and ${}^1\!/\!2\pi\mathrm{i} E_-(\tau\sqrt{2}) + (1 + \alpha t) \exp \frac{\mathrm{i}\tau^2}{1+\alpha\tau}$ otherwise. Then the saddle-point approximation to our convolution is

$$t_+ \exp \frac{\mathrm{i}}{t_+} \exp\left(-\frac{\mathrm{i}\tau}{\sqrt{t_+}}\right) E_{\mathrm{full}}(\tau, \sqrt{t_+}).$$

---

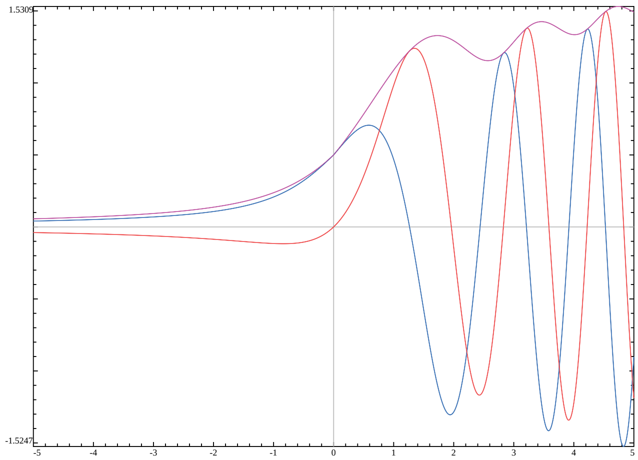
[458] Modulations appear since $E_+(t)$ for $t \gg 0$ is a sum of an oscillating term of constant magnitude and a decaying non-oscillating term. Therefore $|E_+|$ behaves as $|A\mathrm{e}^{\mathrm{i}f(t)} + {}^B\!/t| = |A + B\mathrm{e}^{-\mathrm{i}f(t)}/t|$, hence oscillates with the magnitude B.

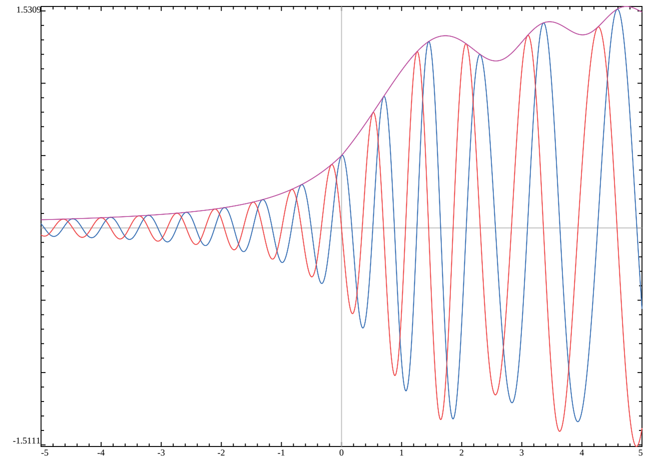
[459] Since $kt_+^2 = 1$, the extra phase term oscillates as $\exp -{}^{\mathrm{i}\tau}\!/\!\sqrt{t_+}$, so in the scale of the plot above it is a very quick oscillation, with the frequency about $\sqrt[4]{k}$.

[460] We also plot the magnitude of terms containing E and $\widetilde{E}$. One can see that the latter magnitude also has an oscillating term, but amplitude of these oscillations is microscopic.

[461] The “apparent” switch in the direction of oscillation on the right of the graphs is an artefact of our quadratic approximation to ${}^{\mathrm{i}}\!/t$. The more terms of Taylor series we keep, the further to the right this artefact would appear.

Two plots below are for E_{full}, as well as it combined with the phase factor $\exp\left(-{}^{\mathrm{i}\tau}/\sqrt{t_+}\right)$ (both with $\alpha = {}^1/_{10}$).





Note that $E_{\text{full}}(\tau, 0) = {}^1/_{2\pi\mathrm{i}} E_+(\tau\sqrt{2})$. Moreover, the first term in Taylor expansion of $E_{\text{full}}(\tau, \alpha)$ in α is similar to $\widetilde{E}_+(\tau)$ together with the contribution of the cubic term considered above.

Unfortunately, combining the factor $\exp -{}^{\mathrm{i}\tau}/\sqrt{t_+}$ above with $E_{\text{full}}(\tau, \alpha)$ when $\alpha := \sqrt{t_+}$ leads to counterintuitive results when one takes the Taylor approximation at $\alpha = 0$. The errors in the derivative of the phase become so large when $\alpha\tau \sim 1$ that the finite sum has *phase reversal*: the phase of the approximation changes the direction of rotation.[462]

Consider now a more general convolution $T \exp {}^{\mathrm{i}L}/_T \star \operatorname{sinc} KT$ in the variable T (above $L = 1$ and $K = k$). A coordinate change $t = {}^T/_L$ reduces this to the case $L = 1$, $k = KL$. Therefore the phase transition happens when $t = t_+ := {}^1/\sqrt{KL}$, or $T = T_+ := \sqrt{{}^L/_K}$. The width of the zone of phase transition around T_+ is $T_+\sqrt{t_+} = \sqrt[4]{{}^L/_{K^3}}$. The analysis above makes sense when $k \gg 1$, hence if it makes sense for $L = L_0$, it also makes sense for $L \geq L_0$.

Note also that the zone between the "±"-phase transitions for L_0 is deep inside the flattened zone for $L \gg L_0$. **Conclusion:** for a sum of a series in L, only the term which has the smallest $L = L_0$ would contribute significantly into the convolution when $|T| \lesssim \sqrt{{}^{L_0}/_K}$. If the Fourier series (with period $2\pi L_0$) for $G(t)$ has coefficients a_n, then in this zone for L_0, the contributions of a_1 is $a_1\sqrt{{}^{L_0}/_K}$ near the ends, and $a_1\sqrt[4]{{}^{L_0^3}/_{K^5}}$ near 0. The contribution of other Fourier terms has the order of magnitude $a_n\sqrt[4]{{}^{L_0^3 n^3}/_{K^5}}$.

Hence the contributions of $n > 1$ are products of

- a coefficient of magnitude $a_n\sqrt[4]{{}^{L_0^3 n^3}/_{K^5}}$;
- a quickly oscillating in T phase factor $\exp(-\mathrm{i}KT)$ (independent of n);
- an oscillating in n phase factor $\exp 2\mathrm{i}\sqrt[4]{nL_0K}$, and
- a slowly changing on $[-(1+\beta)T_+, (1+\beta)T_+]$ (here T_+ is taken for $n = 1$, and $0 < \beta \ll 1$) factor which is ${}^1/_{(1-\xi/\sqrt{n})}$ asymptotically in n; here $T = \xi T_+$.

Recall that in our fractally-symmetric cases $G(t)$ is an antiderivative of the Fourier transform of the sequence (N_n) which grows very slowly (recall that this sequence is related to modular solutions to a polynomial equation). Hence $a_n = {}^{N_n}/_n$.

Note that the variation of the slowly-changing factor is ${}^\xi/\sqrt{n}$ plus a remainder of magnitude ${}^1/_n$. Hence *this remainder* is given by an absolutely converging series, so it may be considered "small". We are going to ignore it in what follows.

[462] In the non-asymptotic formula for E_{full} the phase of $\exp \frac{\mathrm{i}\tau^2}{1+\alpha\tau}$ changes slower when $\alpha\tau$ grows to become ~ 1.— Compare this with the phase reversal in the first-order Taylor approximation in α.

Conclusion: in the flattened region the behaviour of convolution is controlled by two coefficients

$$\boxed{\sqrt[4]{L_0^3/K^5}\sum_n \exp\left(2\mathrm{i}\sqrt[4]{nL_0K}\right)\frac{N_n}{\sqrt[4]{n^m}}, \qquad m=1,3.}$$

at powers 1, ξ of ξ. If we can show that these coefficients are bounded in K, this would explain the appearence of flattened regions in the convolution $TG(1/T) \star \operatorname{sinc} KT$ near $T=0$ (as well as why these regions look very similar to the contribution of the first terms in the Fourier series for G).

These coefficients are, essentially, the Fourier transforms of $\sum_n N_n/\sqrt[4]{n^m}\,\delta_{\sqrt[4]{n}}$. As $\eta := \sqrt[4]{n}$ grows, the density of the points $\sqrt[4]{n}$ grows as η^3. Hence "smoothing" this generalized function leads to η^{3-m} times "the average value of N_n near $n=\eta^4$". Since the sign of N_n changes "chaotically", it is not surprising that the average value of N_n is small; however, to show that the latter Fourier transform is bounded, one needs to estimate this average value to be quite small (for example, for $m=1$ it is enough to show it to decay[463] quicker than $n^{-3/4}$ when averaging[464] over a region of size $n^{3/4}$).

Recall that *what is known* (for almost a century now) is the behaviour of the ζ-function, which is, essentially, the Fourier transform of $\sum_n N_n\,\delta_{\log n}$; this Fourier transform is an entire function. Note that taking $\log n$ "compresses" the points $n\in\mathbb{N}$ "much stronger" than taking $\sqrt[4]{n}$. In particular, the latter Fourier transform being smooth implies that averages of N_n over regions of size βn are rapidly decreasing in n. However, these regions are too large: their size is way larger than $n^{3/4}$ considered above!

On the other hand, we can use the properties of the Fourier transform $G(t)$ of the sequence N_n/n to estimate the required averages. It turns out that this is related to the rate of decay of the antiderivative of $G(t)$ near 0. Indeed, $G(t)$ is the Fourier transform of the generalized function $\widehat{G} := \sum_n N_n/n\,\delta_n$, and the averaged values of N_n on big regions are related to *convolutions* of $\widehat{G}$ with cut-off functions with *big* support; in turn, these are related to products of $G(t)$ with functions concentrated on a *small* region near 0. **Conclusion:** the convergence of framed series above is related[465] to the behaviour of $G(t)$ near 0.[466]

To handle the averaging described above, we essentially need to restrict G to a region of size $\varepsilon := 1/n^{3/4}$ near 0, and show that the nth Fourier coefficient of the restriction is less than $1/n^{1\frac34}$. However, near 0 we can use the fact that $G(t) = tH(1/t)$ with a periodic function H with the average value 0. The image of a period of H under $t\mapsto 1/t$ has a length which decreases quadratically when we get closer to 0; so these images inside $[-\varepsilon,\varepsilon]$ are shorter than $\varepsilon^2 \ll 1/n$.

Hence every period of $\sin nt$ in $[-\varepsilon,\varepsilon]$ is covered by many images of periods of H (hence zones over which the average of G is very close to 0). *This* is the reason why the scalar product with $\sin nt$ is small! Indeed, since on the image of any period of H the function $t\sin nt$ is almost constant, the integral of $H(1/t)\cdot t\sin nt$ is approximately proportional to the average value of $H(1/t)$ on this interval, which is approximately the average value 0 of H on any period.

The formal way to treat this argument is to work with "averages of G" — which is essentially determined by the antiderivative of G — and plug this into the formula of integration by parts. As a change of variable $t\mapsto 1/t$ shows, this antiderivative may be written as $t^3G_1(t)$ with a bounded function G_1. Hence instead of $\int_{-\varepsilon}^{\varepsilon}(t^3G_1(t))'\sin nt\,\mathrm{d}t$ we may consider $n\int_{-\varepsilon}^{\varepsilon}t^3G_1(t)\cos nt\,\mathrm{d}t$ which decays at least as $n\varepsilon^4 = 1/n^2$ indeed.[467]

---

[463] Note that this is equivalent to the corresponding average value of N_n/n decaying quicker than $n^{-1\frac34}$.

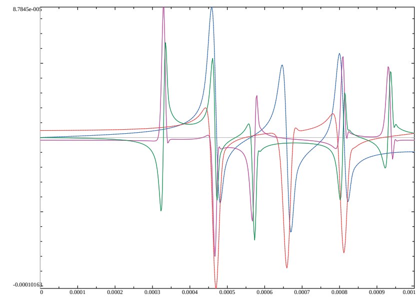
[464] ... with weight smoothly decaying to become 0 near the ends of the region.

[465] Recall that *the shape* of the flattened region (and nearby zones of the graph) is related to periodicity of G, hence to its behaviour near ∞.

[466] This property is automatically satisfied if $tG(1/t)$ is periodic with the average value 0.

[467] In fact, one can continue integrating by parts; this would show that N_n/n averaged over regions of width n^β with $\beta > 1/2$ would rapidly decrease. Hence the same holds for N_n.

This finishes our considerations of flattened zones of $TG({}^1\!/T) \star \operatorname{sinc} KT$ when G is periodic and its Fourier coefficients decay sufficiently quick when averaged appropriately. (Recall[468] that in our examples $L_0 = {}^{4\pi^2}\!/c$ where c is the conductor, and K is at most 16 millions; this is why $k = {}^K\!/L$ is often of order of tens of thousands.)



Remark 103: [469] Note that this explanation works for some non-periodic functions G too; to have a flattened region in $|T| \le T_+$ with K terms of the Fourier series, all we need[470] is the Fourier transform $\widehat{G}(L)$ of G vanishing up to $|L| \approx T_+^2 K$. This may be relevant for the graphs below related to counts of modular roots of polynomials of degree 4; we obtain graphs visually similar to $TG({}^T\!/t)$ with a non-periodic — but very slowly growing — oscillating function G. Since these graphs have clearly visible flattened parts, this suggests that $\widehat{G}(\tau)$ may vanish for τ close to 0. This leads to the question:

$$\boxed{\text{What can we deduce about } \widehat{G} \text{ knowing that flattened parts exist for } K \gg 0?}$$

In this remark we provide a **failed** attempt to do so (kind of fixed in the next remark, which shows that the information about the non-descrete spectrum is actually *lost* in this approach!).

So we omit the assumption that G in $H(T) = TG({}^1\!/T)$ is periodic; now its Fourier transform $\widehat{G}$ (which is a generalized function) is not necessarily a sum of δ-functions δ_{nL_0} at points of $L_0\mathbb{N}$; still, it may be considered as *an integral* of such δ-functions with a certain weight. The asymptotics above describe very well a contribution of one of these δ-functions δ_L in the region $T \in [-(1+\Delta)T_+, (1+\Delta)T_+]$ with $\Delta > 0$, however, before, for *a detailed* consideration we were focussing only on a zone of width about $T_+^{3/2}/\sqrt{L}$ around $\pm T_+$. Since $T_+ = \sqrt{{}^L\!/K}$ depends on L, to consider an integral in L, we need to consider the whole zone $T \in [-(1+\Delta)T_+^{\max}, (1+\Delta)T_+^{\max}]$.

Moreover, to make the description more transparent, we simplify $T_+ = \sqrt{{}^L\!/K}$ by a substitution $L = \lambda^2$. So our goal is to re-analyze our asymptotics for a contribution of one δ-function δ_{λ^2} in an interval of T around 0. Furthermore, the asymptotic series we discussed was in variable $\sqrt{t_+}$ with ${}^1\!/t_+ = \sqrt{KL}$; so write $K = {}^1\!/\varepsilon^4$. Now $\varepsilon \ll 1$. Recall that our asymptotics work for $t_+ \ll 1$, or $\varepsilon^2 \ll \lambda$.

Hence we need to study the convolution[471] $\widetilde{G}_\varepsilon := TG({}^1\!/T) \star \operatorname{sinc} {}^T\!/\varepsilon^4$ with $G(T) = \exp \mathrm{i}\lambda^2 T$. In fact, our asymptotic expressions for this convolution all contain a very quickly oscillating factor $\exp(-{}^{\mathrm{i}T}\!/\varepsilon^4)$; so it *seems* better to investigate $\widetilde{G}_\varepsilon \exp {}^{\mathrm{i}T}\!/\varepsilon^4$, where this factor is cancelled; we start with investigating this approach. Moreover, we saw that for periodic G the width of the flattened zone decreases with ε as $C\varepsilon^2$; the corresponding rescaling $T = \varepsilon^2 x$ leads to consideration of $\widetilde{\widetilde{G}}_\varepsilon(x) := \widetilde{G}_\varepsilon(\varepsilon^2 x) \exp {}^{\mathrm{i}x}\!/\varepsilon^2$.

It turns out that in this context one of the simplest term of our asymptotics is the "remainder" term (due to non-quadratic Taylor terms under the exponent). Recall that it was[472]

$$-\mathrm{i}t_0\sqrt{t_+/8}\, Y\!\left(\frac{t_0 - t_+}{\sqrt{t_+^3/2}}\right) \exp(2\mathrm{i}\sqrt{k} - \mathrm{i}KT) = -\frac{\mathrm{i}}{\sqrt{8}}\varepsilon^3 x\, Y\!\left(\frac{x-\lambda}{\varepsilon\sqrt{\lambda/2}}\right) \exp\frac{\mathrm{i}(2\lambda - x)}{\varepsilon^2}$$

(However, an honest proof would need to also consider what happens outside $[-\varepsilon, \varepsilon]$: the cut-off function is small there, but one needs to estimate *how* small we can make it while keeping its Fourier transform supported between $\pm n^{3/4}$.)

[468] **N.B. (???) Move?**

[469] **N.B. (???) The plots should be moved 1 page later — but `wrapfig` does not work inside `remarks`.**

[470] … in addition to decay of the antiderivative of G near 0…

[471] **N.B. (???) In fact, below we consider only the integral over a half-line!**

[472] **N.B. (???) Coefficient 2π?**

for $k \gg 1$, with $t_+ = \varepsilon^2/\lambda$, $t_0 = T/\lambda^2$ and $k = 1/t_+^2$. For $L \neq 1$ one needs an extra factor $L^{-1/4} = 1/\sqrt{\lambda}$. Removing the oscillating in T factor, we obtain the contribution $\mathrm{const} \cdot \varepsilon^3 x\, Y\bigl((x-\lambda)/\varepsilon\sqrt{\lambda/2}\bigr) \exp 2\mathrm{i}\lambda/\varepsilon^2$ into $\tilde{\tilde{G}}_\varepsilon(x)$.

For any function $y(t)$ with integral y_0 and the asymptotic C/t^2 at infinity, the family $y_m(t) := my(mt)$ has the asymptotic expansion $(y_0 - C/m)\delta_0 + C/m \cdot 1/t^2 + O(1/m^2)$ as $m \to \infty$ (in the space of generalized functions[473]). In the formulas above, the denominator inside $Y(\dots)$ goes to 0 when ε decreases, so the main term in ε in this formula is $\mathrm{const} \cdot \varepsilon^4 \sqrt{\lambda} \exp(2\mathrm{i}\lambda/\varepsilon^2) x \delta_\lambda$.

Since this was the contribution of δ_{λ^2} in the Fourier transform $\widehat{G}$ of G, one may have tried to restate it as

$$\tilde{\tilde{G}}^{\mathrm{cubic}}_\varepsilon(x) \overset{?}{=} \mathrm{const} \cdot \varepsilon^4 \exp(2\mathrm{i}x/\varepsilon^2) x^{3/2} \widehat{G}(x^2) + O(\varepsilon^5)$$

as a generalized function. Unfortunately, multiplication by a quickly oscillating factor (depending on parameter) is not continuous in the topology of generalized functions. (For example, $\varphi_\varepsilon(x) := \exp \mathrm{i}x/\varepsilon$ converges to 0 when $\varepsilon \to 0$, but multiplying this by $\exp -\mathrm{i}x/\varepsilon$ gives 1.) Hence we need to compensate the factor $\exp(2\mathrm{i}x/\varepsilon^2)$ *before* we take the limit.

Note that the removal of the oscillating term leads to oscillations starting to appear above $t_0 \sim t_+ + \sqrt{t_+}$. Hence *if we can kill oscillations by an appropriate smoothing of our function*, the smoothing is going to decay on the right of t_+ as well as it decays on the left of it (here even without smoothing!).

For example, above, at the beginning of this remark, we plotted the case of the smallest (in magnitude) cubic conductor 23 (for $M = 6$) with a cancellation of the phase factor as above, and a weakest possible smoothing to kill the oscillations (the real and imaginary parts, the dashed lines for 8 and solid for 16 million Fourier terms).[474] [475]

The graphs have maxima for $\mathrm{const} \cdot \sqrt{n}$ for $n \in \mathbb{N}$ such that $N_n \neq 0$. For the cubic conductor 23, up to $n = 12$ these are 1, 2, 3, 6, and 8. When we have two graphs with K differing 2 times, the positions of the maxima (denoted as T_+ above) for n and the larger K is the same as for $2n$ and the smaller K.

Conclusion: one can *observe the discrete spectrum* of the function $G(t)$ when one

- Cuts off the high frequencies from $tG(1/t)$;
- Cancels the oscillations near 0 (by multiplying by an appropriate oscillating function);
- Cuts off high frequencies again. (With cut-off frequency tens times smaller than one on the first step.)[476]

(Below we show that the continuous spectrum of $G(t)$ gives only a vanishingly small contribution. — And the smoother it is, the smaller the contribution — provided the cut-off frequency is high.)

---

[473] Recall that the generalized function $1/t^2$ is normalized by $\int_{-1}^{1} dt/t^2 = 0$.

[474] We convolve with the Gaussian kernel with half-width $35/\pi$ of the period of the cut-off oscillation. So for 16 million terms, the half-width is 4.4e-6.

This smoothing is the minimal possible to get at least one maximum on the graph without spikes. (This is the first maximum on the dashed lines.) The other maxima have very visible spikes; the further we go from 0, the more pronounced are the spikes. (To avoid the spikes, the smoothing should depend on the cut-off, and be stronger further away from 0.)

[475] **N.B. (???) Estimate the necessary smoothing! Note that k for these graphs has an unusually high value: 27.5 millions for the higher value of the cut-off frequency.**

[476] The final result of this process is that if Fourier coefficients were a_n, we sum up $\sum_n = 0^{n_0} a_{n_0-n}\sigma(n/n_1)\mathrm{e}^{-nt}$, with $\sigma(x) \sim \mathrm{e}^{-x^2}$, and $1 \ll n_1 \ll n_0$. When one replaces this σ by a step cut-off function, this still results in a certain “smoothing” of “the sum without the factor σ”. However, such a σ results (up to an oscillating factor) in the difference between partial sums of the *intial* Fourier series for cut-off frequencies n_0 and $n_0 - n_1$.

Conclusion: this difference should behave similarly to the graph above. (Recall that the graphs in the main part of these notes *combine* two partial sums, one drawn in blue, another in red, so the places where the difference is significant are visible as blue species on the graph.) In particular, this explains why the blue species tend to appear in groups.

Therefore we define $G_\varepsilon(x) := \widetilde{G}_\varepsilon(\varepsilon^2 x)\exp(-{}^{\mathrm{i}x}/_{\varepsilon^2})$. Now the contribution of the "cubic" term is

$$\mathrm{const}\cdot\varepsilon^3 x\, Y\left(\frac{x-\lambda}{\varepsilon\sqrt{\lambda/2}}\right)\exp\left(\frac{2\mathrm{i}(\lambda-x)}{\varepsilon^2}\right).$$

Now for $\varepsilon \ll 1$ the oscillating term has period much smaller than $\varepsilon\sqrt{\lambda^3/2}$, hence the limit when $\varepsilon\to 0$ is described very differently than before. Consider the limit of $my(mt)\mathrm{e}^{\mathrm{i}mnt}$ when $m,n\to\infty$. If the Fourier transform of $y(t)$ is $\widehat{y}(\tau)$, then the Fourier transform of $y(t)\mathrm{e}^{\mathrm{i}nt}$ is $\widehat{y}(\tau+n)$, and the Fourier transform of $my(mt)\mathrm{e}^{\mathrm{i}mnt}$ is $\widehat{y}({}^{\tau}/_{m}+n)$. **Conclusion:** this Fourier transform becomes[477] "more and more constant" as $m,n\to\infty$, hence $\lim my(mt)\mathrm{e}^{\mathrm{i}mnt}/\widehat{y}(n) = \delta(t)$.

Since the asymptotic of $\widehat{y}$ depends on the singularities of y, what is important about Y is that it has a jump of the third derivative at 0; hence its Fourier transform $\widehat{Y}(\tau)$ decays as $\mathrm{const}\cdot\tau^{-4}$. Therefore the expression above has the limit $\mathrm{const}\cdot{}^{\varepsilon^8}/_{\sqrt{\lambda}}\delta_\lambda$.

Likewise, $E_{\mathrm{full}}(t,0)$ has a jump of the third derivative at 0, and $E_{\mathrm{full}}(t,\alpha)$ for $\alpha\neq 0$ has a jump of the first derivative proportional to α. Hence the expression for the main term of the convolution

$$t_+\exp\frac{\mathrm{i}}{t_+}\exp\left(-\frac{\mathrm{i}\tau}{\sqrt{t_+}}\right)E_{\mathrm{full}}(\tau,\sqrt{t_+}) = \frac{\varepsilon^2}{\lambda}E_{\mathrm{full}}\left(\frac{x-\lambda}{\varepsilon\sqrt{\lambda/2}},\frac{\varepsilon}{\sqrt{\lambda}}\right)\exp\frac{\mathrm{i}(2\lambda-x)}{\varepsilon^2}$$

after multiplication by the correcting phase factor $\exp-{}^{\mathrm{i}x}/_{\varepsilon^2}$ has a limit $\mathrm{const}\cdot{}^{\varepsilon^6}/_{\lambda^2}\delta_\lambda$.

Conclusion: the main term[478] of asymptotic of $G_\varepsilon(x)\exp-{}^{\mathrm{i}x}/_{\varepsilon^2}$ in $\varepsilon\to 0$ *as a generalized function* is $\mathrm{const}\cdot{}^{\varepsilon^6}/_{x^2}\widehat{G}(x^2)$. Hence calculating the "averaged"[479] behaviour of the phase-corrected convolution $(TG({}^{1}/_{T})\star\operatorname{sinc}KT)\exp(-\mathrm{i}KT)$ with precision much better than $K^{-3/2}L^{-2}$ in a region of width $2\sqrt{L/K}$ about 0 gives information about the Fourier transform $\widehat{G}(\tau)$ of G for[480] $\tau\in[0,L]$.

In particular, the argument above shows[481] that this function of T is $o(K^{-3/2})$ for $|T|<\sqrt{L/K}$ when the cut-off frequency $K\to\infty$ if and only if $\widehat{G}(\tau)$ vanishes for $\tau<L$.

Remark 104: In fact, the calculations above were tacitly exchanging the order of two limits — which is a no-no-no in analysis (as far as one can do it honestly — otherwise the answer may still have some *heuristical* value!). To proceed honestly, note that $\int_0^\infty x\mathrm{e}^{-A({}^{1}/_{x}+x)}\mathrm{d}x$ is transformed by a coordinate change ${}^{1}/_{x}+x=2y$ to[482] $\int_1^\infty(4\sqrt{y^2-1}+{}^{2}/_{\sqrt{y^2-1}})\mathrm{e}^{-2Ay}\mathrm{d}y = 2K_0(2A)+2{}^{K_1(2A)}/_{A}$. Hence the Fourier transform of the function $t\mathrm{e}^{-B/t}$ with support on the positive semiaxis is ${}^{2B\mathrm{i}}/_{\tau}(K_0(2\sqrt{-\mathrm{i}B\tau})+{}^{K_1(2\sqrt{-\mathrm{i}B\tau})}/_{2\sqrt{-\mathrm{i}B\tau}})$. For $B=-\mathrm{i}$ this is ${}^{2}/_{\tau}(K_0(-2\mathrm{i}\sqrt{\tau})+{}^{\mathrm{i}K_1(-2\mathrm{i}\sqrt{\tau})}/_{2\sqrt{\tau}})$ with $\operatorname{Im}\sqrt{\tau}\geq 0$.

Denote by $G_{\lambda,\varepsilon}(t)$ the function $G_\varepsilon(t)$ for $G(t)=\exp\mathrm{i}\lambda^2 t$. The asymptotics above show that to analyse dependence of G_ε on the continuous part of the Fourier transform of G, it makes sense to consider the limit of $G_{\lambda,\varepsilon}(x\varepsilon^2)\mathrm{e}^{x/\varepsilon^2}$ as a generalized function[483] of λ.

Recall that often a limit $\lim f_n$ in the topology of generalized functions corresponds to a pointwise limit of the Fourier transforms.[484] This leads to investigating the asymptotic in ε of the Fourier

---

[477] … unless $\log\widehat{y}$ changes pathologically quick near ∞.

[478] **N.B. (???) Check the term for $-T_+$.**

[479] **N.B. (???) Explain!**

[480] **N.B. (???) Prohibit negative τ?**

[481] **N.B. (???) Heuristically only, since we tacitly exchange the order of limits!**

[482] Here K_0, K_1 are the Bessel functions.

[483] Then if this limit exists in a certain Banach topology (of the scale of Banach topologies on the space of generalized functions), then we can predict the point-wise behaviour of $G_\varepsilon(x\varepsilon^2)$ when the Fourier transform of G is in the dual Banach space.

(Note that when δ-functions are not in the dual space, the formulas may look very different in the case of periodic G and the case of the preceding paragraph!)

[484] Since the operator of Fourier transform is continuous on the space of (tempered) generalized functions, the limit above would lead to a limit of the Fourier transforms *in the sense of generalized functions*. However, if the functions f_n

transform of $G_{\lambda,\varepsilon}\mathrm{e}^{x/\varepsilon^2}$ in λ (with $t, x > 0$):

$$\int \frac{t}{t - x\varepsilon^2} \exp \mathrm{i}\left(\frac{\lambda^2}{t} + \frac{t}{\varepsilon^4} + \lambda\mu\right) \mathrm{d}t\, \mathrm{d}\lambda = \varepsilon^2 \int \frac{T}{T - x} \exp \frac{\mathrm{i}}{\varepsilon^2}\left(\frac{\lambda^2}{T} + T + \varepsilon^2\lambda\mu\right) \mathrm{d}T\, \mathrm{d}\lambda$$

(with the paths of integration in t and T going below the poles at $t = x\varepsilon^2$ and $T = x$). The integral in λ gives

$$\sqrt{\mathrm{i}\pi}\varepsilon^3 \int_0^\infty \frac{T^{3/2}}{T - x} \exp \mathrm{i}T\left(\frac{1}{\varepsilon^2} - \tfrac{1}{4}\varepsilon^2\mu^2\right) \mathrm{d}T = \sqrt{\mathrm{i}\pi}\varepsilon^3 x^{3/2} \int_0^\infty \frac{\tau^{3/2}}{\tau - 1} \exp \mathrm{i}\tau x\left(\frac{1}{\varepsilon^2} - \tfrac{1}{4}\varepsilon^2\mu^2\right) \mathrm{d}\tau.$$

Recall that[485] $\int_0^\infty t^A/(t+B)\; \mathrm{e}^{-pt}\mathrm{d}t = \Gamma(A+1)B^A\Gamma(-A, Bp)\mathrm{e}^{Bp}$ if $B \notin \mathbb{R}_{\leq 0}$ with $\Gamma(k, z)$ being the *incomplete* Γ-*function* $\int_z^\infty t^{k-1}\mathrm{e}^{-t}\mathrm{d}t$. We should take the limit value when $B = -1 - 0\mathrm{i}$, giving $-\tfrac{3}{4}\sqrt{\mathrm{i}\pi}\Gamma(-3/2, -p)\mathrm{e}^{-p}$.

Considering $u(z) := \Gamma(-3/2, z)\mathrm{e}^z$, this leads to

$$-\tfrac{3}{4}\mathrm{i}\pi\varepsilon^3 x^{3/2} u\left(\frac{\mathrm{i}x(1 - \tfrac{1}{4}\varepsilon^4\mu^2)}{\varepsilon^2}\right).$$

Recall that $u(z)$ has the main terms of asymptotic[486] $3/2 \cdot z^{-3/2}$ for $z \approx 0$ and $z^{-5/2}$ for $|z| \gg 0$, and has the corresponding asymptotic series expansion. In particular, all terms in Taylor expansion in ε are polynomials in μ.

Conclusion: For a fixed x, $G_{\lambda,\varepsilon}(x\varepsilon^2)\mathrm{e}^{x/\varepsilon^2}$ has a Taylor expanstion in ε as a generalized function in λ. The Taylor coefficients are polynomials in x with coefficients being (even) derivatives of δ-functions[487] at 0 in[488] λ. Hence the contribution of continuous parts of the Fourier transform of G into G_ε (away from the 0 in the frequency region) decay quickier than any power of ε. In other words: observing $G_\varepsilon(t)$ for small t gives immediate[489] information about the discrete spectrum of G, but not the continuous spectrum.

ζ-functions

So far, we mentioned several times (but completely ignored all the details) one of the major players on the scene of investigation of “arithmetic sequences” (such as our numbers $\widetilde{N}_n^{\mathrm{res}}$, or better, $\widetilde{N}_n^{\mathrm{Gal}}$): the ζ-functions. (Sometimes these functions are called L-functions.) Our approach was to encode the sequence using its Fourier transform $F(t)$; however, some of the properties of the sequence become much more accessible if one uses a different way to encode this sequence into a function.

Start with returning to the context of Remark 38 on p. 61, where we began investigating the statistical properties of the numbers $\widetilde{N}_n^{\mathrm{res}}$ for the polynomial “tetrahedral numbers + 2”. (These numbers are related to the sequence of colors,

1 **2** **3** 4 **5** 6 **7** 8 9 10 **11** 12 **13** 14 15 16 **17** 18 **19** 20 21 22 **23** 24 25 26 27 28 **29** 30 **31** 32 33 34 35 36 . . .

have nice asymptotics at ∞, the Fourier transforms are going to be continuous (outside of a few points); moreover, if the asymptotics are uniform, then the limit is going to exist locally in C^0 (outside of the exceptional points).

This is why the limit of Fourier transform is going to exist pointwise so often!

485 Formula 9.162 in Ditkin–Prudnikov’s *Integral transforms and operational calculus*.

486 In fact, $u(z) = U(5/2, 5/2, z)$ with U the Tricomi’s *confluent hypergeometric function*.

487 **N.B. (???) Is it possible to compare this with our asymptotics in the case of periodic G?!**

488 Paired with $\widehat{G}(\lambda^2)$, they give the derivatives of $\widehat{G}(L)$ at 0, which are (regularized!) pairings of $G(t)$ with powers of t. In turn, these are regularized integrals $\int tG(1/t)t^{-k}\mathrm{d}t$ for $k \geq 3$. If $tG(1/t)$ is periodic, these integrals can be expressed as sums involving the Fourier coefficients — and these sum is easy to recognize as values of the corresponding ζ-function at $-k$.

Moreover, the Hecke functional equation shows that in the “modular form” case these values vanish; in the Maass case, they vanish for even k. **N.B. Maybe the contributions of two saddle points in the Maass case cancel each other?! Check! ???**

489 **N.B. (???) Discuss smoothed “reversed” graphs we made?**

on p. 19.)

With the discriminant being -4×971, for all prime numbers except 2 and 971 the count $\widetilde{N}_n^{\mathrm{res}}$ is one of 0, 1 or 3 (with 0 related to the "red" color above).[490] It turns out that the count 3 appears less often than the others; in the part of the colored sequence shown above (on p. 19), it appears only for the prime 3. The first few other occurrences are for the primes 37, 61, 83, ...

Remark 105: As we mentioned in Remark 48 on p. 77, the Chebotaryov's density theorem predicts that asymptotically, exactly ⅓ of the primes are going to be red, half of the primes is assigned the count 1, and the remaining ⅙ is assigned the count 3. Hence asymptotically, one of four green primes is assigned the count 3.

On the other hand, 3 is the second of the green primes, 37 is the 6th, 61 is the 10th, and 83 is the 15th. For the "1 out of 4" law, the "expected" positions are 2½, 6½, 10½, 14½ etc. **Conclusion:** the observed prime numbers with count 3 follow the expected law $4n - 1½$ — however, from the statistical point of view, it is suspicious that they follow this law "too close"!

Well, I know no nice explanation for this, however, the plot thickens yet more for larger primes — and *that* has an explanation! Observe the following table (for $p < 600$):

Prime with count=3	3	37	61	83	151	167	257	263	281	353	389	409	433	457	461	563	...
Position among green primes	2	6	10	15	23	26	37	38	41	46	51	53	56	59	60	67	...
"Expected" position	2½	6½	10½	14½	18½	22½	26½	30½	34½	38½	42½	46½	50½	54½	58½	62½	...

For primes $p \geq 83$, the mismatch is always in the same direction: the primes with count 3 are consistently rarer than expected from Chebotaryov density theorem! Moreover, typically the mismatch is quite large...

Conclusion: from this data, it *seems* that for primes up to 300, the fraction is closer to 1 in 4⁵⁄₉; up to 600, it is 1 in 4.375. For primes up to 1,000, this decays again to less than 1 in 4½. Going up to 10,000, this proportion is still above 4⅛. Only when going to 100,000 the proportion starts goes down almost to 4: 1 in 4.038.

Such an effect is well-known as "a large bias for small primes".[491] The larger are the prime numbers, the smaller the bias. However, there is another effect which is independent of the magnitude of primes!

Observe for which numbers p_0 the proportion above taken for the primes $p \leq p_0$ is larger than the average value ¼, and for which it is smaller. It turns out that the first situation appears *much* more frequently than the second one — and that such numbers p_0 appear in long groups ("runs") separated by relatively narrow intervals.

A calculation shows the runs $83 \leq p_0 \leq 4{,}877$ and $5{,}669 \leq p_0 \leq 30{,}029$, then $73{,}243 \leq p_0 \leq 748{,}597$ and $811{,}387 \leq p_0 \leq 5{,}371{,}783$, $5{,}384{,}887 \leq p_0 \leq 15{,}954{,}779$ and $17{,}588{,}033 \leq p_0 \leq 24{,}622{,}691$, then $28{,}655{,}951 \leq p_0 \leq 581{,}827{,}369$, then $729{,}069{,}533 \leq p_0 \leq 1{,}045{,}122{,}007$ and $p = 2{,}341{,}051{,}859 \leq p_0 \leq 9{,}447{,}449{,}639$, etc.[492] The simplest incarnation of this effect is known as "prime races". In these races one compares count of primes $p \leq p_0$ which are $a \mod D$ to a similar count for $b \mod D$ for chosen a, b and D. Same as in our case, the effect *does not decay* as p_0 grows: for some triples (D, a, b) one of the classes "wins" (has more elements) for much longer runs of p_0 than the "intervening intervals".

[490] A more detailed analysis shows that 2 is not exceptional. To see this, one needs to consider the "*field discriminant*" instead of the discriminant of the polynomial — it turns out to be −971. Compare with Footnote 440 on p. 144.

Moreover, $\widetilde{N}_{971}^{\mathrm{res}} = 2$, so it is truly "exceptional". Since 971 is so large, treating 971 as "non-exceptional" (as we do in this section) does not affect the statistical properties described below.

[491] In a situation very close to one we consider in these notes, when Kevin Buzzard notices that "Non-zero a_p are rather sporadic and take a while to start", he just comments that this happens "for the usual reasons".

[492] There is another run starting at 13,103,095,877 and going at least to $2 \cdot 10^{10}$. (These calculations take about 5 hours on an old slow notebook using the code at the end of these notes.)

Note that most of the intervening intervals are quite short ("relatively", — or in log-scale). Two notable exceptions are the intervals after 30,029 and after 1,045,122,007.

Anyway, *on average* the "runs" are (relatively) much wider than the intervening intervals.

While the "inequality" in prime races mod 4 (which is the simplest precursor of the effect we observed above) was first noticed in 1853 by Chebyshyov, the appearence of "very long runs" (as above) in prime races were first quantitatively explained by Sarnak and Rubinstein in 1994). Moreover, it turns out that the prime races are directly connected to sequences $\widetilde{N}_n^{\mathrm{res}}$ related to modular roots of *abelian* polynomials.[493] The generalization good enough for our sequences $\widetilde{N}_n^{\mathrm{res}}$ was investigated in the Ph.D. thesis of Nathan Ng in 2000.

The simplest way to formulate these results is to select two types of prime numbers: in Chebyshyov case, these are primes which are $1 \mod 4$, and $-1 \mod 4$; in our case, it is primes p with $\widetilde{N}_p^{\mathrm{res}} = 3$, and $\widetilde{N}_p^{\mathrm{res}} = 1$. These types should be related to the Chebotaryov's density theorem. (In other words, the type should be determined by a certain cycle decomposition of the permutation we associated to a prime number in Footnote 322 on p. 118.[494]) Then one compares "the relative size" of these classes: calculate how many prime numbers up to p_0 are in these classes, as a function of p_0, and take a quotient of these functions.

As Riemann–Weil explicit summation formula shows[495] (in the context of Artin L-functions), the relative size is asymptotically $C_0 - (C_1 + C_2 + C_3(\log p_0) + O(1/\log p))p_0^{-1/2}$. Here C_0, C_1 and C_2 are constants, while C_3 is an oscillating function with average 0.[496]

The Chebotaryov theorem describes C_0: in the main term, the relative size of a type is proportional to the size of the corresponding conjugacy class in the Galois group.[498] In addition to this, Ng described the term C_1 as the size of the preimage of the conjugacy class under the map $g \mapsto g^2$ of the Galois group.

In the plot below, we cancel the term C_0 by taking $n_{\mathrm{green}} - 4n_3$, with $n_{\mathrm{green}}(p_0)$ the count of green primes up to p_0, and $n_3(p_0)$ the similar count of primes p with $N_p = 3$. We also rescale appropriately to In this case, $C_2 = 0$ for both types of primes, and C_1 are ???[499]

[493] For example, the race discovered by Chebyshyov is related to possible divisors of numbers $n^2 + 1$.

[494] This description assumes that the corresponding Galois group is the full group S_d of permutations of the roots of a certain polynomials of degree d. In general, one considers the *conjugacy classes* in the Galois group. — One can also consider unions of such classes.

[495] **N.B. (???) More details!**

[496] Moreover, C_3 is not only oscillating, but is in a fact a precisely described *almost periodic function*.[497] It is the description of terms C_2 and $C_3(\tau)$ where the discussion of ζ-functions is not avoidable.

In turn, C_2 is easy to calculate in terms of the ζ-function (the multiplicity of the root at ½) — and for most of polynomial equations, C_2 is expected to be 0. On the other hand, the properties of $C_3(\tau)$ are controlled by the other roots of ζ, hence by the Grand Riemann Hypothesis.

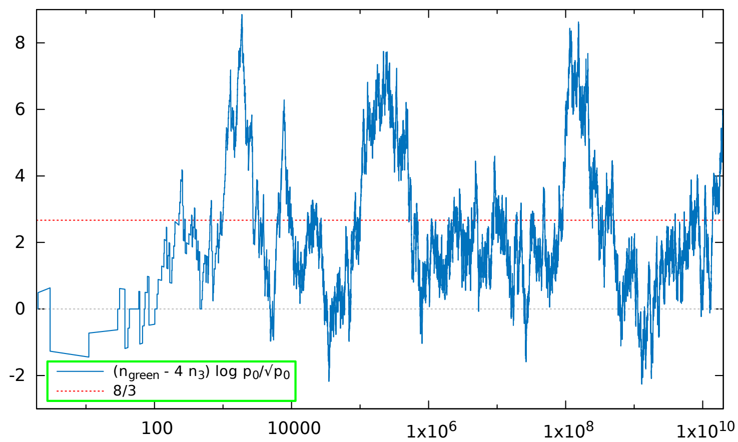
[497] Almost periodicity is conditional (and in fact equivalent!) on the Grand Riemann Hypothesis.

Moreover, one should keep in mind that oftentimes, in their definitions of almost periodic functions, people require that the *coefficients* are in ℓ_1. (This ensures that the function is continuous and bounded.)

However, the class of almost periodic functions we need is much larger: the coefficients are in ℓ_2. (For example, these functions have a jump at $\log p$ for many prime numbers p, so cannot be continuous.)

[498] For example, in S_3 there are 3 conjugacy classes: {id} (of relative size ⅙), "transpositions" (of relative size ½), and "3-cycles" (of relative size ⅓). These matches the values of $\widetilde{N}_p^{\mathrm{res}}$ of 3, 1 and 0 (as the numbers of fixed points).

[499] **N.B. (???) More details: zeros of L-function, Dyson's quasi-crystals, log-scale-Fourier=Dirichlet. The observed average is closer to 7/3, not 8/3? Almost-periodicity?**



8
6
4
2
0
-2
100
10000
1x10^6
1x10^8
1x10^10
(n_green - 4 n_3) log p_0/√p_0
8/3

Appendix: On verification, — and the future

In construction! (Lousy — and not fully complete [see N.B.] — exposition.)

The adelic completion

As we saw in Remark 30 on p. 47, combining 2π-periodicity of $F(t)$ with its fractal symmetry at $t = 0$, we can get a big collections of fractal symmetries of $F(t)$ which give a huge "hyper-family" of horizon-self-similar points. Since the fractal symmetry at $t = 0$ is a corollary of Hecke's functional equation, which was known decades before the Langlands program, we can treat these horizon-self-similar points as "the trivial cases" of Langlands program.

For cubic equations, what the Langlands program claims is that *every* rational multiple of π is a horizon-similar point, and that "a big fraction" of the set of these points is horizon-self-similar. A significant part of this statement is the claim that the tensor field $F(t)(\mathrm{d}t)^{½}$ is preserved by strongly-congruence fractional-linear transformation.[500] In other words, if c is the conductor, $A, B, C, D \in \mathbb{Z}$ and $A - 1, C, D - 1$ are divisible by c, then the transformation $T \mapsto {}^{(AT+B)}\!/_{(CT+D)}$ preserves $F(t)(\mathrm{d}t)^{½}$; here $T = 2\pi t$.

Note that $t = 0$ being horizon-similar is, essentially, equivalent the coordinate change $T \overset{\varphi}{\mapsto} -{}^{1}\!/_{cT}$ in $F(T)$ leading to a periodic function. In other words, F is preserved by the transformation ${}^{1}\!/_{T} \mapsto {}^{1}\!/_{T} + c$. Together with the periodicity of F, these two symmetries, when combined, lead to a giant group of "trivial symmetries". All these symmetries are strongly congruence; however, as we saw (say, in Remark 30 on p. 47, and in Footnote 202 on p. 78), for $c > 4$ the strong congruence group is much larger than this group of "trivial" symmetries.

In other words: it is easy to show that any combination of the "trivial symmetries" $T \overset{\varphi}{\mapsto} -{}^{1}\!/_{cT}$ and $T \mapsto T + 1$ can be written either as Ψ or as $\varphi \circ \Psi$ with Ψ in a subgroup generated by $T \mapsto T + 1$ and ${}^{1}\!/_{T} \mapsto {}^{1}\!/_{T} + c$. Since the latter one may be written as $T \mapsto {}^{T}\!/_{(cT+1)}$, it is clear that all these transforms Ψ are strongly-congruence transforms.[501]

Here we investigate *how many symmetries* should be added to the "trivial symmetries" discussed in the beginning of this section to get the whole collection of strongly-congruence symmetries.

Start with the basics of adelic approach: we say that two numbers n and n' are *adelically close* to each other if $n - n'$ is divisible by many integers. (For example, we may say that they are N-close if $N! | (n - n')$.) Likewise, two matrices with integer coefficients are close to each other if their matrix coefficients (at corresponding positions) are close to each other.

The same language is applicable to fractional-linear transforms: we say that two such transforms are close to each other if they may be both written as $T \mapsto {}^{(AT+B)}\!/_{(CT+D)}$ with integer A, B, C, D so that the corresponding coefficients are adelically close to each other.[502] Now we can formulate which "extra condition" is needed:

$$\boxed{\text{If a fractional-linear transform is sufficiently close to identity, it preserves } F(t)(\mathrm{d}t)^{½}.}$$

---

[500] Above, we saw how this claim may be generalized: in addition to strongly-congruence transforms, *all* congruence transforms act likewise, with a possible "extra" factor ± 1 equal to $\left(\frac{A}{c}\right) = \left(\frac{D}{c}\right)$. So one drop the (equivalent) conditions $c|(A-1)$, $c|(D-1)$.

[501] Another way to say this is that φ *normalizes* the strong congruence group.

[502] Note a certain similarity to being strongly-congruence: in the same sense as "the adelic distance" is related to divisibility by $N!$ for $N \gg 0$, for strongly-congruence transforms we require that A, C, D are near the coefficients $A = 1$, $C = 0$, $D = 1$ of the *identity transform* "in the sense of divisibility by c". However, we do not require anything about B, and we do not care about divisibility by prime numbers which are not divisors of c.

(Here "closeness" is understood adelically.)

First of all, if we know the property stated above, of preservation by strongly-congruence transforms, then the framed statement holds trivially. Indeed, the defining property of the adelic closeness is that given an integer c,

$$\boxed{\text{Any integer sufficiently close to 0 (adelically!) is divible by } c.}$$

In particular, any transform c-close to identity is strongly congruence.

In the other direction, given a strongly-congruence transform Φ, it is enough to show that

$$\boxed{\text{One can find } \Psi \text{ as above which is arbitrarily adelically close to } \Phi.}$$

(Recall that Ψ is in the subgroup generated by $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\c&1\end{smallmatrix}\right)$.)

First, note that if $c \leq 4$, then one can achieve $\Phi = \Psi$. We will use this fact only for $c = 1$, when it[503] follows from the Gauss elimination process: essentially, (together with the Euclid process) this process shows that multiplying by matrices $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)$ (and their inverse matrices) on the right and on the left, one can reduce any matrix with integer coefficients and determinant 1 to either identity, or to the "exchange-rows/columns" matrix $\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$. On the other hand, the latter matrix is also a similar product:

$$\begin{pmatrix}0&1\\-1&0\end{pmatrix} = \begin{pmatrix}1&0\\-1&1\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}\begin{pmatrix}1&0\\-1&1\end{pmatrix}.$$

Reducing this statement mod M, we can see that if α, β are mutually prime with M, then matrices $\left(\begin{smallmatrix}1&\alpha\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\\beta&1\end{smallmatrix}\right)$ generate $\mathsf{SL}_2(\mathbb{Z}/M\mathbb{Z})$. (Indeed, suitable powers of these matrices coincide with $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)$ modulo M.)

Using Gauss elimination mod c^k, it is easy to see that $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\c&1\end{smallmatrix}\right)$ generate a subgroup of $\mathsf{SL}_2(\mathbb{Z}/c^k\mathbb{Z})$ consisting of matrices $\left(\begin{smallmatrix}A&B\\C&D\end{smallmatrix}\right)$ with c dividing $A-1$, C and $D-1$. Indeed, since $A \mod c^k$ is invertible, we can kill B by Gauss column-transforms, then kill C by row-transforms; this leaves a diagonal matrix with determinant 1. To finish, use the "Whitehead lemma" identity:

$$\begin{pmatrix}u&0\\0&v\end{pmatrix} = \begin{pmatrix}1&1\\0&1\end{pmatrix}^v \begin{pmatrix}1&0\\1&1\end{pmatrix}^{1-u} \begin{pmatrix}1&1\\0&1\end{pmatrix}^{-1} \begin{pmatrix}1&0\\1&1\end{pmatrix}^{1-v} \qquad \text{if } uv = 1.$$

If $1 - u = cU$ and $1 - v = cV$, then $\left(\begin{smallmatrix}1&0\\1&1\end{smallmatrix}\right)^{1-u} = \left(\begin{smallmatrix}1&0\\c&1\end{smallmatrix}\right)^U$, likewise for V, so the expression on the right-hand side is a combination of the required form.

Combining two last results and with the Chinese remainder theorem, it follows that in the last framed statement we can approximate Φ by Ψ modulo c^kM with $(c, M) = 1$, hence modulo any number. This finishes the proof of the double-framed statement.

Conclusion: Strongly-congruence matrices are the adelic closure of matrices generated by $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\c&1\end{smallmatrix}\right)$. "The strongly-adelic part" of the Langlands program would follow from the adelic continuity and the functional equation.

Essentially, this approach allows to break the investigation of fractal symmetries of $F(t)$ into two completely independent parts. First, we may ask whether F satisfies the double-framed condition above.[504] Second, we may ask whether F is horizon-similar at $t = 0$. Assuming that the transformation

---

[503] It (or very similar statements) has many names: the Smith normal form, or the fundamental theorem of finite abelian groups, or the "basic calculation of $K_1\mathbb{Z}$ in algebraic K-theory.

[504] Such functions (or tensor fields) F may be called *adelically analytic*. Here "analyticity" is understood in slightly different way than in real analysis. The adelic approach defines a topology on the set of fraction-linear transforms. However, this topology happens to be *totally disconnected* (as one on the Cantor sets). In these settings, being "locally constant" (as in the double-framed condition) turns out to be the best subtitute for analyticity (at least for complex-valued functions).

However, the most often used name for such F is *automorphic*. (While technically it is more correct to call such an F an automorphic *form*, we are going to abuse notations, and call F an automorphic *function*.)

$T \mapsto {}^{1}\!/_{T}$ in $F(T)$ leads to a function with an integer period c, the arguments above show that if F is adelic, the strongly-congruence transforms (for *that* value of c) preserve F.

As these notes show, for cubic polynomials this approach may look as an obfuscation only — in the preceding chapters we treated $F(t)$ without using any trace of the adelic language. However, for more general cases it turns out that it is much easier to work separately with the "automorphic" properties of F, and with "the 'trivial' fractal symmetry at $t = 0$".

Remark 106: Note that any automorphic function is k-periodic for an appropriate value of $k \in \mathbb{Z}$. Moreover, for any rational numbers A, B, C, D with $AD \neq CD$ the coordinate change $T \mapsto {}^{(AT+B)}\!/_{(CT+D)}$ sends an automorphic function to an automorphic function. Applying this to $-{}^{1}\!/_{T}$, it shows that any automorphic function is horizon-similar near $T = 0$.

The same argument shows that any automorphic function is fractally-symmetrical: the horizon-similar points are dense.

In the following section, we consider the analytic details of having "the 'trivial' fractal symmetry".

The behavior near $t = 0$ and the relation to θ-factors

To begin, note that the most visual feature of the graphs is the horizon-similarity at $t = 0$. Indeed, the horizon-similarity at other rational multiples of π requires more zooming to be seen, so may be considered as "visual features of the second order".

We saw that this horizon-similarity is the geometric counterpart of Hecke's functional equation for the ζ-function (see p. 82). To have this counterpart, we need to choose a suitable "geometry corresponding to the functional equation", and this choice is governed by Hecke's θ-factors. Recall that the θ-factors depend on two parameters: the degree of the polynomial equation we solve, and the number of the real roots of the polynomials. For degree 3, two possible forms of θ-factors lead to two simple symmetries (one with absolute value, one without) of tensor fields on $\mathbb{RP}^1$ (in other words, they lead to 1-dimentional projective geometry), producing the required horizon-similarity near $t = 0$.

However, we saw that for degree larger than 3, the corresponding θ-factors do not seem to allow a similar translation of Hecke's functional equation to symmetries of $F(t)$ (or similar functions of one variable) near $t = 0$. Moreover, while the scoop of the Langlands program is the existence of "a certain fractally-symmetric function $F^{\mathrm{geom}}(t_1, \ldots)$ corresponding to our polynomial equation", to *actually provide* $F^{\mathrm{geom}}(t_1, \ldots)$ seems, in general, much less straightforward than for cubic equations.

On one hand, $F^{\mathrm{geom}}(t_1, \ldots)$ should be a function of several variables. On the other hand, in general, *the relation* of this function to our "counting sequence" N_n is quite indirect: instead of Fourier transform it seems unavoidable to consider "the Hecke eigenvalue problem" of the preceding section. (In other words: for degree 3 we were very lucky that the eigenvalue problem could be *explicitly* solved through *a shortcut* of Fourier transform.)

On γ-factors and ϑ-terms

For these notes, we used the simplest possible examples in which "both sides" of Langlands correspondence allow "an elementary exposition". In fact, it may be that these are the only such cases, and any further progress into understanding the Langlands program may *require* studying *much* more esoteric topics.

Why the cases we consider here are so special? The corresponding Langlands symmetries were *directly applicable* to the Fourier transform $F(t)$ of the sequence N_n. Recall that this sequence was, more or less, a slightly "massaged" point-counting function $\widetilde{N}_n^{\mathrm{res}}$ (or better, $\widetilde{N}_n^{\mathrm{Gal}}$ which needs very little massaging; see p. 60). However, the reason *why* this Fourier transform was so special turns out to be very delicate.

Since I do not qualify to discuss gory details of Langlands program, let us focus on something much simpler: the symmetry which we considered before as "almost trivial", one due to a precursor of Langlands program: Hecke's functional equation (see p. 82). This symmetry sends $G(T)$ to $G(-1/cT)/T$. We saw that when $t = 2\pi T$, these symmetries multiply $F_\mathbb{C}$ by a constant. **Question:** how come this transformation is a symmetry of $F_\mathbb{C}$?

What the functional equation claims about the counting functions $\widetilde{N}_n^{\mathrm{Gal}}$ of a polynomial equation in 1 variable is[505]

- There is a function $\varkappa = \varkappa_{d,r_1}(\sigma)$ defined for $\sigma > 0$ and depending only on the degree d and the number of real roots r_1 of the equation;
- There are numbers $c \in \mathbb{N}$, α and C such that[506] the sum $K(\sigma) := \sum_n \widetilde{N}_n^{\mathrm{Gal}} \varkappa(n\sigma)$ is symmetric: $K(-1/c\sigma) = C/\sigma^\alpha \cdot K(\sigma)$. (In our cases, $\alpha = 1$.)

With distillation (see p. 75), we replace $\widetilde{N}_n^{\mathrm{Gal}}$ by the sequence N_n. For the new sequence a very similar statement continues to hold, only with a different[507] function $\varkappa(\sigma)$.

Additionally, the way the function $\varkappa$ is constructed implies that $\Pi(\sigma\partial_\sigma)\varkappa(\sigma) = B \cdot \sigma^2 \varkappa(\sigma)$ for a certain polynomial Π of degree d and[508] a constant B. Moreover, for $d = 2$, $r_1 = 0$ this may be simplified[509] to $\partial_\sigma \varkappa = -2\pi\varkappa$. (Likewise, there is a similar equation on $\varkappa(\sigma)\sigma^{\alpha_0}$ which differs by a shift in the polynomial Π.)

Note that these ordinary differential equations (ODEs) are preserved by coordinate changes $\sigma \mapsto c \cdot \sigma$ up to a change in the constant B.

This ends the "functional equation" part of our story. On the other hand, note that even "the trivial symmetries" we were considering, say, in the section on p. 43, were obtained by *combining* the symmetry $t \mapsto -1/Ct$ with periodicity in t. The symmetry $\sigma \mapsto 1/c\sigma$ above is very similar to $t \mapsto -1/Ct$, but there is no trace of periodicity in the examples above. **Conclusion:** even the "trivial" symmetries are not *only* due to the functional equation, but require extra arguments.

The main ingredient of this "additional argument" is that instead of $\varkappa(\sigma)$, one can consider the function $\varkappa(\sigma)\sigma^{\alpha_0}\mathrm{e}^{2\pi \mathrm{i} T}$ of σ, T with a certain constant α_0 (in most examples, $\alpha_0 = 0$). It turns out

[505] The Langlands program predicts that a similar statement works for any system of polynomial equations — at least after a suitable distillation (which may be much less trivial than what we did above).

[506] Note that defining $\varkappa_{d,r_1,c}(\sigma) := \varkappa_{d,r_1}(\sigma/\sqrt{c})$ and likewise $K_c(\sigma) := K(\sigma/\sqrt{c})\sigma^{\alpha/2}$ would lead to a simpler equation: $K_c(1/\sigma) = \pm K_c(\sigma)$. However, it is easier to keep c outside of the function K (we already saw this in the rest of these notes, and would see it down below too).

[507] For example, if distillation corresponds to "removing a 'trace' of an equation" of degree d' with r_1' real roots, then one replaces $\varkappa_{d,r_1}$ by $\varkappa_{d-d',r_1-r_1'}$. (Sometimes this may result in negative $r_1' - r_1$ — however, the formulas for $\varkappa_{\bullet,\bullet}$ allow this.)

In the examples we considered in these notes, $d' = r_1' = 1$. (Remark 78 on p. 130 may be considered as an exception — but we did not investigate the *completely distilled* cases of this remark in any detail.)

[508] We could merge B into Π, but it is more convenient to keep it separate.

[509] The function $\varkappa(\sigma)$ is defined by the condition that the Fourier transform of $\varkappa(e^x)$ is $\gamma(\mathrm{i}\xi)$ (hidden inside this description is the notion of the *Mellin transform*). Here $\gamma(z) := \Gamma_\mathbb{R}^{r_1}(z)\Gamma_\mathbb{C}^{r_2}(z)$ is *the gamma-factor*, r_2 is defined by $d = r_1 + 2r_2$, and $\Gamma_\mathbb{R}(z) = \pi^{-z/2}\Gamma(z/2)$, while $\Gamma_\mathbb{C}(z) := \Gamma_\mathbb{R}(z)\Gamma_\mathbb{R}(z+1) = 2(2\pi)^{-s}\Gamma(z)$. (See also the article on `LMFDB.org`.)

One can immediately see that $\Gamma_\mathbb{R}$ satisfies the equation[510] $z\Gamma_\mathbb{R}(z) = 2\pi\Gamma_\mathbb{R}(z+2)$, hence $z^{r_1+r_2}(z+1)^{r_2}\gamma(z) = (2\pi)^d\gamma(z+2)$, and that for $r_1 = 0$ one can simplify this to $z^{r_2}\gamma(z) = (2\pi)^{r_2}\gamma(z+1)$. Since $z = \mathrm{i}\xi$ corresponds to $-\partial_x = -\sigma\partial_\sigma$, this leads to the relations $(-\sigma\partial_\sigma)^{r_2}\varkappa(\sigma) = (2\pi)^{r_2}\sigma\varkappa(\sigma)$ if $r_1 = 0$, otherwise to $(-\sigma\partial_\sigma)^{r_1+r_2}(-\sigma\partial_\sigma + 1)^{r_2}\varkappa(\sigma) = (2\pi)^d\sigma^2\varkappa(\sigma)$.

[510] In the standard expositions, the principal "reason for existence" of γ-factors is "to make the functional equation work". However, all the condition above would obviously be satisfied if one changes $\gamma(z)$ to $\gamma(z)\widetilde{\gamma}(z)\widetilde{\gamma}(1-z)$ for an arbitrary entire function $\widetilde{\gamma}$. On the other hand, the discussed equations mark the "classical" choices of γ-factors as "those satisfying the simplest possible equations".

(For another reason why the "classical" choices are the best possible, see Keith Conrad's answer to the question *Why does the Gamma-function complete the Riemann Zeta function?* on **MathOverflow** — as well as the resulting discussion.)

that in the simplest examples we considered above (and in a few additional examples!), this function satisfies remarkable conditions. Below, we use a "weight function" $W(\sigma) := \sigma^\alpha$ with[511] $\alpha = 1$.

- In the ODE for $\varkappa(\sigma)\sigma^{\alpha_0}$, one can substitute certain constants 2π by the operator $-\mathrm{i}\partial_T$ to obtain a partial differential equation (PDE) in coordinates (σ, T) with a large group of symmetries.
- In particular, this equation is preserved by the "n-rescaling" coordinate changes: $\sigma' = n\sigma$, $T' = n^\beta T$, for an appropriate β.
- Moreover, this equation is preserved by a coordinate change $\mathscr{M}$ which sends the subset $T = 0$ to itself, inducing on it the transformation $\sigma' = {}^1\!/\sigma$, and the transformation of dependent variable[512] $\varkappa' = w(\sigma, T)\varkappa$ with $w(\sigma, 0) = W(\sigma)$.
- Composing the last two shows that one can do the same for $\sigma' = {}^1\!/c\sigma$.

One can immediately see that this means:

- If the degree of this PDE is not too large, one can consider $\varkappa(\sigma)\sigma^{\alpha_0}\mathrm{e}^{2\pi\mathrm{i}T}$ (or maybe $\varkappa(\sigma)\sigma^{\alpha_0}\cos 2\pi T$) as a "natural extension" of $\varkappa(\sigma)\sigma^{\alpha_0}$. This extension is similar to finding a solution of PDE given "a boundary condition on $T = 0$". (Compare to the Cauchy–Kovalevskaya theorem.)
- Likewise, $k(T, \sigma) := \sum_n N_n \varkappa(n\sigma)\sigma^{\alpha_0}\mathrm{e}^{2\pi\mathrm{i}n^\beta T}$ (or a similar sum with $\cos\varphi$ instead of $\exp \mathrm{i}\varphi$) is "a natural extension" of "the boundary value" $K(\sigma)\sigma^{\alpha_0} = \sum_n N_n \varkappa(n\sigma)\sigma^{\alpha_0}$.
- Since "the boundary value" $K(\sigma)\sigma^{\alpha_0}$ "is preserved" by $\sigma \mapsto {}^1\!/c\sigma$ (the quotes mean "up to multiplication by $C \cdot W(\sigma)\sigma^{-2\alpha_0}$"), and this transformation extends to (σ, T) *while preserving the PDE*, the "extension" $k(T, \sigma)$ of this function is also "preserved" by this extension of $\sigma \mapsto {}^1\!/c\sigma$.

(In the examples we saw, related to cubic equations, we had $\beta = 1$.)

Now if the transformation $\mathscr{M}$ preserves the line $\sigma = 0$, and $\beta \in \mathbb{N}$, then[513]

$$\boxed{\text{The restriction } \widetilde{F}(T) \text{ of } k(T,\sigma) \text{ to } \sigma = 0 \text{ is periodic, and "is preserved" by } \mathscr{M}|_{\sigma=0}.}$$

If $\varkappa(\sigma)$ has non-0 limit for $\sigma \to +0$, then $\widetilde{F}(T)$ is $\sum_n N_n \mathrm{e}^{2\pi\mathrm{i}n^\beta T}$. If the same holds for $\sigma^{-\alpha'}\varkappa(\sigma)$, then instead of restriction one should take the main term of asymptotic in σ, and $\widetilde{F}(T)$ becomes $\sum_n n^{\alpha'} N_n \mathrm{e}^{2\pi\mathrm{i}n^\beta T}$. In applications below, $\mathscr{M}$ induces the transformation $T \mapsto -{}^1\!/c^\beta T$ on $\{\sigma = 0\}$, and $w(0, T) = T^{\alpha/\beta}$. (Sometimes it is convenient to restate this as preservation of $\widetilde{F}(T)(\mathrm{d}T)^{\alpha/2\beta}(\mathrm{d}\sigma)^{\alpha'}$. Indeed, $\mathscr{M}$ sends $\mathrm{d}T \mapsto {}^1\!/c^\beta T^2 \mathrm{d}T$. In examples below, $(\mathrm{d}\sigma)^{\alpha'}$ behaves under $\mathscr{M}$ the same as $(\mathrm{d}T)^{\alpha'/\beta}$.)

This leads to the following reformulation of the functional equation:

$$\widetilde{F}(T) = \text{const} \cdot T^A \widetilde{F}(-{}^1\!/cT) \qquad \text{with } A = {}^\alpha\!/\beta + {}^{2\alpha'}\!/\beta.$$

Essentially, we introduced $\varkappa(\sigma)$ as an inverse Mellin transform of the gamma-factor $\gamma(z)$ (see Footnote 509); since in simplest examples this transform would eventually lead to the ϑ-functions (sometimes called by a misnomer *theta constants*), we call $\varkappa(\sigma)$ *the ϑ-term*. **Conclusion:** "the purpose of $\varkappa(\sigma)$ in life" is[514] to combine its rescales $\varkappa(n\sigma)$ as $\sum_n N_n \varkappa(n\sigma)$.

---

[511] For general systems of polynomial equations, the value of α depends on the number $\mathbf{d}$ of parameters for the solutions: $\alpha = 1 + \mathbf{d}$. In these notes we focus on the case of 1 equation with 1 unknown, so $\mathbf{d} = 0$, and $\alpha = 1$.

[512] In other words, if $k(\sigma, T)$ is a solution, then $w(\sigma, T)k(\sigma', T')$ is also a solution.

[513] This tacitly assumes that $k(T, \sigma)$ extends (as a solution to our PDE) from $T = 0$, $\sigma > 0$ up to $\sigma = 0$ and any T. In our examples related to cubic equations, this was not completely true: it could be extended to $\sigma > 0$ honestly, but did not have a continuous extension to $\sigma \geq 0$. However, the extension made perfect sense as a generalized function on $\sigma = 0$. (In the context of PDEs in question this is equivalent to the growth near the boundary $\sigma = 0$ being not quicker than polynomial in σ^{-1}.)

[514] **N.B. (???) Eventually, we would need to take into account that N_n is not a Fourier coefficient, but the eigenvalue of the Hecke operators!**

Examples of ϑ-terms

In many examples we are going to have $\alpha_0 = 0$. This is the assumed value unless we state it otherwise.

Start with an example which should be very familiar from what we investigated in the main part of these notes. By Footnote 509 on p. 163, for $r_1 = 0$, $r_2 = 1$ the ODE on $\varkappa$ is $\partial_\sigma \varkappa(\sigma) = -2\pi\varkappa(\sigma)$ with the solution proportional to $\mathrm{e}^{-2\pi\mathrm{i}\sigma}$ (so $\alpha' = 0$).

The corresponding PDE is $\partial_\sigma - \mathrm{i}\partial_T$. This is the condition of being holomorphic in $T + \mathrm{i}\sigma$, so $k(T,\sigma)$ is just the holomorphic extension of $K(\sigma)$. Since the holomorphic extension is usually considered as something "very natural", even without the scheme introduced above, one "could have just observed" that $K(\sigma)$ has holomorphic extension from imaginary values of ζ to $\operatorname{Im}\zeta > 0$ (here $T = \operatorname{Re}\zeta$, $\sigma = \operatorname{Im}\zeta$).[515]

Conclusion: $\varkappa(\sigma)\mathrm{e}^{2\pi\mathrm{i}T}$ is proportional to $\mathrm{e}^{2\pi\mathrm{i}\zeta}$. One can immediately see that n-rescalings are $\varkappa(n\sigma)\mathrm{e}^{2\pi\mathrm{i}nT}$, so $A = \beta = 1$. Hence putting $s = 2\pi\sigma$, $t = 2\pi T$ and $F(t) := \widetilde{F}(T)$, $f(s,t) := k(T,\sigma)$ coincides with our description of $f(s,t)$ as of "regularization" of $F(t)$ in the section on p. 84. (This means that the restriction to $\sigma = 0$ is $\sum_n N_n \mathrm{e}^{\mathrm{i}nt}$.)

As we explained in the preceding section, this form of $\varkappa$ works for a distilled cubic equation with exactly 1 real root. It is also suitable for non-distilled quadratic equation without real roots.[516]

As we repeated it many times already, in this case not only do we get periodicity of $F(t)$ and the "fractal transform" at $t = 0$, but also fractal transforms in a dense collection of points of the form $t = 2\pi{}^{R}/_{D}$.

The following example is almost as simple as the preceding one.

For $r_1 = 1$, $r_2 = 0$ the ODE on $\varkappa$ is $\partial_\sigma \varkappa(\sigma) = -2\pi\sigma\varkappa(\sigma)$. The solution $\varkappa(\sigma)$ is proportional to $\mathrm{e}^{-\pi\mathrm{i}\sigma^2}$ (hence $\alpha' = 0$). So introduce $S = {}^{\sigma^2}/_2$; then the differential operator becomes $\partial_S + 2\pi$, and we extend it to a PDE $\partial_S - \mathrm{i}\partial_T$. This is again condition of being holomorphic (but in a different coordinate $T + \mathrm{i}S$ than above — it is a different conformal structure!).

The n-rescalings take the form $\varkappa(n\sigma)\mathrm{e}^{2\pi\mathrm{i}n^2T}$. In particular, here $\beta = 2$. Hence the restriction of $k(T,\sigma)$ to $\sigma = 0$ is $\sum_n N_n \mathrm{e}^{\mathrm{i}n^2t}$, and $A = ½$. (Here for similarity with the case above we write $T =: 2\pi t$.)

There are two main applications of this. First, one can consider the simplest polynomial equation: $x = 0$; its solutions correspond to $\widetilde{N}_n^{\mathrm{Gal}} \equiv 1$. Since this is already "the simplest of equations", there is no distillation needed, so $N_n \equiv 1$, and $F(t)$ is the ϑ-function $\vartheta(t) := \sum_n \mathrm{e}^{\mathrm{i}n^2t}$. The arguments above show that it has nice transformation properties w.r.t. $t \mapsto -{}^{1}/_{4\pi^2 t}$. In fact, it is also a modular form (in other words, it allows such a transform not only "near $t = 0$", but also near other points $\pi{}^{R}/_{S}$).[517]

In fact, Riemann in his famous proof ("the second proof") of the functional equation for the Riemann's ζ-function[518] applied exactly the same method, but backwards: he used the known behaviour of ϑ-function under transformation $s \mapsto {}^{1}/_{s}$ to deduce the functional equation.

As another application of this case ($r_1 = 1$, $r_2 = 0$), we can consider a distilled quadratic equation with 2 real roots. In the section on p. 62 we already saw that in this case N_n is a periodic function of n, the Legendre symbol (from p. 208): $N_p = \left(\frac{p}{D}\right)$ (here D is the discriminant of the polynomial). Moreover, the Euler formulation of Quadratic Reciprocity (on p. 15) shows that in this case N_n is an

[515] This is the modus operandi with holomorphic modular forms: until Maass, people did not pay attention that one may need a different procedure of continuation from $\zeta \in \mathrm{i}\mathbb{R}_{>0}$. Other examples of this section give a zoo of such alternative continuations.

[516] One may say that a distillation is similar to the difference between L-function of a field vs. the corresponding motivic L-function. Compare with the section on p. 75.

[517] Although it should be considered as a tensor of rank ¼, and not ½ as we had in the preceding example.

[518] Done in 1859, but the equation itself was conjectured by Euler 110 years before this! (See "Two lectures…" by A. Weil of 1974.)

even periodic function of n. In many cases[519] this implies that it is actually a periodic function of n^2. **Conclusion:** if D has no prime divisors $\equiv_4 -1$, then $\sum_n N_n e^{in^2 t}$ is obtained from the ϑ-function by multiplying its Fourier coefficients[520] a_m by a periodic function of m. Recall that in this case $A = 1/4$.

The next case to consider is good for a distilled quadratic equation without real roots: $r_1 = -1$, $r_2 = 1$. Then the ODE on $\varkappa$ is $(\sigma\partial_\sigma - 1)\varkappa(\sigma) = -2\pi\sigma^2\varkappa(\sigma)$; writing $\varkappa(\sigma) = \sigma\widetilde{\varkappa}(\sigma)$ leads to the same equation on $\widetilde{\varkappa}$ as above: $\partial_\sigma\widetilde{\varkappa}(\sigma) = -2\pi\sigma\widetilde{\varkappa}(\sigma)$. (In particular, $\varkappa(\sigma)$ is proportional to $\sigma e^{-\pi i\sigma^2}$, and $\alpha' = 1$.) This means that $\sigma_0 = -1$.

So we again introduce $S = \sigma^2/2$ etc., and $\widetilde{\varkappa}$ becomes a holomorphic function.

The n-rescalings of $\sigma\widetilde{\varkappa}(\sigma)e^{2\pi iT}$ take the form $n\sigma\widetilde{\varkappa}(n\sigma)e^{2\pi in^2T}$ (here $\beta = 2$). Hence the main term of the expanstion of $k(T,\sigma)$ for $\sigma \approx 0$ is $\sigma F(T)$ with $F(t) := \sum_n nN_n e^{in^2t}$. This leads[521] to $A = 3/4$.

Our arguments show that $F(t)$ is preserved by $t \mapsto -1/Ct$ (as a tensor field). For a distilled quadratic equation without real roots, N_n is odd periodic in n. In fact, in this case $F(t)$ is also a modular form: it is horizon-similar near any $t = \pi R/S$.

(The last two cases are also applicable to any abelian polynomials. According to the Class Field Theory, the corresponding sequence $\widetilde{N}_n^{\mathrm{Gal}}$ may be distilled into several components, and each of these components N_n is a multiplicative periodic sequence. By multiplicativity $N_{-1} = \pm 1$, hence N_n is either even, or odd — so one of the cases above is applicable.[522] In fact, multiplicativity implies also that N_n is either 0, or a root of 1 — and the cases above cover the situations when these roots are ± 1 — in other words, when numbers N_n are real.)

Above, we exhausted the cases when the degree of the ODE is 1. In the Maass case the ODE has degree 2; this corresponds to $r_1 = 2$, $r_2 = 0$. The ODE is the modified Bessel equation $(\sigma\partial_\sigma)^2\varkappa(\sigma) = 4\pi^2\sigma^2\varkappa(\sigma)$ ("of Bessel-order 0") in the coordinate $2\pi\sigma$. The solution we need is proportional to $K_0(2\pi\sigma)$. (This explains appearence of Bessel functions in Remark 35 on p. 58; we expand on this below.)

Note that $\widetilde{\varkappa}(\sigma) := \sqrt{\sigma}\varkappa(\sigma)$ satisfies the equation $(\sigma\partial_\sigma - 1/2)^2\widetilde{\varkappa}(\sigma) = 4\pi^2\sigma^2\widetilde{\varkappa}(\sigma)$, or $\sigma^2(\partial_\sigma^2 - 4\pi^2)\widetilde{\varkappa}(\sigma) = -1/4\,\widetilde{\varkappa}(\sigma)$.

The key observation is that this ODE allows an extension to a PDE in (σ, T) such that

- $\sqrt{\sigma}K_0(2\pi\sigma)e^{2\pi iT}$ is a solution, and
- the PDE has a symmetry $(\sigma, T) \mapsto (n\sigma, nT)$ for $n > 0$.
- the PDE has a symmetry inducing $\sigma \mapsto 1/c\sigma$ on $T = 0$.

This PDE is the eigenvalue problem for the Laplace operator in the (Lobachevsky) metric $\sigma^{-2}(d\sigma^2 + dT^2)$ with the eigenvalue $1/4$. Indeed, this metric is preserved by the dilation above, and by the Lobachevsky 180°-rotations (the compositions of the reflection in the line $T = 0$ and the inversions in the circles centered at 0). In other words, in the notations above, we can take $\alpha_0 = 1/2$, and the action of $\mathscr{M}$ involves only the arguments of $k(\sigma, T)$, but not its values (in other words, it acts on $k(\sigma, T)$ as on a scalar-valued function).

For any K, there is a unique extension of $\widetilde{K}(\sigma) := \sqrt{\sigma}\widetilde{K}(\sigma)$ to an even in T solution $k(\sigma, T)$ of the PDE. If $K(\sigma) = \sum_n N_n\varkappa(n\sigma)$, then $\widetilde{K}(\sigma) = \sum_n N_n/\sqrt{n}\sqrt{n\sigma}\varkappa(n\sigma)$, and the arguments above show that the extension must be $k(\sigma, T) := \sum_n N_n/\sqrt{n}\sqrt{n\sigma}\varkappa(n\sigma)\cos 2\pi nT = \sqrt{\sigma}\sum_n N_n\varkappa(n\sigma)\cos 2\pi nT$.

---

[519] This is related to the bottom-multiplicativity of the Legendre symbol (see p. 210).

Essentially, we need to check that $\left(\frac{P}{D}\right) \neq \left(\frac{P'}{D}\right)$ then $P^2 \not\equiv_D P'^2$. This boils down to $a^2 \equiv_D 1$ implying $\left(\frac{a}{D}\right) = 1$. With the assumptions above this is automatically satisfied when D is prime (since $D \equiv_4 1$, $D > 0$); however, if $D = D_+D_-$ with $D_\pm \equiv_4 -1$, $D_\pm > 0$, then taking $a \equiv_{D_+} 1$ and $a \equiv_{D_-} -1$ leads to a counterexample: $\left(\frac{a}{D}\right) = -1$.

[520] Note that a_m is 1 or 0 depending on whether m is a square.

[521] Note that the transformation for $F(t)$ shows that the zone of horizon-self-similarity behaves as $1/\sqrt{c}$ instead of $1/c$ we saw for polynomial of degree 3 (this is due to $\beta = 2$). Compare this with the Hasse–Artin conductor/discriminant formula.

[522] ... with a minor correction. If the sequence N_n is not real, then $\sigma^\alpha k(1/c\sigma)$ is not $k(\sigma)$ *for the sequence* N_n, but $k(\sigma)$ for the complexly conjugated sequence.

The logarithmic asymptotic of $K_0(s)$ when $s \to 0$ together with the slower growth of the "other" solution suggest that the main term of asymptotic of ¼-eigenvalue $k(\sigma, T)$ of Laplace on the Lobachevsky plane when $\sigma \to 0$ is $F(T)\sqrt{\sigma}\log\sigma$ — and it is easy to show this by a direct calculation (here we take asymptotic in σ of in the sense of generalized functions of T). The action of the transformation $\mathscr{M}$ described above on a function with such an asymptotic leads to a function with the main term $F(-1/cT)\sqrt{\sigma/cT^2}\log\sigma/cT^2$, which is, up to negligible terms, $\sqrt{c}|T|F(-1/cT)\sqrt{\sigma}\log\sigma$. **Conclusion:** since we know that $\widetilde{K}(\sigma)$ is preserved by $\mathscr{M}$, the function $|T|F(-1/cT)$ is proportional to $F(t)$.

So this explains appearence of functions $\sqrt{s}K_0(s)$ in Remark 35 on p. 58, as well as why the "actual fractal transform" in this case involves $|T|$, and the invariance of F under this transform. Moreover, as we saw, and as the Langlands program predicts, a similar fractality happens near every point $\pi P/D$.

There is one more case when one gets the ODE of the second order: $r_1 = 0$, $r_2 = 2$. The ODE is $(\sigma\partial_\sigma)^2\varkappa(\sigma) = 4\pi^2\sigma\varkappa(\sigma)$. The coordinate change $\sigma =: 4\Sigma^2$ with $\omega(\Sigma) := \varkappa(4\Sigma^2)$ makes it $(\Sigma\partial_\Sigma)^2\omega(\Sigma) = 4\pi^2\Sigma^2\omega(\Sigma)$, which is the same equation as above (with Σ replacing every σ used in the preceding example). The symmetries described above become $(\Sigma, T) \mapsto (n\Sigma, nT)$, or $(\sigma, T) \mapsto (n^2\sigma, nT)$, so $\beta = ½$.

Proceeding as above, we conclude that $F(T) := \sum_n N_n \cos 2\pi n^\beta T = \sum_n N_n \cos 2\pi\sqrt{n}T$ satisfies the same relation as above. However, since $\beta \notin \mathbb{Z}$, the function F is not periodic, so this argument does not lead to combinable symmetries of the kind we saw before. In particular, there is a fractal transform sending $0 \mapsto \infty$, but not only we cannot expect the fractal behaviour at any point $\pi P/D$, but not even at the points of the "trivial" Cantor hyper-family.[523]

This case is applicable to distilled polynomials of degree 5 with one real root, and non-distilled case of degree 4 without real roots.

The Hecke operators

As we saw in the preceding section, what in degree 3 was the "trivial" fractal symmetry due to the functional equation is always preserved as a symmetry of the function $k(\sigma)$, but one cannot translate it to a symmetry of a periodic function $F(t)$ if the degree is above 3. How the Langlands theory avoids this problem?

In degree 3, we described a simple procedure to translate the (distilled) sequence N_n into function $F(t)$: take the Fourier series with coefficients N_n. Likewise, given such a fractal function $F(t)$, one can find N_n as Fourier coefficients of $F(t)$. However, there is another, *very circumstantial* connection between F and N_n. In higher degrees, only this circumstantial connection survives. In particular, I never heard of the direct recipe to find F given the sequence N_n: as far as I know, one can put certain conditions on F which define it uniquely, but there is no immediate way to find this unique solution.

This indirect connection goes through the recurrence relations we (essentially) used to define N_n. The Steps (c), (d) on p. 60 show that for a prime p, the sequence $\nu_p(k) := N_{p^k}$ satisfies a simple recurrence relation $\nu_p(k) = A_1\nu_p(k-1) + \ldots + A_K\nu_p(k-K)$ for a certain value of K (for cubic polynomials, $K \le 2$; if we allow trailing 0␣s in the sequence (A_k), we may put $K = 2$). (Here the coefficients A_k depend on p, so we may write it as $A_{p,k}$.)

---

[523] Moreover, not only F is non-periodic (so to expect the behavior of $F(T)$ we need graphs on long intervals of T, but also the series converges much slower. For example, "the total weight" of coefficients N_n contributing to terms $\cos\tau T$ with $\tau \in [\tau_0, \tau_1]$ with bounded $|\tau_1 - \tau_0|$ was growing very slowly for cubic polynomials, but it grows quickier than $\sqrt{\tau_0}$ in this case.

Adding to this the absense of suitable polynomials of degree 5 with discriminant below 1,500 and the fact that it is the discriminant *squared* which leads to the zone of the horizon-self-similarity (due to $\beta = ¼$) it may happen that plotting such $F^{(-1)}(T)$ with any reasonable precision may be not possible with only "naive" algorithms.

Combining with multiplicativity of N_n (Step (e) on p. 60), this may be rewritten in a more general form:

$$N_{np^K} = A_1 N_{np^{K-1}} + A_2 N_{np^{K-2}} + \ldots + A_K N_{np^0}.$$

One can also write similar relations for a rational number $n = \widetilde{n}/p^L$ with $L \leq K-2$, if we assume that $N_n = 0$ unless $n \in \mathbb{N}$.

Obviously, the conditions $N_k = 0$ for $k \leq 0$, and $N_1 = 1$, and numbers $A_{p,k}$ for primes p (with $A_{p,1} = N_p$) uniquely determine the sequence N_n, and its Fourier transform $F(t)$. Moreover, it is easy to rewrite the condition above in terms of the function $F(T)$.

Indeed, given an integer s and a $2\pi s$-periodic function $f(t)$, consider the function

$$\operatorname{Av} f(t) := \operatorname{Av}_s f(t) := \bigl(f(t) + f(t+2\pi) + \ldots + f(t+2(s-1)\pi)\bigr)/s;$$

it is now 2π-periodic. Consider Fourier coefficients f_n of f with $n \in {}^1\!/\!_s\mathbb{Z}$; then Fourier coefficients $(\operatorname{Av} f)_n$ of $\operatorname{Av} f$ satisfy $(\operatorname{Av} f)_n = f_n$ (for $n \in \mathbb{Z}$). While a T-periodic function f is also kT-periodic, $\operatorname{Av} f$ is well-defined since Av_s and Av_{ks} coincide on $2\pi s$-periodic functions.

For example, if f is 2π-periodic, then $f(t/s)$ is $2\pi s$-periodic, and

$$\operatorname{Av} f(t/s) = \bigl(f(t/s) + f((t+2\pi)/s) + \ldots + f((t+2(s-1)\pi)/s)\bigr)/s$$

is 2π-periodic. Call its Fourier coefficients $f_n^{[s]}$; obviously, $f_n^{[s]} = f_{sn}$.

Now one can rewrite the recursion relation on N_n as

$$\operatorname{Av} F(t/p) = A_1 F(t) + A_2 F(pt) + \ldots + A_K F(p^{K-1}t).$$

Indeed, this says that $N_{np} = A_1 N_n + A_2 N_{n/p} + \ldots + A_K F(n/p^{K-1})$ if $n \in \mathbb{N}$ — which coincides with the conditions above. One can immediately recognize this as a claim that F is an A_1-eigenvector of the operator $\mathbf{T}_p := \operatorname{Av} \circ \mu_{1/p} - A_{p,2}\mu_p - \ldots - A_{p,K}\mu_{p^{K-1}}$ (here p is prime).

This operator is called *the Hecke operator*. Note that since in our examples $K \leq 2$, this operator is just $\operatorname{Av} \circ \mu_{1/p} - A_{p,2}\mu_p$. Moreover, inspecting Footnote 184 on p. 74 shows that $A_{p,2}$ is always ± 1 or 0, and 0 appears only for exceptional p; so essentially, in our cases the Hecke operator $\mathbf{T}_p$ takes one of 3 different forms: $\operatorname{Av} \circ \mu_{1/p}$ or $\operatorname{Av} \circ \mu_{1/p} \pm \mu_p$.

Furthermore, comparing these recursion relations with the table on p. 73 suggests that the sign of $A_{p,2}$ (when it equals ± 1) depends only on $p \mod c$; here c is the conductor. In fact, more is true: $A_{p,2} = 0$ iff p divides c, otherwise the sign depends on whether $X^2 = p \mod c$ has solutions; one can write this using the Legendre symbol from p. 210 as $A_{p,2} = -\left(\frac{p}{c}\right)$.

Conclusion: for polynomials of degree 3 the coefficient $A_{p,2}$ in the Hecke operators $\mathbf{T}_p$ for prime p depend only on c and on $p \mod c$; here c is the conductor. Eigenvalues of these operators for the (common!) eigenvector F are integer. Moreover, these eigenvalues match the coloring ???[524]

In fact, it is easy to see that the operators $\mathbf{T}_p$ and $\mathbf{T}_{p'}$ commute for mutually prime p, p' when we apply them to 2π-periodic functions. Since we are going to consider $\mathbf{T}_p$ only for prime p, this leads to a commuting system of operators with a common eigenvector F.

This translation process provides an alternative description of our way to construct the function F:

- We chose a symmetrical geometric object X: the line $T \in \mathbb{R}$ completed by $T = \infty$, with fractional-linear transforms as symmetries.

[524] **N.B. (???) Ref?**

- We consider a certain type of tensor fields on X; the symmetries above act on such tensor fields.[525]
- Given the conductor c, we define the Hecke operators $\mathbf{T}_p$ as above.
- There exists a 2π-periodic[527] tensor field F which is an eigenvector of all the Hecke operators with eigenvalues matching the coloring ???.[528]

However, this description *also* encompasses the fractal symmetries of $F(t)$: before, we constructed F explicitly, then started to check whether it is fractal-symmetric. As we show in the following section, the translation to the eigenvalue-problem gives us a possibility to *start* looking for a suitable $F(t)$ among fractal-symmetric functions.[529]

The Hecke operators and higher degrees in Langlands program

The analisys above and our construction of the sequence N_n from the section on p. 59 completely determine the properties of common eigenvectors of Hecke operators. Given any numbers $A_{p,2}$ (above they are always -1, 0 or 1) and any eigenvalues, there is a function with period 1 which is a common eigenvector (here we ignore questions of convergence of the Fourier series). Moreover, this determines the Fourier coefficients with positive indices uniquely up to a multiplicative constant; likewise for the negative indices.[530]

This leads to the following reformulation of the statement of Langlands program (in the case of cubic polynomials):

- Choose the eigenvalues to match the red/green coloring of the sections ???.
- Then the common eigenvector constructed above is automorphic.[531]

Another way to reformulate this in a very compact form is to note that Hecke operators send an automorphic function to an automorphic function.[532] Therefore one can consider the action of

---

525 For example, $-1/T$ sends $F(T)(dT)^k$ to $F(-1/T)(dT)^k/T^{2k}$. From this, one can recognize the "actual" fractal transform as matching $k = ½$ in the modular case, and fields $F(T)|dT|^{½}$ in the Maass case.[526]

526 In fact, it is easier to formalize the description of this type of tensor slightly differently by using the standard models of action of $\mathsf{SL}_2\mathbb{R}$ on tensor fields. Note that the argument T takes value on "a circle", $T \in \mathbb{RP}^1$; consider a coordinate ϑ on $\mathbb{RP}^1$ such that functions on $\mathbb{RP}^1$ are identified with π-periodic functions of ϑ. (For example, $T = \tan\vartheta$.)

In the coordinate ϑ, one can use fields $G(\vartheta)|d\vartheta|^{½}$ (or ½-densities) for both cases of tensors, but say that in the first case the coefficient $G(\vartheta)$ is π-*anti*-periodic (in other words, G is 2π-periodic, and $G(\vartheta+\pi) = -G(\vartheta)$).

(This is another context in which the Maass form case is a bit easier than the modular form case: it does not need "the twist" of anti-periodicity. Compare with Remark 99 on p. 143.)

527 **N.B. (???) Periodicity: is it Hecke-like too?**

528 **N.B. (???) Ref?**

529 **N.B. (???) Why eigenvalue is fractal-symmetric?**

530 In applications, there are additional conditions on F (preserved by Hecke operators) which determine the negative Fourier coefficients given the positive ones. (For example, $F(t)$ may continue holomorphically to $\{\operatorname{Im} t > 0\}$; or F may be even.) If all eigenvalues equal $1 - A_{p,2}$ (which never happens in examples considered in these notes), then, additionally, $F \equiv \mathrm{const}$ is a solution. This means that under suitable conditions, the common eigenvector is unique.

531 Note that the operator Av is defined only on periodic functions with integer periods. Hence this eigenvector is automatically periodic.

532 Moreover, the same holds for automorphic functions with period of length 1.

Recall that any automorphic function is periodic with a period of integer length. Moreover, if for a given p one of $A_{p,k}$ with $k \geq 2$ or the eigenvalue is non-0, then for the corresponding eigenvalue, p cannot divide the (numerator of) length of the minimal period. In particular, it turns out that if c is square-free (so the previous sentence applies for every p), any eigenfunction has a period of length 1. For ps of the other type, the Fourier coefficients $N_{p^k n}$ vanish for $k \neq 0$ if $(p,n) = 1$, and the length of the period is 1. Allowing functions with length of the period more than 1 does not add "essentially new" eigenfunctions. (This is quite similar to the relation between newforms and oldforms in Atkin–Lehner–Li–Miyake theory.[533])

Because of this, below we consider the Hecke operators in automorphic functions with period 1 only.

533 **??? N.B.: only we do not use orthogonal complement??? Compare with this preprint for Γ_0?**

Hecke operators in the vector space of automorphic functions; again, one can consider the common eigenvector problem *in this subspace*. Call the restricted operators *the automorphic Hecke operators*.

We saw that in the whole space, any collection of eigenvalues matches an eigenvector; however, it is not surprising that very few of collection survive in the subspace:

For a certain common eigenvector of automorphic Hecke operators, the eigenvalues matches the coloring of ??

Call such a sequence of eigenvalues $A_{p,1}$ *automorphic*.

Furthermore, for every $c \in \mathbb{N}$ we can consider the automorphic functions F which have a period of length 1 and such that the change of variables $T \mapsto -{}^1\!/_T$ gives a function with period of length c. (We saw that any automorphic function with period of length 1 is of this form for a suitable c.) They form a finite-dimensional subspace. If $p|c$ for any prime p such that $A_{p,2} = 0$, then the Hecke operators preserve this subspace, so there is only a finite collection of common eigenvectors (up to a multiplicative constant).

Moreover, there is only a finite collection of[534] sequences $A_{p,2} = 0, \pm 1$ (for prime p) depending only on $p \mod c$; for each choice we have a finite number of choices of eigenvectors. This gives a finite collection of automorphic sequences; they are called sequences *on level* c. Hence[535] any automorphic sequence lives on a certain level.[536] [537] **Conclusion:** for every c, there is a finite number of automorphic sequences of leven c, and the sequence of colors for a cubic polynomial is one of them.

In the geometric context of p. 168, we can generalize this approach as:

- We chose a symmetrical geometric object X with a certain collection of symmetries of X.
- Inside the set of symmetries, we consider only a suitable subset of "symmetries with rational coefficients", so that it is possible to say that "two symmetries are adelically close".
- We consider a certain type of tensor fields on X (so the symmetries act on these tensor fields).
- An *automorphic* tensor field is one which is preserved by any "rational" symmetry close to id.
- We define the "X-Hecke operators $\mathbf{T}_p$ acting on automorphic tensor fields on X (as suitable linear combinations of the action of symmetries; the coefficients may depend on the conductor c).
- Look for common eigenvectors of $\mathbf{T}_p$, and call the corresponding sequence of eigenvalues X*-automorphic*.

One of the approaches of the Langlands program is that for any point-counting sequence (such as $\widetilde{N}_n^{\mathrm{Gal}}$)[538] [539] can be "distilled", and any distilled part gives a sequence N_n such that numbers N_p coincide with a certain sequence of X-automorphic eigenvalues.[540] Moreover, the geometric object X matching a given point-counting problem may be chosen from one of the few series of known examples.

In fact, any eigenvector F obtained this way is automatically "fractally symmetric" tensor field on X. This is a part of being X-automorphic!

However, note a major difference with the case when X is 1-dimensional (as the cases considered so far were: F was a function of one variable): above, given any collection of coefficients $A_{p,2}$ of Hecke operators, and any collection of eigenvalues A_p, we could immediately *calculate* Fourier coefficients of the common eigenvector F. However, with more complicated X␣s (apparently) this problem becomes

[534] **N.B. (???) Which?**

[535] **N.B. (???) Is $\chi(p)$ always of this form, for an automorphic sequence?**

[536] Atkin–Lehner–Li–Miyake theory shows that for a given sequence, the eigenvectors form a 1-dimensional space *if we consider the minimal possible level.*

[537] **N.B. (???) Is not uniqueness shown above anyway?!**

[538] One can do more: when one considers polynomials of many variables, one can count not only the points, but also "algebraic cycles".

[539] **N.B. (???) Ref?**

[540] Moreover, (N_{p^k}) for $k \geq 2$ are the coefficients in the definitions of Hecke operators.

much harder. (At least, I do not know cases when F for a particular distilled sequence N_n is described with X of dimension > 1 and "as explicitly" as the recipe of Steps (a)–(e) (from p. 60) — followed by the Fourier transform.)

Conclusion: to every arithmetic "counting problem" (for solutions of a system of polynomial equations, or for "families" of such solutions — a.k.a. "cycles"), one can assign a geometric object X with a collection of "generalized Hecke operators" acting on appropriate tensor fields on X such that the solution to the counting problem is an X-automorphic sequence.

However, this is not all! "The other approach" of Langlands program goes in the opposite direction: it claims that *any* X-automorphic sequence is related to a certain arithmetic "generalized counting problem". It defines *exactly* what a generalized counting problem is — however, I do not know whether *every such problem* may be related to a distillation of a cycle-counting problem for a system of polynomial equations.

Verification and further examples

While we exhausted the examples we may easily plot with our tools, it makes sense to mention what else has a chance to be handled in a similar way. First, we must explain *why* the fractality observed for cubic polynomials *should* take place. The key ingredient is our description of Weil denominators as characteristic polynomials of matrices for Frobenius permutations (see Footnote 322 on p. 118). From another point of view, Frobenius elements permute 3 roots, inspect their 3×3 permutation matrices. A permutation matrix can be thought of as permuting *real weights* assigned to the permuted points (the roots of the polynomial!). However, since these permutations preserve the *total weight*, the eigenvalues of these matrices split into two parts: the eigenvalue 1 "for the total weight", while the rest *matches the action of the permutation* on the distributions of weights with total 0. (The latter vector space is 2-dimensional, hence this action is, essentially, given by 2×2 matrices.)

On the other hand, having the numerator $1 - u$ essentially *cancels* the eigenvalue 1; so all that remains is the second action. Summarizing:

- For every permutation of 3 roots, we inspect how it acts on "real weights assigned to roots" with total 0.
- The characteristic polynomial of this 2×2 matrix has degree 2.
- To every non-exceptional prime number p we assign a particular Frobenious permutation.[541]
- Consider the characteristic polynomial of the 2×2 matrix of the action of Frobenius permutation.
- The coefficients of this characteristic polynomial can be considered as coefficients of the recurrence relation.[542]
- Our numbers[543] $N_{p^k} =: a_k$ are defined by this recurrence relation. (We start with $a_0 = 1$, and a_k=0 for $k < 0$.)

People who have heard of[544] Artin L-function can immediately recognize[545] that our numbers N_m are exactly the coefficents of this function (for our assignment of 2×2 matrices).[546]

[541] Well, only a conjugacy class — but all permutations in a class have the same characteristic polynomial.

[542] For example, a polynomial $1 - 3u + 2u^2$ gives a recursion relation $a_k - 3a_{k-1} + 2a_{k-2} = 0$.

[543] ... from the section on p. 59.

[544] **N.B. (???) This is a duplicate of the section on p. 131.**

[545] In addition to what we did in the section on p. 115, one needs to check that the standard definition of *the Frobenius permutation* gives a 3-cycle if there are no roots mod p (the "red" primes), a transposition in the case of 1 root, and the trivial permutation in the case of 3 roots.

[546] Since our language is not good enough for a *general* description of what happens in *exceptional* primes, this does not verify the match if m is divisible by an exceptional prime. Still, in our particular case one can check such matches as well.

Finally, recall that *in the simplest cases* this part of Langlands program is *already known*:

$\boxed{F(t) \text{ has required fractal properties when } N_m \text{ are coefficients of an "uncomplicated" Artin } L\text{-function.}}$

According to Langlands–Tunnell results (of ≈ 1980)[547] a case is "uncomplicated" if the matrices are 2×2, and it is not the "icosahedral" case: products of these 2×2 matrices do not match the composition laws of the symmetries of an icosahedron.[548]

One can try to proceed as above for polynomials of degree $d > 3$. Basically, there are two strategies: start as above, with weights with total 0, which leads to $(d-1)\times(d-1)$-matrices; or proceed with appropriately chosen 2×2-matrices.[549]

In the first case, one still gets $N_p := \widetilde{N}_p - 1$ for non-special p. However, start with recalling that in the cubic case, cyclic polynomials would lead to 2×2-matrices which can be simultaneously diagonalized — and that this made the fractality patterns more complicated (see Remark 78 on p. 130). So, first of all, one may need to exclude a similar situation: when the $(d-1)\times(d-1)$-matrices above can be simultaneously block-diagonalized.[550][551]

Moreover, in the 2×2 case Gelbart explicitly states[552] how to translate Artin's L-function to a particular "automorphic form" (which is F, in our language), and the properties of this form. While in the case of general $n\times n$ matrices Cogdell *apparently* says that this would work too:[553] "Surprisingly, the technique is exactly the same as Hecke's, i.e., inverting the integral representation", I could not find any exposition which would result in anything "explicit", such as our generalized functions F.

In the second case, where we assign 2×2-matrices, the Langlands program has sufficiently explicit formulations[554] to ensure *the same* fractality properties for $F(t)$ as what we saw in our examples. — However, in this case the numbers N_p need to be given by more complicated formulas[555] than

[547] In Knapp's notes in the Edinburgh Proceedings *Representation Theory and Automorphic Forms*, 1997, this is Theorem 8.9 (together with the paragraph after Theorem 8.7).

[548] The icosahedral case is also known, but only in the "even" case (meaning: 1 real root) since 2009. See Khare–Wintenberger paper *Serre's modularity conjecture. I*.

[549] These two cases should be eventually connected by Langlands functoriality (for an introduction, see Friedberg's AMS Notes).

Note that for functoriality to be *immediately* applicable, in a particular direction, one needs an extra property: if one strategy assigns to two Galois symmetries g and g' the matrices M_g and $M_{g'}$ which have the same eigenvalues with the same multiplicies, then the other strategy must do likewise. (For non-cyclic cubic polynomials this works with our first strategy and an arbitrary strategy in place of the second strategy.)

[550] For 2×2-matrices, block-diagonalization and diagonalization are equivalent.

[551] For example, this happens for abelean polynomials in degree ≥ 3 (as we already saw in Remark 78 on p. 130), and, in degree ≥ 4, for polynomials with the Galois group being the dihedral group D_k.

[552] In Section 4.2 (and Remark 2.5.5) of his chapter in *Modular Forms and Fermat's Last Theorem*. I could not find similarly explicit and general statements elsewhere!

[553] In *Analytic Theory of L-Functions for* GL_n.

On the other hand, Bump writes "The form of the Gamma factors in the functional equation show that a complex Galois representation can be associated with an automorphic form in this way only if the automorphic form is a modular form of weight one or a Maass form of weight zero with a Laplacian eigenvalue of ¼" (in *Automorphic forms and representations*, CUP 1998).

[554] Compare with Footnote 548 on p. 172.

[555] Moreover, arguments of these formulas may include the counts $\widetilde{N}_p^{\text{extra}}$ for some *ancillary* polynomials! (This is somewhat similar to what we did in Remark 40 on p. 72.) Still, the resulting numbers remain remote cousins of our original coloring of numbers into red/green on p. 19: the possible coefficients N_p assigned to red and green prime numbers p are different.

These "ancillary" polynomials may be "symmetric powers" of the initial polynomial P: if P has roots x_k, the second symmetric power would have roots $x_1 + x_2$, $x_1 + x_3$, etc.

$N_p := \widetilde{N}_p - 1$ even for non-special p. Moreover, only a few "flavors" of polynomials allow such inclusions into 2×2-matrices.[556]

For example, in degree 4 the Galois group is a subgroup of S_4, which is a group of rotations of a cube, hence may be included into[557] $\mathsf{SO}_3\mathbb{R} \simeq \mathsf{PSU}_2 \subset \mathsf{PSL}_2\mathbb{C} \subset \mathsf{PGL}_2\mathbb{C}$. The same[558] happens in degree 5 when the discriminant is a complete square: the Galois group is a subgroup of A_5 which is a group of rotations of icosahedron. (Compare with the section on tetrahedral/etc. cases in WikiPedia.)

With the first scenario, we are dealing with a sequence of numbers N_n obtained by the *essentially the same* rules as the rules for cubic polynomials in the beginning of this section. Moreover, the fractal transform at 0, given by $F(^1/_{Ct})/t$, still multiplies $F(t)$ by a constant (due to the "Hecke's functional equation"); therefore the same happens at t in the Cantor hyper-family (see p. 83). However, I could not find what the Langlands program could predict about the fractality near *other* rational multiples of 2π. This leads to a question about coefficients N_n of the Artin's L-function of the $(d-1)\times(d-1)$ matrices defined above:

$$\boxed{\text{Would the Fourier transform of } N_n \text{ be an exact fractal?}}$$

This question is kind of remote from Langlands program: when we assign $(d-1)\times(d-1)$ matrices, this gives a mapping from the Galois group of an irreducible polynomial into $\mathsf{GL}_{d-1}\mathbb{C}$. By Langlands program, this is related to objects whose symmetries contain the "Langlands dual" group, which is also GL_{d-1} — but (in principle!) we need to consider the "adelic flavor" of this group.[559]

Fortunately, the Galois group we started with was finite, so this is the so called "Artin case",[560] and — to make the long story short — we can ignore the "adelic" part and substitute something much simpler. In the Artin case, the "complicated part of adelic GL_{d-1}" should act trivially! After we take this into account, what we need is a geometric object $\widehat{T}$ with the action of the group $\mathsf{GL}_{d-1}\mathbb{R}$ and a tensor field $\widehat{F}(\widehat{t})$, $\widehat{t} \in \widehat{T}$, which is preserved by a certain congruence-subgroup of $\mathsf{GL}_{d-1}\mathbb{Z}$. (The choice of the congruence-subgroup is the place where the conductor enters the picture!)

With $d-1=2$ the geometric object is the t-line (completed by $t=\infty$), and the tensor field is our function $F(t)$ (considered as a tensor-field on this line). The fact that it is preserved by a congruence-subgroup led to fractal symmetries of $F(t)$ and $F^{(-1)}(t)$. However, for $d>3$ the group $\mathsf{GL}_{d-1}\mathbb{R}$ does not act on the projective line! This is why the fractality of $F(t)$ requires a separate consideration. Anyway, a question remains:

$$\boxed{\text{Is there a recipe for } \widehat{F}(\widehat{t}) \text{ in terms of } d \text{ and the sequence } N_n\text{?}}$$

At the very end, the "arithmetic" part of the Langlands program happens to have two facets: reciprocity and functoriality; above, we used reciprocity only. **Question:**

$$\boxed{\text{Is it possible to extract any additional info about functions } F(t) \text{ from Langlands functoriality?}}$$

[556] One needs to consider *inclusions* since the kernel would lead to us, essentially, studying a polynomial of a smaller degree.

[557] Note that this mapping does not lift to a mapping to $\mathsf{GL}_2\mathbb{C}$. While there is *another* mapping into $\mathsf{GL}_2\mathbb{C}$, it passes through $\mathsf{S}_4 \to \mathsf{S}_3$, hence has a kernel. — Therefore the corresponding sequence N_n corresponds to a related polynomial (the "*cubic resolvent*") of degree 3 (from .

[558] ... only in this case there is no non-trivial homomorphism into $\mathsf{GL}_2\mathbb{C}$ whatsoever.

[559] The adeles in question are *rational* adeles, provided our polynomial had rational coefficients (so the Galois group is defined *over rationals*). (*This* is why we eventually arrive at the *real* flavor of GL_{d-1}.)

[560] Essentially, this means that we consider "motives of dimension 0" — indeed, our polynomial is 1 equation with 1 unknown.

The bird's eye view and the Grothendieck group of manifolds

In these notes, we chipped off a tiny chunk from the Langlands program; this chunk shows that the "point counting function" $\widetilde{N}_m$ (see p. 60) for polynomials of degree up to 3 is in no way "random", but has a very strong "pattern" (when restricted to prime m; otherwise, one needs to consider $\widetilde{N}_m^{\mathrm{Gal}}$). Essentially, the Langlands program "at large" goes in the direction of stating "something similar"[561] for general systems of polynomial equations.

Suppose that the last (fuzzy) statement is literally true. What would be the corollaries for arithmetic? Consider the vector space spanned by all possible sequences $\widetilde{N}_p$ (indexed by prime p). Then:[562]

- This vector space has an increasing filtration indexed by a certain "complexity degree".[563]
- So far, we encountered three levels of complexity: if we have no equations and d unknowns, then $\widetilde{N}_p = p^d$, so we have a polynomial sequence. For a quadratic equation with 1 unknown, we get a periodic sequence. (Likewise for other abelian polynomials.) For a non-abelian cubic equation with 1 unknown, we get $\widetilde{N}_p = 1 + N_p$ where N_p are coefficients of a modular or a Maass form (and 1 is a polynomial — so it sits in "a simpler" level of filtration!).[564]
- The existence of our "distillation process" suggests that the filtration above can be refined to a grading.
- The Langlands program suggests that the index set of the grading is related to complex reductive groups.[565]

Moreover, "joining systems of equations together" shows that the vector space above is actually closed under pointwise multiplication of sequences. It is not very hard to check that when we multiply coefficients of a modular form by a periodic function, the result is again a sequence of coefficients of a modular form.[566] This suggests[567]:

$$\boxed{\text{The filtration above is closely related to pointwise multiplication.}}$$

(The first non-trivial examples of such multiplicativity were discovered by Rankin and Selberg about 1940.)

Finally, the vector space above is a very close relative of the K-group (actually, it has a structure of a commutative ring) of algebraic manifolds:[568] given a submanifold $Z \subset X$, we can introduce a relation $[X] = [Z] + [X \smallsetminus Z]$ in the abelian group generated by classes $[X]$ of isomorphism of such manifolds. The K-group is the quotient by these relations. Direct product of manifolds gives a structure of a ring on this group. The vector space above is a quotient of this ring by a certain ideal.[569]

---

[561] Unfortunate, my almost complete illiteracy in these topics does not let me say anything more precise.

[562] I suspect that this approach should be well-known to specialists in the Langlands program — but I never saw it mentioned explicitly.

[563] This measure of complexity is "orthogonal" to dimension, degree, discriminant and height.

[564] In fact, for elliptic curves the answer is quite similar to the last one (only *the weight* of the form is different, and 1 is replaced by $1 + p$).

[565] Moreover, for every group there is an additional filtration *by conductor* (ordered by divisibility). For example, inside the vector space of periodic sequences (here the group is $\mathsf{GL}_1\mathbb{C}$) one considers subspaces of c-periodic sequences.

[566] Although I did not see it written this way! The resulting conductor is typically much harder; it is a divisor of cK^2; here c is conductor of a modular form, and K is the length of the period.

[567] I cannot follow it close enough, but I suspect that Cogdell's paper (see Footnote 553 on p. 172) investigates what happens in this directions.

[568] Warning: this should not be confused with the (completely unrelated) K-group of an *individual* manifold!

[569] Very little is known about the K-group. However, it *is* known that the affine line ("zero equations with one unknown") is a divisor of zero, hence this ideal is non-trivial!

The filtration conjectured above can be lifted to the K-group — but the ideal remains unfiltered. This leads to a question:

Can the lifted filtration be "meaningfully refined" so that the result subdivides the ideal?

Exercises on Fourier transform

The following exercises are not tuned/debugged yet.[570]

Exercises G: Fourier transform as black box — and other approaches

Only a very minimal knowledge of Fourier series is required for the main body of these notes. It seems that for this purpose, on the first reading it is possible to treat Fourier transform as "a black box", with only two particular properties of this "black box" important for our purposes:

- No information is lost by the (direct and inverse) Fourier transforms.

  In other words, they are *bijections* between sequences and 2π-periodic functions — so they *recode* the information contained in the series into a function — and back.
- "The speed of decay" of the sequence *matches* "the degree of smoothness" of the function.

  Here "the speed of decay" of the terms of the sequence measures how quickly they go to 0. One may say that "the degree of smoothness" of the periodic function essentially measures how many derivatives of the function "make sense".

These two properties are easiest to understand[571] provided we consider sequences which "decay 'sufficiently' fast", and functions which are "'sufficiently' smooth". Moreover, the second property may be qualified numerically: consider two operations:

- Multiplication of a sequence (a_n) by n (which decreases the "rate of decay"), and
- Taking a derivative of a function (which decreases the "degree of smoothness").

Fact: these operations "match each other" (up to a factor $\mathrm{i} = \sqrt{-1}$) "under Fourier transform". In other words, the derivative $(\mathscr{F}a)'$ of the Fourier transform of a sequence a_n is the Fourier transform $\mathscr{F}(\mathrm{i}na_n)$ of the sequence $(\sqrt{-1}na_n)$. So:

Each step in "how many derivatives make sense" matches speedup by the factor $1/|n|$ of "the rate of decay".

To finish the "black box" treatment: keep in mind that the Fourier transform was invented more than a century before "the hidden symmetries" we discuss here were conjectured/understood — and it has an enormous usability outside of the topics of these notes. So its role in "unraveling" the hidden symmetries of number theory is not of "a specialized tool invented for this particular purpose", but should be considered as a mathematical miracle.

The exercises below are for the readers who are not satisfied with the "black box approach" and want to understand "all the details". — Use them as a tool to control one's understanding. **Warning:** These excercises may be quite heavy on "the analytic skills": they assume a working dexterity in dealing with integrals and estimates.[572] [573] An exercise may depend on the results of the preceding exercises, but in most cases one should be able to solve them "independently" when one "just takes the conclusions of preceding exercises as given".[574]

---

[570] **N.B. (???) Check!**

[571] We inspect the points of view which make this valid in *the other cases* as well in the group L of exercises on p. 191. These points of view are important for our purposes since the sequences we consider are "one or two steps worse" than what can be covered by the "easiest" approaches.

[572] On the other hand, it is probably much more important to be able to understand "what these exercises are about" than to be able to actually *solve* them. *This* is why we promote them only as "a tool of control".

[573] The hints are not yet field-tested: I'm not sure that they are really helpful.

[574] Or one may use plotting software to "cheat" and *look* what is the result of missed problems. (I expect that the values of the constants in the estimates below are "correct" — but one should be ready to change them a bit if they do not actually "fit". Please let me know about every such case!)

The story of Fourier transform of 2π-periodic functions goes like this:

- Given such a function F, it may be "represented in a certain sense by a series" $\sum_{k\in\mathbb{Z}} F_k e^{ikt}$.
- The numbers F_k may be found integrating $\frac{1}{2\pi}F(t)e^{-ikt}$ over the period.

The devil is in the details: what does it mean to be "represented in a certain sense by a series", and (a bit less intriguing) what "integration over period" actually means.

In the simplest case, the meaning is plain. Repeat that a *trigonometric polynomial* is an expression of the form $c_0+c_1\cos t+c_2\cos 2t+c_3\cos t+\ldots+c_m\cos mt+s_1\sin t+s_2\sin 2t+s_3\sin 3t+\ldots+s_n\sin nt$.

Exercise G0000: For any polynomial $P(x,y)$ of two variables, $P(\cos t,\sin t)$ is a trigonometric polynomial. Any trigonometric polynomial F may be written in the form $F_{-u}e^{-iut}+F_{-u+1}e^{-i(u-1)t}+\ldots+F_{v-1}e^{i(v-1)t}+F_v e^{ivt}$ with $u,v\in\mathbb{Z}$.

Moreover, one can find F_k with $k\in\mathbb{Z}$ as *the k␣th Fourier coefficient of* F defined as $\frac{1}{2\pi}\int_0^{2\pi}F(t)e^{-ikt}dt$. In particular, if $k<-u$ or $q>v$, the latter integral is 0.

Hint: It is enough to consider the case $k=0$.

In order of complexity, the next case is when F is "smooth enough". We start with two observations recalling the usual criteria of convergence:

EXERCISE G000 **(Cauchy's convergence test):** Given a subset S of an interval $[-M,M]$, the *hull of* S is the smallest closed interval[575] containing S. Say that a sequence S_n with $|S_n|$ bounded by M is a *Cauchy sequence* if the length of the hull of the set $\{S_n,S_{n+1},S_{n+2},\ldots\}$ decreases to 0 as n grows. (Denote this hull by $[m_n,M_n]$.)

Show that the sequence m_n is non-decreasing, and the sequence M_n is non-increasing, and there is a number L such that $m_n\le L\le M_n$ for every n. Show that for a Cauchy sequence, such L is unique. Show that if L is unique, it is[576] the limit of S_n.

Hint: Consider the hull of $\{m_1,m_2,\ldots\}$.

EXERCISE G00 **(Bounds for power law sums):** Recall Lagrange's Mean value theorem expressing a difference via derivative: $f(x+1)-f(x)=f'(x+c)$ for a certain $0<c<1$ (provided $f'(x+c)$ makes sense for all such c). Hence $f(x+1)-f(x)>f'(x+1)$ if f' decreases.

Show that $1/(n+1)^r+1/(n+2)^r+\ldots+1/m^r<\frac{1}{(r-1)n^{r-1}}$ for any $m>n\ge 1$ and $r>1$. (Hence given $r>1$, by increasing n one can make this sum as small as necessary for all $m>n$ at once.)

Hint: $1/n^{r-1}-1/(n+1)^{r-1}>(r-1)/(n+1)^r$ for $r>1$.

Exercise G0: If G has a continuous m␣th derivative, then its Fourier coefficients $|G_k|<C/(1+k^m)$ for a certain number C.

Given a sequence (f_k) such that $|f_k|\le C/|k|^r$ for $k\ne 0$, the corresponding Fourier series $F(t):=\sum_{-\infty}^{\infty}f_k e^{ikt}$ converges at every t to a continuous function F provided $r>1$. Likewise, if $r>n+1$, then F has n continuous derivatives. (Moreover, the corresponding Fourier series converges uniformly[577] in t.)

Hint: Show that $|F_{\le K}(t_1)-F_{\le K}(t_2)|\le 2S_K\sin\min(\pi/2,K|t_1-t_2|/2)$ with $S_K:=\sum_{k=-K}^{K}|f_k|$.

Note *a gap of size 2* in the conclusions of this exercise: starting with m, we allow $r\le m$, hence the round-trip gives a knowledge of n continuous derivatives provided $n<m-1$. This is essentially the same as putting $n=m-2$.

Exercise G1: The Fourier coefficients F_k of the continuous function F defined in the preceding exercise coincide with numbers f_k.

Hint: Use uniform convergence.

Anyway, one can use the description above as[578] a way to reconstruct a continuous function F from its Fourier coefficients F_k:

---

[575] Usually, one denotes its ends by $\inf S$ and $\sup S$. To get this interval, take the intersection of *all* the closed intervals containing S. (This intersection is closed and convex, hence an interval.)

[576] It should be obvious that if S_n has a limit l, then L is unique and $L=l$.

[577] This means that by choosing N one can make $|F_{\le N}(t)-F(t)|$ small *simultaneously* for all t.

[578] **N.B. (???) Generalize to integrable? sum-then-differentiate?**

Exercise G2: If F is continuous, then the series $G(t) := -\sum_{k\neq 0} F_k/k^2 \cdot e^{ikt}$ converges at every t (uniformly!), and G has a continuous second derivative. Moreover,[579] $F_0 + G''(t)$ coincides with F.

Hint: Consider the second antiderivative of $F - F_0$.

Exercises H: Fourier transform and "generalized functions"

Here we show how one can translate the results of a very naive approach above into an extremely powerful machinery covering the majority of situations[580] where Fourier series appear in analysis. However, this approach requires a tiny bit of abstraction. People for whom this bit it too large to digest on the first bite may want to switch first to the exercise sections K on p. 187 and L on p. 191 where we treat things in a very down-to-Earth terms.

The exercises of the preceding section show the importance of the following situation: we have a function G whose Fourier series converges to G (pointwise, or uniformly — does not matter). It might be hard to see from the Fourier coefficients G_k of G, but it turns out that G has an mth derivative $G^{(m)}$ ($\boxed{\text{"in a certain sense"}}$ — for example, a continuous mth derivative). Nevertheless, we want to work with $F := G^{(m)}$ "as if" it matches the Fourier series $i^m \sum_{-\infty}^{\infty} k^m G_k e^{ikt}$.

So we call the numbers $i^m k^m G_k$ *the **formal** Fourier coefficients* of F. Note that if F is actually a continuous function, then these numbers coincide with the "usual" Fourier coefficients of F; in other words, one can find numbers F_k as "certain integrals involving F" (as above). (In particular, dropping the word "formal" does not lead to any confusion; — so we do this below.) However, if the framed "sense" above is more delicate, the calculation of F_k in terms of F may require a delicate analysis.[581]

The *most general* approach to the "sense" of the mth derivative (as above, in the framed phrase) is to do it *formally*. So say that a pair (F, m) with F a continuous function and $m \in \mathbb{Z}_{\geq 0}$ represents *an mth formal derivative derivative of* F. Write this pair down as $F^{[m]}$. To fix obvious defects, say that $F^{[m]}$ and $G^{[m]}$ are *indistinguishable* when $F - G$ is a polynomial of degree $\leq m - 1$. In addition, say also that $F^{[m]}$ and $G^{[m-1]}$ are indistinguishable when G is the derivative of (differentiable) function F.

Exercise H1: Suppose that $n \geq m$ and one can join formal derivatives $F^{[n]}$ and $G^{[m]}$ by a chain of indistinguishable formal derivatives. Then F has $n - m$ continuous derivatives, and $F^{(n-m)} - G$ is a polynomial of degree $\leq m - 1$.

Hint: By induction, construct $k \geq 0$ and a continuous function H with $k + n - m$ continuous derivatives s.t. $F = H^{(k)}$ and $H^{(k+n-m)} - G$ is a polynomial of degree $\leq m - 1$.

If one can join $F^{[0]}$ and $G^{[0]}$ by a chain as above, then $F = G$.

Now we may consider *formal repeated derivatives "up to undistinguishability"*. In other words, we *identify* formal derivatives which may be connected by a chain as in the exercise. The short name for a "formal repeated derivative of a continuous function up to undistinguishability" is a *generalized function*.[582] [583]

---

[579] **N.B. (???) Compare with Cesàro summation.**

[580] There is a bit more general theory of "Fourier transform of hyperfunctions". The situations where it is *actually* inavoidable are rare — and we do not mention them in these notes except Footnote 603 on p. 186.

[581] **N.B. (???) Compare with the discussion after Exercise H5 on p. 179 and "multiply-by-C^m???.**

[582] Note that the usual definition of generalized functions uses not the fact that every generalized function is repeated derivative of a continuous function, but the fact that every generalized function has well defined F-averaged values for a smooth weight function F. Compare with the discussion after Exercise L25 on p. 196.

[583] For people who did not see similar approaches before, this definition may seem abstract. — However, this is *exactly* the way one deals with rational numbers!

There, instead of differentiation, one *divides by a positive integer*. So "a rational number is described by its numerator R and denominator S" (written as R/S), but the numbers written as R/S and R_1/S_1 are *indistinguishable* if $R_1 = kR$ and $S_1 = kS$ with $k > 0$.

Without these "generalized numbers", one could divide an integer R by a positive integer S "only sometimes". (Like a continuous function not always having Sth derivative.) Now one can always "proceed with this division" by "considering a formal expression R/S". (So keep this analogy in mind when dealing with operations defined below.)

Define addition of $F^{[n]}$ and $G^{[m]}$ by writing them down (up to undistinguishability!) as $\underline{k}$th formal derivatives (with $k \geq \max(m,n)$) and defining $H^{[k]} + J^{[k]} := (H+J)^{[k]}$. Finally, one can define the derivative as $(F^{[n]})' := F^{[n+1]}$ and parallel translation by T as $(F^{[n]})(t-T) := (F(t-T))^{[n]}$.

It is clear that this addition — and subtraction, and multiplication by a constant — makes sense on "derivatives up to undistinguishability", and satisfy all the usual properties of these operations.[584] Moreover, the second claim of the last exercise shows that one can consider continuous functions as particular cases of generalized functions.

Exercise H2: Taking derivative of continuously differentiable functions is compatible with the inclusion of continuous functions into "generalized functions". Same for addition of continuous functions and multiplying them by a constant.

If F is a formal repeated derivative and F' coincides up to undistinguishability with 0, then there is a number C such that F coincides up to undistinguishability with the constant (continuous) function C.

Conclusion: these compatibility properties make it possible to use the notation $F^{(n)}$ for the $\underline{n}$th formal derivative of a continuous function F up to undistinguishability. Indeed, if this derivative coincides with a continuous[585] function G, all our operations with the formal derivative (including considering the antiderivatives!) coincide with operations on G.

Conclusion: we defined a collection of "objects" called "generalized functions"; they allow the usual operations of addition, multiplication by a constant, differentiation, and antiderivative (defined up to addition of a constant function). Every continuous function (and later we include L^1-functions and measures, see Exercise M3) is identified with a unique generalized function.

Say that a generalized function F is T-periodic if $F(t-T)$ coincides with F. Say that two generalized functions $F^{(n)}$ and $G^{(n)}$ (with continuous functions F and G) coincide on an open set S if

- If S is an interval: if $F-G$ coincides with a polynomial of degree $\leq n-1$ on S.
- In general: if $F^{(n)}$ and $G^{(n)}$ coincide on any subinterval of S.

(This makes sense[586] since both parts give compatible results if S is an interval.)

Exercise H3: Define δ as $\frac{1}{2}|t|''$. Then δ coincides with 0 on $\mathbb{R} \smallsetminus \{0\}$.

Exercise H4: Consider a generalized function F which coincides with 0 on $\mathbb{R} \smallsetminus \{0\}$. Then F is a linear combination of δ-function and its derivatives.

Exercise H5: If a generalized function F is T-periodic, and is an $\underline{M}$th derivative of a continuous function, then there is a number F_0 such that for any $m \geq M$ one can write $F - F_0 = G^{(m)}$ with G a T-periodic continuous function. Moreover, G is defined uniquely up to addition of a constant.

Hint: Write $F = H'$ and consider the generalized function $H(t-T) - H(t)$.

Call such number F_0 *the average value of* F. This exercise shows that the argument after Exercise G2 is applicable to $F - F_0$ — provided one understands $\boxed{\text{a certain sense}}$ as "in the sense of generalized functions". In other words, we *define* Fourier coefficients F_k of F for $k \neq 0$ as $\mathrm{i}^m k^m G_k$; here G_k are the Fourier coefficients of G. The 0th Fourier coefficient of F may be defined as F_0.

However, to "give enbodiment" to this definition, we need

- To describe the relation of F_k to the expressions $\frac{1}{2\pi}\int_{\text{period}} F(t)\mathrm{e}^{-\mathrm{i}kt}\mathrm{d}t$.

Warning: for rational numbers, there is "the best" ("reduced") representation as R/S. Hence to compare rational numbers for equality, one can reduce the representations, then compare numerators and denominators separately. For generalized functions, this is not possible — hence to check equality of $F^{[n]}$ and $G^{[m]}$, one *is forced* to do something like Exercise H1.

[584] Such collections are called "*vector spaces*".

[585] In classical analysis one can also consider situations of "derivatives which make sense as non-continuous functions" of a special form, in Exercise M3 we show how such functions (e.g., L^1-functions, or measures) can be included in our formal framework.

[586] To be pedantic, one should have checked that the definition above is not changed if $F^{(n)} = \widetilde{F}^{(\widetilde{n})}$ and $G^{(n)} = \widetilde{G}^{(\widetilde{n})}$ and one replaces n by $\widetilde{n}$.

- To describe the relation of F to the expression $\sum_{-\infty}^{\infty} F_k e^{ikt}$.

We address the first issue in the rest of this section, and the second issue in the next section I on p. 180.

Exercise H6: Show that for every $n \geq 0$ there are polynomial expressions E_k, $k = 0, \dots, n$, in G and derivatives of Φ (up to $\Phi^{(n)}$) such that $\Phi G^{(n)} = E_0 + E_1' + E_2'' + \dots + E_n^{(n)}$ provided Φ and G are functions such that all the involved derivatives make sense.

Hint: Guess E_n and proceed by induction.

Consider a formal derivative $F = G^{(n)}$ with a continuous G. Then for every n times continuously differentiable Φ, the expressions E_k give continuous functions, so we can *define* ΦF using the RHS of the expression in the exercise.

Exercise H7: Show that different (indistinguishable!) representations $F = G^{(n)} = \widetilde{G}^{(\widetilde{n})}$ of a generalized function F lead to the same generalized function ΦF provided Φ has at least $\max(n, \widetilde{n})$ continuous derivatives. Moreover, for $n = 0$ one gets "the usual multiplication" of continuous functions.

Show that the Leibniz's rule $(\Phi F)' = \Phi' F + \Phi F'$ holds provided Φ has sufficiently many continuous derivatives for all the products to make sense. Show that $\Psi(\Phi F) = (\Psi\Phi)F$ provided Φ and Ψ have sufficiently many continuous derivatives. Show that $1 \cdot F = F$, and $(\Phi F)(t - T) = (\Phi(t - T)) \cdot (F(t - T))$.

Show that if Φ is 0 on an interval $[a, b]$, then ΦF is 0 on the interval (a, b).

Exercise H8: Show that $t\delta = 0$, and $t^n \delta^{(n)} = (-1)^n n! \delta$.

Hint: One can calculate $t\delta^{(n)}$ using Leibniz's rule.

Suppose that Φ has $n + 2$ continuous derivatives,[587] and $\Phi(0) = \Phi'(0) = \dots = \Phi^{(n-1)}(0) = 0$. Find $\Phi(t)\delta^{(n)}$.

Exercise H9: Consider $M > 0$. Given a generalized function G which is 0 on $(-\infty, -M)$ and on (M, ∞), define $\int_{-\infty}^{\infty} G(t) dt$ as $C_+ - C_-$. Here C_- and C_+ are the (constant!) values on $(-\infty, -M)$ and on (M, ∞) of an arbitrary antiderivative of G.

Show that the average value of a 2π-periodic generalized function F coincides with[588] $\int_{-\infty}^{\infty} \Phi(t) F(t) dt$ provided Φ is smooth enough for the product to make sense, Φ vanishes outside of $[-2\pi, 2\pi]$, and $\Phi(t) + \Phi(t - 2\pi) = 1/2\pi$ on $[0, 2\pi]$.

Hint: Check separately the case of constant F, and of $F = H^{(n)}$ with a periodic continuous H.

Exercise H10: Assume that F is a 2π-periodic generalized function. Show that the average value of F' is 0.

Show that the Fourier coefficient F_k of a 2π-periodic generalized function F coincides with the average value of $e^{-ikt} F(t)$.

Hint: Same as in preceding exercise.

Exercises I: Convergence of generalized functions

Two preceding exercises address the first shortcoming mentioned after Exercise H5 on p. 179. To cover the second, consider the set C^{-m} of generalized functions consisting of mth (formal) derivatives of continuous functions. Say that a sequence $F_{[k]}$ of generalized functions from C^{-m} *converges in the*

[587] In fact, since δ is a measure, it is enough to have n continuous derivatives — but to define the product and verify this, one needs a delicate statement from analysis. Compare with the discussion after Exercise L17 on p. 195.

[588] One cannot replace this blindly with $\int_0^{2\pi} F(t) dt$: this expression is of the form $\int_{-\infty}^{\infty} \Psi(t) F(t) dt$ with Ψ being 1 between 0 and 2π and 0 outside of $[0, 2\pi]$. However, the product does not make sense unless the generalized function F is of a special form — because of the jumps of Ψ. For example, even though this expression can be assigned a certain sense when F is a measure, but then the result depends on the choice of the values of Ψ at 0 and 2π.

sense[589] [590] *of* C^{-m} to $F \in C^{-m}$ if on any finite interval $\mathscr{I}$ there is a choice of mth antiderivatives $G_{[k]}$ and G of $F_{[k]}$ and F such that $G_{[k]} \to G$ uniformly on $\mathscr{I}$.

Exercise I1: Given a sequence $F_{[k]} \to F$ as above and an m times continuously differentiable Φ, the sequence ΦF_k converges to ΦF in the sense of C^{-m}. If all $F_{[k]}$ are 0 outside of $[-M, M]$, then F has the same property.

If in addition the sequence of numbers $\int_{-\infty}^{\infty} F_{[k]}(t)\mathrm{d}t$ converges to L, then $\int_{-\infty}^{\infty} F(t)\mathrm{d}t = L$.

Convergence of 2π-periodic generalized functions in the sense of C^{-m} implies convergence of their kth Fourier coefficient for every k.

Exercise I2: The partial Fourier sums $\Sigma_{k=-K}^{K} F_k \mathrm{e}^{\mathrm{i}kt}$ of a 2π-periodic generalized function $F \in C^{-n}$ converge in the sense of C^{-n-2} to F.

Hint: Compare the case $n = 0$ with Exercise G0 on p. 177.

Similarly to the convergence of sequences in C^{-n}, one can consider convergence of families in C^{-n} parameterized by subsets of $\mathbb{R}$.

Exercise I3: If F is in C^{-n}, then the family $D_s := (F(t) - F(t-s))/s$ (with $s \neq 0$) in C^{-n} converges to F' in the sense of C^{-n-1}.

In fact, this works also when n is negative!

Exercise I4: If a family converges in the sense of C^{-n}, then it converges in the sense of $C^{-(n+1)}$.

This shows that it makes sense to use a figure of speach "converges in the sense of C^{-n} for a suitable n" (and then the convergence holds for larger n too). The shortcut is *converges in the sense of generalized functions.*

Exercise I5: A sequence F_k with $k \in \mathbb{Z}$ consists of Fourier coefficients of a 2π-periodic generalized function F iff there are numbers C, N such that $|F_k| \leq C \cdot (1 + |k|)^N$. Moreover, then $F \in C^{-N-2}$.

If $F \in C^{-m}$, then one can put[591] $N = m$.

One calls such sequences *tempered*. Note that this not only fully describes how the Fourier transform of 2π-periodic generalized functions behaves, but also identifies the set of 2π-periodic generalized functions with an easy-to-describe set of sequences.

Exercises J: Visualization of generalized functions

The introduction of the notion of generalized functions[592] completely changed the landscape of discussions in many domains of math. This makes it very important to have tools to *visualize* the given generalized function F (of the form $G^{(n)}$ with a continuous function G).

In the simplest case, F is continuous.

Exercise J1: Recall that any continuous function C can be considered as a formal 0th derivative $C^{[0]}$, so defines a generalized function. Say that a generalized function F is m times continuously differentiable on an interval (a, b) if there is an m times continuously differentiable (continuous) function C such that F coincides with C on (a, b). (Here $m \in \mathbb{Z}_{\geq 0}$.)

Show that then F' is $m - 1$ times continuously differentiable on (a, b) provided $m \geq 1$. Show that any antiderivative of F is $m + 1$ times continuously differentiable on (a, b).

---

[589] For a general sequence of generalized functions, it *converges* if on every finite interval there is a corresponding $m = m(\mathscr{I})$ such that there is convergence in the sense of C^{-m}.

In fact, one can even generalize the notion of a generalized function, by allowing its "degree of smoothness $-m$" to "depend on an interval": on different finite intervals one may have different generalized functions, but they should be undistinguishable where several make sense.

[590] **N.B. (???) Need to clarify the previous footnote.**

[591] Note the gap of 2 between translations $N \rightsquigarrow m$ and $m \rightsquigarrow N$. It is the same gap as in Exercise G0 on p. 177.

[592] ... and their multi-dimensional analogues — where one considers repeated (formal) partial derivatives of continuous functions, like $\partial^{k+l}F/\partial x^k \partial y^l$ with a continuous functions F (in the case dim = 2). Here F is considered up to the functions like $P(x)C(y) + c(x)p(y)$ with P, p polynomials of suitable degrees, and C and c continuous.

In such a case, the graph of F makes perfect sense — hence a visualization of F exists in the most "convincing" form.

Likewise for the case when F may be represented by a piecewise-continuous function C. (Here one can take the — continuous! — antiderivative G of C, and we assume that F is represented by the formal derivative $G^{[1]}$.) Already here one "annoyance" appears: G does not depend on the values of C at the points of jumps — so these points should better be excluded from the plot "since they do not contribute into F".

There are three workaround for the latter annoyance by normalizing the value at the points of jump:

Exercise J2: Suppose that a function F has the left and right limits at every point. Show that there is a function ${}_lF$ which is left-continuous (so ${}_lF(t_0) = \lim_{t\to t_0-} {}_lF(t)$) and which has the same left and right limits as F. Likewise for ${}_rF$ which is right-continuous.

Alternatively, one may replace the conditions above by ${}_mF(t_0) = \frac{1}{2}\Big(\lim_{t\to t_0-} {}_lF(t) + \lim_{t\to t_0+} {}_lF(t)\Big)$.

The last flavor is convenient [593]

The other flavors are especially important in the theory of random processes.[594]

One can show that if F allows an antiderivative G, then G is continuous, and ${}_lF$, ${}_rF$ and ${}_mF$ also allow G as an antiderivative. **Conclusion:** chosing one of l- or r- or m-flavors allows one to visualize *some of* (formal) derivatives of continuous functions. (Indeed, a function which has left and right limits at every point has a "non-pathological" graph — at least, the closure of the graph — taken as a subset of $\mathbb{R}^2$ — is [595] [596].)

On the next level of the complexity hierarchy, F is not of the forms considered above, but $F = G'$, and G has a non-pathological graph.

Exercise J3: Consider $G' = F = F_{\text{nice}} + F_\delta$ with F_δ a linear combination of shifts of δ-functions. If F_{nice} has left and right limits at every t then the coefficients c_k and shifts t_k of summands $c_k\delta(t-t_k)$ can be reconstructed from the jumps of G.

If F_{nice} has the left limit at t_0, then this limit equals the slope of left tangent ray to the graph of G at $(t, G(t-0)$. Likewise for the right limit.

Note that the context of this exercise is very close to what happens on the plot on p. 68. This plot has jumps, and it seems that at every point there is the "lower" and the "upper" left (and right) tangent rays. In other words, here instead of the left limit there is what seems to be a lower left limit and the upper left limit. Likewise, other "typical" fractal plots in the main body of our notes do not have jumps, but they also seem to have a "lower" and the "upper" left and right tangent rays.[597]

In yet higher levels of the complexity hierarchy, all we can plot is a *repeated* antiderivative of F. In such cases, the interpretation of features of these plots in terms of behaviour of F becomes more and more remote. For the second antiderivative, the corners match the δ-components of F and the jumps match δ'-components — but to say more is way harder; as the order of the antiderivative grows, the matching becomes harder and harder.

Another convenient way to visualize is applicable to generalized functions which are "δ-like" in the sense that they vanish outside of 0 (compare with Exercise H4 on p. 179). To proceed, define

[593] **N.B. (???) . . . due to Exercise L6 on p. 193.**

[594] . . . in the context of Doob's notion of "time-dependent filtration of σ-algebras".

[595] **N.B. (???) What? A very small subset of $\mathbb{R}^2$?**

[596] **N.B. (???) In fact, this works even when we allow the left and right limits to be $+\infty$ and $-\infty$.**

[597] One should be very careful here. These "lower" and "upper" tangent lines appear not due to F_{nice} having a lower left and the upper left limits, but because *suitably averaged* values of F having such limits. Even if we could assign values at points to our functions F, these values would be[598] unbounded on any interval.

[598] **N.B. (???) Convert to an exercise?**

the operation $\Re_s$ of rescaling s times; it acts on formal derivatives as $\Re_s(F^{(n)}) := F^{(n)}_{s,n}$ with[599] $F_{s,n}(t) := s^{n-1}F(t/s)$ (with $s \neq 0$).

Exercise J4: Show that rescaling indistinguishable formal derivatives gives indistinguishable formal derivatives.

Hence it makes sense to rescale generalized functions.

Exercise J5: Consider a continuous function c which vanishes outside of $[-M, M]$. Show that the family $\Re_s c$ has a limit when $s \to 0$ in the sense of C^{-m} for a suitable m. Find the smallest possible m. Show that the result of rescaling does not depend on c (up to a multiplicative constant).

This is how a lot of people "imagine δ-function": as something approximated by a very high and very narrow peak.

Exercise J6: Consider a continuous function F which vanishes outside of $[-M, M]$ and such that $\int_{-\infty}^{\infty} F(t)t^k \mathrm{d}t = 0$ for $0 \leq k < K$. Show that the family $\Re_s F/s^K$ has a limit when $s \to 0$ in the sense of C^{-m} for a suitable m. Find the smallest possible m. Show that the result of rescaling does not depend on F satisfying the conditions above (up to a multiplicative constant).

Remark 107: Note that if c in Exercise J5 has a continuous Kth derivative, then in Exercise J6 one can take[600] $F := c^{(K)}$. Comparing with Exercise I3, the limit of $\Re_s F/s^K$ must be the Kth derivative of the limit of $\Re_s c$.

Exercise J7: Show that the results of two preceding exercises hold (with possibly a different choice of m) if F is a generalized function.

Hint: Subtract a suitable continuous function $\widetilde{F}$.

Exercise J8: In the context of two preceding exercises, put $F(t) := \sum_{k=0}^{K}(-1)^k\binom{K}{k}\delta(t-k)$. Show that $\Re_s F/s^K \to \delta^{(K)}$.

Show that a piecewise-constant function $\overline{F}$ which vanishes outside of $[0, K+1]$ and is $(-1)^k\binom{K}{k}$ between k and $k+1$ gives the same limit as F.

Show that a continuous piecewise-linear function $\widehat{F}$ which vanishes outside of $[-1, K+1]$, whose graph has corners only over points $t \in \mathbb{Z}$ and such that $\widehat{F}(k) = (-1)^k\binom{K}{k}$ for $0 \leq k \leq K$, $k \in \mathbb{Z}$ gives the same limit as F. $k+1$ gives the same limit as F.

Remark 108: The functions $\overline{F}$ and $\widehat{F}$ can be described as *convolutions* of F "with suitable kernels": $\overline{F} = F \star H_1$ and $\widehat{F} = F \star H_2$ (moreover, H_2 is a translation of $H_1 \star H_1$). It turns out that the limits like $\Re_s F/s^K$ are compatible with convolutions; if one believes this, then to prove the last exercise, it is enough to apply Exercise J5 to H_1.

For numerically better convergence, one can translate F as $F(t + K/2)$. (This gives the same limit, but the convergence is — in many respects — much better.)

Remark 109: If fact, F itself may be described as a repeated convolution $f \star f \star f \star \ldots \star f$ (with K copies of f); here f is F constructed for $K = 1$.

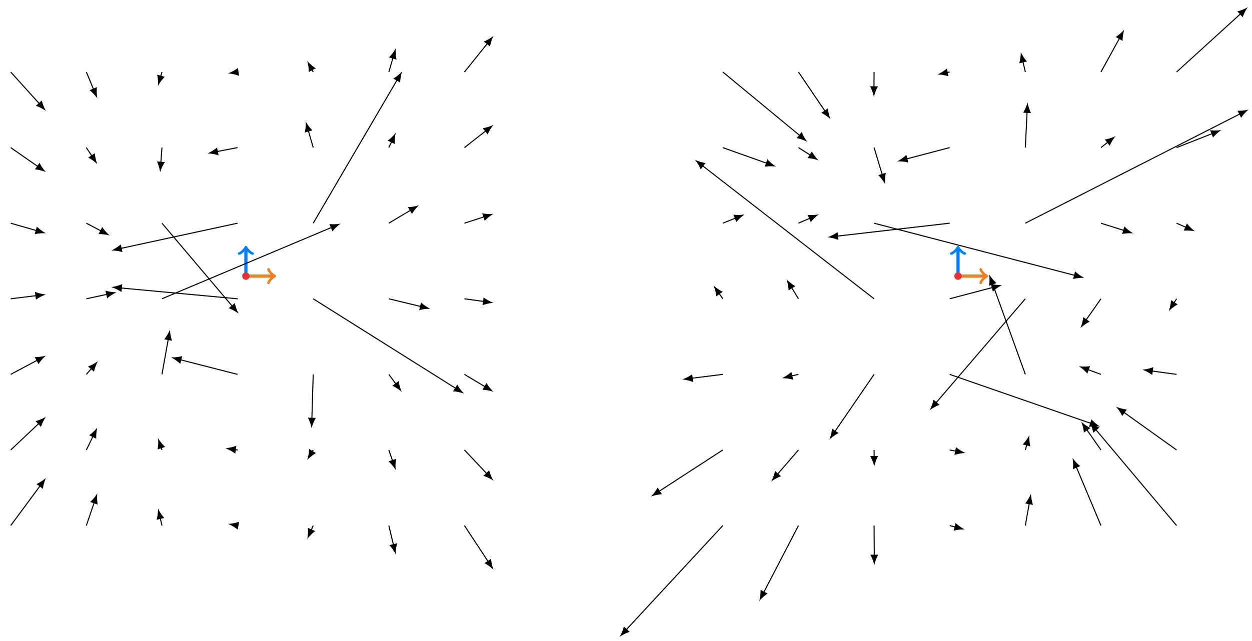
The physicists could use yet another way to visualize a generalized function. For example, in electrostatic the distribution of charges would often be described as a generalized function. Here pointwise charges correspond to δ-functions; simple-layer distributions of charges correspond to "δ-functions of a part of coordinates". Likewise, "dipoles" or double-layer distributions of charges correspond to derivatives of such δ-functions; "quadrupoles" correspond to (partial) 2nd derivatives of δ-functions.

The physicists visualize such distributions of charges by plotting "the electric force" created by these charges, or by plotting the "potential function" for these charges. In 1-dimensional cases this

[599] The extra -1 in the exponent $n-1$ "corresponds" to considering F "as a density $F(t)\mathrm{d}t$".

[600] Or one can proceed in the other direction: F must have a Kth antiderivative c which vanishes outside of $[-M, M]$.

corresponds to taking the 1st and the 2nd derivatives; however, in multidimensional cases this matches not the operators of taking a mixed derivative, but so-called "elliptic operators".



Exercise J9: On the plots above we show electric fields of two of three distributions of charges: a horizontal dipole ($-\partial\delta/\partial x$ or a rescaled $\ominus\oplus$); a vertical dipole ($-\partial\delta/\partial y$ or a rescaled $\genfrac{}{}{0pt}{}{\oplus}{\ominus}$); a quadrupole ($\partial^2\delta/\partial x\partial y$ or a rescaled $\begin{smallmatrix}\ominus\oplus\\ \oplus\ominus\end{smallmatrix}$). Recall that a δ-charge generates the electrical field $\mathbf{r}/r^3$ (here $\mathbf{r} := (x, y)$ is the radius-vector, and $r = |\mathbf{r}|$). Which of the plots matches which distribution?

(On these plots the direction of the field is correct, but the dependence of the magnitude on the distance is mollified a bit. Pay attention that the charges are at the red dot — which is not on the grid.)

Hint: Symmetries!

One more way of visualization is very relevant to the discussion in the main body of these notes. Jump to the next way of visualization if you are not confortable with analytic functions.

So consider an analytic function φ on $U \subset \mathbb{C}$ (such as $\mathfrak{H} := \{z \mid \operatorname{Im} z > 0\}$). Say that φ is *tempered on U* if there is a continuous function $C(x)$ on $\mathbb{R}$ such that $|\varphi(x + \mathrm{i}y)| < C(x)/y^n$ for an appropriate n when $0 < y < 1$. Below, "*tempered*" means "tempered on $\mathfrak{H}$". Recall that every analytic function on $\mathfrak{H}$ has an antiderivative — hence has repeated antiderivatives of arbitrary orders.

Exercise J10: If φ is tempered, then its antiderivative is tempered too. Moreover, one of the repeated antiderivatives of φ has a continuous extension to $\overline{\mathfrak{H}} := \{z \mid \operatorname{Im} z \geq 0\}$.

Exercise J11: Suppose that φ is tempered; write it as $\psi^{(n)}$ with ψ continuous on $\overline{\mathfrak{H}}$. Denote by G the (continous) restriction of ψ to $\mathbb{R} \subset \overline{\mathfrak{H}}$. Show that the generalized function $G^{(n)}$ does not depend on the choices of n and ψ.

The generalized function $G^{(n)}$ is called *the boundary trace value* of φ. Note that if φ itself allows a continuous extension to $\mathbb{R}$, then the boundary trace coincides with the limit values of φ:

Exercise J12: Suppose that φ is tempered and φ allows a continuous extension φ^+ to $\mathfrak{H} \cup \mathscr{I}$; here $\mathscr{I}$ is an open interval in $\mathbb{R}$. Then the boundary trace of φ is continuous on $\mathscr{I}$ and coincides with $\varphi^+|_{\mathscr{I}}$.

Note that one can treat $-\mathfrak{H} := \{z \mid \operatorname{Im} z < 0\}$ in the same terms.

Exercise J13: Consider a smooth function $S(t, w)$ vanishing for t outside of $[-M, M]$. For a continuous function $C(t)$ consider the generalized function $C^{(n)}(t)$ and the function $I(w) := \int_{-\infty}^{\infty} S(t, w) C^{(n)}(t) \mathrm{d}t$. Show that $I(w)$ is a smooth function of w, and $I^{(m)}(w)$ can be calculated by plugging $\partial^m S(t,w)/\partial w^m$ into the integral instead of S.

Hint: One can write $S(t,w) = \partial^n S_+(t,w)/\partial t^n + S_0(t)\alpha_0(w) + S_1(t)\alpha_1(w) + \ldots + S_{n-1}(t)\alpha_{n-1}(w)$ with a smooth $S_+(t,w)$ vanishing for $t \notin [-M,M]$ and suitable smooth S_k, α_k.

This leads to what is called "differentiation under the sign of integral".

Exercise J14: Consider a smooth function S on $\mathbb{R}$ which vanishes outside of $[-M,M]$. Define a function $S_z(x) := S(x)/(z-x)$, here $z \in \mathbb{C}$; obviously, it is a smooth function of x unless $z \in [-M,M] \subset \mathbb{R}$. Given a generalized function F, define $\varphi_{SF}(z) := \int_{-\infty}^{\infty} S_z(x)F(x)\mathrm{d}x$. Show that this is a continuos function of $z \in \mathbb{C} \smallsetminus [-M,M]$.

Differentiation in z under the sign of integral shows that $\varphi_{SF}(u+\mathrm{i}v)$ depends on u and v "in no more complicated way[601] than $1/(u+\mathrm{i}v-x)$". One can apply this to solve:

Exercise J15: Show that $\varphi_{SF}(z)$ is an analytic function of $z \in \mathbb{C} \smallsetminus [-M,M]$. Show that its restrictions to $\mathfrak{H}$ and to $-\mathfrak{H}$ are tempered functions.

Exercise J16: Denote by φ_+ and φ_- the boundary traces of restrictions of $\varphi_{SF}(z)$ to $\mathfrak{H}$ and to $-\mathfrak{H}$. Show that[602] $\varphi_+ - \varphi_- = -2\pi\mathrm{i}SF$.

Show that for any generalized function F and any interval $(-L,L)$ one can find tempered analytic functions Ψ_+ on $\mathfrak{H}$ and Ψ_- on $-\mathfrak{H}$ such that the difference $\varphi_+ - \varphi_-$ of their boundary traces φ_+ and φ_- coincides with F on $(-L,L)$.

In fact, this statement may be inverted. Given a tempered analytic function Ψ on $\mathbb{C} \smallsetminus \mathbb{R}$, consider its restrictions Ψ_+ to $\mathfrak{H}$ and Ψ_- to $-\mathfrak{H}$. Call the difference $\varphi_+ - \varphi_-$ of their boundary traces "*the jump of* Ψ *on* $\mathbb{R}$". The "magic of analytic function theory" implies:

Exercise J17: Consider a finite open interval $\mathscr{I} \subset \mathbb{R}$ and a tempered analytic function Ψ on $\mathbb{C} \smallsetminus (\mathbb{R} \smallsetminus \mathscr{I})$. Then the jump of Ψ on $\mathbb{R}$ vanishes on $\mathscr{I}$.

If φ_+ and φ_- in the context of the preceding exercise coincide on $\mathscr{I}$, then there is Ψ as above such that Ψ_+ is its restriction to $\mathfrak{H}$ and Ψ_- its restriction to $-\mathfrak{H}$.

Hint: Consider first the case when Ψ'_+ and Ψ'_- extend continuously to $\mathscr{I}$. Then apply "C^1 and Cauchy–Riemann equations imply analiticity".

Consider a finite open interval $\mathscr{I} \subset \mathbb{R}$ and tempered functions Ψ and $\widetilde{\Psi}$ on $\mathbb{C} \smallsetminus \mathbb{R}$. Say that these functions have "*manifestly the same jump on* $\mathscr{I}$" if $\Psi - \widetilde{\Psi}$ is a restriction of a function analytic on $\mathbb{C} \smallsetminus (\mathbb{R} \smallsetminus \mathscr{I})$. A "*manifestly defined jump on* $\mathscr{I}$" is described by a tempered function Ψ on $\mathbb{C} \smallsetminus \mathbb{R}$ *up to* changing Ψ to a function with manifestly the same jump on $\mathscr{I}$. Such jump *vanishes on* $\mathscr{I}$ if Ψ may be taken to be 0.

Exercise J18: Consider a manifestly defined jump on $\mathscr{I}$ described by a tempered analytic function Ψ on $\mathbb{C} \smallsetminus \mathbb{R}$. Associate to this description the generalized function F on $\mathbb{R}$ which is the jump of Ψ. Show that a different description $\widetilde{\Psi}$ of the same manifestly defined jump would lead to a generalized function $\widetilde{F}$ which coincides with F on $\mathscr{I}$.

Moreover, if tempered analytic function Ψ and $\widetilde{\Psi}$ on $\mathbb{C} \smallsetminus \mathbb{R}$ have jumps which are the same on $\mathscr{I}$ (as generalized functions), then Ψ and $\widetilde{\Psi}$ have manifestly the same jumps on $\mathscr{I}$.

Hint: Exercise J17.

Furthermore, for any generalized function F_0 on $\mathbb{R}$ one can find a manifestly defined jump on $\mathscr{I}$ whose associated generalized function F coincides with F_0 on $\mathscr{I}$.

Conclusion: manifestly defined jumps of tempered functions *are the same* as "restrictions of generalized functions to intervals". Since a generalized function is uniquely defined by its restrictions

[601] **N.B. (???) How to quantify this?**

[602] **N.B. (???) Need a hint? Compare with Footnote J19 on p. 186.**

to finite open intervals, one *could define* a generalized function as suitable equivalence classes of tempered functions[603] on $\mathbb{C} \setminus \mathbb{R}$. Such approach is very fruitful in many contexts.

Remark 110: For example, in the main body of the text we consider certain particular sequences encoding information from arithmetic problems, and take their Fourier transforms F. However, the typical approach is to consider F as a trace of a certain analytic function Ψ_+ on $\mathfrak{H}$, and work with Ψ_+ instead. (Note that F can be "made into a jump" by putting $\Psi_- := 0$ on $-\mathfrak{H}$.)

Exercise J19: Consider *the main branch* of the analytic function $\log z$ defined on $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$. Show that its antiderivative $z \log z - z$ allows continuous extensions to $\overline{\mathfrak{H}}$ and $-\overline{\mathfrak{H}}$, and its jump on $\mathbb{R}$ is $2\pi \mathrm{i} z$ on $(-\infty, 0]$ and vanishes on $[0, \infty)$.

Show that the jump of $\log z$ on $\mathbb{R}$ is piecewise-constant and is 0 on $(0, \infty)$ and $2\pi\mathrm{i}$ on $(-\infty, 0)$, and the jump of $1/z$ on $\mathbb{R}$ coincides with $-2\pi\mathrm{i}\delta$. Show that the boundary traces of $1/z$ from $\mathfrak{H}$ and from $-\mathfrak{H}$ are linear combinations of the δ-measure and the generalized function $1/x := (x \log|x| - x)''$.

Remark 111: Note that the latter representation "loses a bit of degree of smoothness". Indeed, $-2\pi\mathrm{i}\delta$ is a (complex-valued) measure — but when we represent it as a difference $\varphi_+ - \varphi_-$ of traces from $\mathfrak{H}$ and from $-\mathfrak{H}$, the components $\varphi_\pm$ are linear combinations of δ and the generalized function $1/t$ which is not a measure.

Exercise J20: If the jump on $\mathbb{R}$ of a tempered analytic function Ψ on $\mathbb{C} \setminus \mathbb{R}$ coincides as a generalized function with a real-analytic function φ on an interval $\mathscr{I} \subset \mathbb{R}$, then both the restriction Ψ_+ of Ψ to $\mathfrak{H}$ and the restriction Ψ_- to $-\mathfrak{H}$ allow analytic continuations through $\mathscr{I}$.

Hint: It is enough to find a particular representation of a manifestly defined jump.

Finally, we examine yet another way of visualization; it is applicable to certain generalized functions from C^{-2}. For these generalized functions, their antiderivative has left and right limits, and it is "not oscillating too much". The former conditions allow one to assign "weight" to any interval; the latter condition ensures that splitting an interval into a disjoint union of subintervals corresponds to addition of weights.[604]

Such generalized functions are called "measures". We study them in the group L of exercises on p. 191. (While "weights" as above cannot be immediately translated into pictures, they still provide a very tangible *corporeal* way of representing these generalizaed functions.)

Exercise J21: Given a closed interval $\mathscr{I}_\alpha$, divide it into 3 equal parts, and call the left and the right ones $\mathscr{I}_{\alpha 0}$ and $\mathscr{I}_{\alpha 2}$ correspondingly (assume them closed). Let $\mathscr{E}_\alpha$ be the "excluded" (open) middle interval. If the interval $\mathscr{I}$ has a certain weight w, assign to $\mathscr{I}_{\alpha 0}$ and $\mathscr{I}_{\alpha 2}$ weights $w/2$.

Now proceed with the intervals $\mathscr{I}_{\alpha 0}$ and $\mathscr{I}_{\alpha 2}$ likewise, then with the corresponding subintervals $\mathscr{I}_{\alpha 00}$ and $\mathscr{I}_{\alpha 02}$ and $\mathscr{I}_{\alpha 20}$ and $\mathscr{I}_{\alpha 22}$. Etc.

Start with $\mathscr{I} = [0, 1]$ with weight 1 and proceed as above. This assigns non-negative weights to subintervals like $\mathscr{I}_{0020022}$ etc. Call such intervals $\mathscr{I}_\alpha$. Show that for any t there is a (finite or infinite — or empty) sequence of indices $\alpha_1, \alpha_2, \ldots$ such that all intervals $\mathscr{I}_{\alpha_k}$ are inside $(-\infty, t)$, do not intersect, and any interval $\mathscr{I}_\alpha$ contained in $(-\infty, t)$ is a part of one of $\mathscr{I}_{\alpha_k}$.

Let $C(t)$ be the total weight of all intervals $\mathscr{I}_{\alpha_k}$. Show that C is between 0 and 1, and depends continuously on t.

Denote by $\mathfrak{C}$ *the Cantor set*: the points of $[0, 1]$ which are not in any one of intervals $\mathscr{E}_\alpha$. Show that $1/4 \in \mathfrak{C}$ and find $C(1/4)$. Show that the generalized function C' vanishes on all the intervals $\mathscr{E}_\alpha$, and the total length of these (non-intersecting!) intervals is 1. (Also, C' vanishes on $(-\infty, 0)$ and on $(1, \infty)$.)

---

[603] Weakening the assumption of "being tempered" leads to yet more general objects than generalized functions. Dropping this assumption altogether leads to the huge class of *hyperfunctions*. In multidimensional case, these require a more delicate description via " the edge of the wedge" approach.

[604] The *non-trivial* part is that this should work for infinitely (countably!) many subintervals too.

Show that points of $\mathfrak{C}$ are in 1-to-1 correspondence with infinite fractions $0.d_1d_2d_3d_4\dots$ in base 3 such that $d_k = 0$ or $d_k = 2$ for every k.

Hint: Pay attention that fractions 0.122222... and 0.2000... in base 3 denote the same real number.

(In fact, C' is a measure, and it vanishes on any interval which does not intersect $\mathfrak{C}$. The weight of any interval with ends a and b is $C(b) - C(a)$. In particular, the weight of any point is 0 — for measures, this is equivalent to the antiderivative being constant.)

Exercise J22: Do as in the preceding exercise, but for every interval $\mathscr{I}_\alpha = [a, b]$ choose arbitrarily[605] two non-intersecting subintervals $\mathscr{I}_{\alpha 0} = [a, x]$ and $\mathscr{I}_{\alpha 2} = [y, b]$ correspondingly. Assign them weights w_0 and $w_2 := w - w_0$ with an arbitrary $w_0 \neq 0, w$.

Proceeding as in the preceding exercise, now [606]

This provides examples of continuous functions C on $[0, 1]$ which have a well-defined derivative on a collection of subintervals with the total length 1. However, this derivative is always 0 — but C still grows from 0 to 1. (And still, C has no jumps!)

Such measures C' are called *singular measures*. In fact, a nice fact holds: for any measure μ one can find a converging linear combination μ_δ of translations of δ-measures, and a measure μ_0 with a density F such that "the rest" $\mu - \mu_\delta - \mu_0$ is a singular measure.[607]

(Unfortunately, this very strong and beautiful fact of analysis is not helpful for what we do in these notes: the functions we plot are more complicated than antiderivatives of measures.)

Exercises K: Meander wave

The exercises of this section focus on summing the Fourier series of the "rectangular=meander wave" (see Exercise K0 on p. 188). They form the first half[608] of explanation why "partial sums" of Fourier transform lead to pictures like what we can see on p. 62. On these plots one can see that the "spikes" on the blue and on the red plot have the same height, but are twice as wide on the blue plot. (Recall that the blue plot sums ½ of Fourier terms of the red plot.) The appearence of such spikes is called "Gibbs phenomenon".

The exercises demonstrate the utility of switching between derivatives and antiderivatives in dealing with Fourier series of non-smooth functions, and may suggest why "the sine integral" $\mathrm{Si}(x)$ and its derivative $\mathrm{sinc}(x)$ play so important roles in the theory of Fourier series. The base is the formulas for how the Fourier transform (the summation of Fourier series) interacts with taking derivative and antiderivative. When the Fourier transform is $F(t) := \sum_n a_n e^{int}$, we denote by $F_{\leq N}$ the result of summation with n restricted to $-N \leq n \leq N$ ("partial sums" of Fourier series). For such partial sums, the formulas for derivatives/antiderivatives in t hold trivially: applying these operations to $F_{\leq N}(t)$ corresponds to[609] multiply/divide a_n by in.

Hint: To deal with integration constants below, it may help to be able to recognize when the sum of a Fourier series is odd, and when it is even.

Depending on one's tastes, it may be easier to do generalized functions first, or this section (and the next) first. Another approach is to skim through this (and the next) section to understand the *style* of these problems, then switch to the group H of exercises on Fourier transform of generalized

[605] In fact, one can allow these intervals to not cantain a and b — but this way one does not get any "extra" Cantor sets.

[606] **N.B. (???) When the process converges and C is continuous?**

[607] Moreover, F coincides with the "usual" derivative of the antiderivative M of the measure μ. In particular, this means that singular measures coincide with measures whose antiderivative is continuous (this implies $\mu_\delta = 0$), but its derivative is "equal to 0 almost everywhere" — or "0 in the sense of L^1-functions".

[608] **N.B. (???) The second half?**

[609] Here a_0 is exceptional. For the anti-derivative to have a Fourier series it must be periodic, hence $a_0 = 0$ must hold. — Hence one may *assume* that $a_0 = 0$. Then in the formula for the antiderivative one gets $0/i0$ when $n = 0$ — which is convenient to resolve as 0. (For example, this allows taking repeated antiderivatives. — In other respects, a particular choice for $0/i0$ in this formula does not matter: any RHS will result in a "correct" antiderivative!)

functions on p. 178 — and only after covering them, return to *solving* exercises here. (Moreover, keep in mind that we do not use the spirit and the results of these exercises directly in the main body of our notes.)

Exercise K0: Show that "the meander wave" $M(x)$ (or "the rectangular wave") is the Fourier transform of $a_n := -\mathrm{i}/n$ for odd n, with $a_{2k} := 0$. Here the 2π-periodic function M is given by $M(x) = \pi/2$ for x in $(0, \pi)$, $M(x) = -\pi/2$ for x in $(-\pi, 0)$, $M(x) = 0$ for $x = 0, \pi$.

Hint: The easiest way is to use the *inverse* Fourier transform.

Exercise K1: Show that $M_{\leq 100}(x)$ is odd, and $M_{\leq 100}(x + \pi/2)$ is even.

For the following exercises it makes sense to inspect $M_{\leq 100}(x)$ with plotting software on intervals $[0.01, 0.1]$, $[0.01, 0.4]$, and $[1, 2]$. (In GP/PARI, one can use

```
plot(x=0.01,0.4,sum(n=-100,100,if(n%2,real(-I*exp(I*n*x)/n))))
```

— or one can use `ploth()` instead.) In the exercises below it is OK to estimate values of elementary functions using calculators (as opposed to "the real aces", who would estimate these values from the definitions ;-)).

Exercise K2: Write a formula for $M'_{\leq 100}(x)$ in terms of trigonometric functions. Show that $M'_{\leq 100}(x)$ oscillates for $x \in [1, 2]$ with amplitude close to 1. Find "the period" of the oscillations (this is a figure of speech — they are not periodic in the strict sense).

Exercise K3: (Here we assume a "heuristic" approach. We will ask the same question later[610] requiring complete explanations — but with a lot of hints!) Explain why $M_{\leq 100}(x)$ for $x \in [1, 2]$ oscillates about a certain value with amplitude close to 0.01.

Exercise K4: Using $\sin x \approx x$ for small x, sketch the plot of $M'_{\leq 100}(x)$ for $x \in [-0.07, 0.07]$.

Exercise K5: Using the result of the preceding exercise, sketch (very roughly is OK!) the plot of $M_{\leq 100}(x)$ for $x \in [-0.07, 0.07]$.

Exercise K6: Show that for $a = \pi k/100$, $b = \pi(k + 1)/100$ with k such that $a, b \in [1, 2]$, one can write $\int_a^b M'_{\leq 100}(x)\mathrm{d}x = \pm E/100$ with E between 2 and 3.

Exercise K7: How much can one improve the estimate 3 in the preceding exercise? (This assumes "with minimal modifications of the arguments"!)

Exercise K8: Denoting the integral in Exercise K6 as I_k, show that $|I_k + I_{k+1}| \leq 0.00045$ under the assumptions of that Exercise.

Hint: Use the formula for $\sin x - \sin y$.

Exercise K9: Show that $|\int_a^b M'_{\leq 100}(x)\mathrm{d}x| \leq 0.04$ for $a, b \in [1, 2]$.

Exercise K10: Show that $M_{\leq 10{,}000}(x)$ for $x \in [1, 2]$ is within 0.0002 of a certain constant.

In the exercises below, we say that a shape S is "within distance ε" from a shape T if for any point of T, there is a point of S at the distance $\leq \varepsilon$ (and the same for S and T exchanged). The minimal working value of ε is usually called "*the Hausdorff distance*" between S and T.

Exercise K11: Given 3 shapes R, S, T, show that Hausdorff distance satisfies the triangle unequality. In other words, $\mathrm{distance}(R, T) \leq \mathrm{distance}(R, S) + \mathrm{distance}(S, T)$.

Exercise K12: A certain "simple" shape is within distance 0.0002 from the plot of $M_{\leq 10{,}000}(x)$ on the interval $[1, 2]$. Find this shape (it is OK to find it "up to a small parallel translation in vertical direction").

Exercise K13: Proceeding as in Exercises K7, K8, but for $M'_{\leq m}(x)$ instead of $M'_{\leq 100}(x)$, show that $|I_k + I_{k+1}| \leq 1/2k^2$, and $|I_k| \leq 4/(\pi k)$.

Exercise K14: Show that $\left|\int_a^b M'_{\leq m}(x)\mathrm{d}x\right| \leq (3t + 1)/m$ for $a, b \in [1/t, 2]$; here $t \geq 1$.

[610] Starting with Exercise K6.

Exercise K15: Using the results of Exercise K14 on p. 188, how can one extend the interval $[1, 2]$ in Exercise K12 on p. 188 if one allows being "within distance 0.02"?

Exercise K16: With notations as in Exercise K8 on p. 188, show that $|I_{2k-1}| > I_{2k}$.

Exercise K17: Show that $\max M_{\leq 100}(x)$ on $[0, \pi/2]$ is achieved at $x = \pi/100$.

Exercise K18: Show that for a certain C the union of the solution to Exercise K15 and the interval $[0, C]$ on the y-axis is within distance 0.02 of the graph of $M_{\leq 10{,}000}(x)$ on $[0, 1]$. Find an appropriate C.

Exercise K19: Show that $C = \int_0^\pi \sin(x)\mathrm{d}x/x$ works as a solution of Exercise K18.

Exercise K20: Show that there is a shape S such that for every given distance $\varepsilon > 0$, the graphs of $M_{\leq m}(x)$ for $x \in [-10, 10]$ is within distance ε of S provided that m is large enough.

In fact, if one requires that S "has no holes" (in the language of topology, this is "closed") then such S in Exercise K20 is unique. In topology, such a shape as S is called "the limit of graphs in Hausdorff topology". The effect illustrated in Exercise K20 (more precisely, that S is "not a rectangular wave") is called "*the Gibbs phenomenon*".

Above, we essentially described the Gibbs phenomenon up to precision $2/\sqrt{m}$. In fact, there is a much finer description which works up to precision about $4/m$.

Exercise K21: Show that $f(x) := 1/\sin(x) - 1/x$ is a smooth function on the interval $[0, \pi/2]$. (This assumes knowledge of Taylor series — or a very skillful integration by parts. People who do not know this may skip this exercise and consider its result as "given" for what follows.)

Below, we use notations $D_0 := \max |f(x)|$, $D_1 := \max |f'(x)|$, $D_2 := \max |f''(x)|$ on $[0,\pi/2]$ (with f from the preceding exercise). (Using plotting software, one can "be convinced" that the maxima are achieved at $\pi/2$, and are less than ½. In fact, one can omit the absolute values in the descriptions above.)

Exercise K22: Show that there is a function $g(x)$ such that $|M_{\leq m}(x/m) - g(x)| \leq D_0/m + (D_1 + D_2 x/m)/m^2$ for any $m \geq 1$ and x in $[0, m\pi/2]$. Using the estimates above, show that $|M_{\leq m}(x/m) - g(x)| \leq 2/m$. (Again, if one is too lazy to do the required integrations by part, one may try to guess $g(x)$, check with plotting software, and consider this result as "given" for what follows.)

Hint: Differentiate.

In the following exercise, use the fact[611] that if M is smooth near x_0, then the Fourier approximtion $M_{\leq m}(x_0)$ must get closer to $M(x_0)$ as $m \to \infty$.

Exercise K23: Show that $\int_{-\infty}^{\infty} \sin(x)\mathrm{d}x/x = \pi$.

Hint: Choose x_0 and use the result of the preceding exercise and Exercise K0 on p. 188.

Inspecting the plot of "the sine integral" $\operatorname{Si} x := \int_0^x \sin t\mathrm{d}t/t$ in a plotting software, one can notice that the maxima and minima of this plot are close to the curves[612] $y = \pi/2 \pm 1/x$. So this function is "close to" being squeezed between these curves. Below, assume that this holds: in other words, that $|\operatorname{Si} x - \pi/2| \leq 1/(x - \alpha)$ for all $x > \alpha$, while $|\operatorname{Si} x - \pi/2| \geq 1/(x + \alpha)$ for an appropriate value of x in every interval $[x_0, x_0 + 2\pi]$ with $x_0 \geq 0$. Here we assume that α is an appropriate "small" number (it is about 0.3).

Remark 112: One can do this in `GP/PARI` by `ploth(x=1,10,[real(Si(x)),Pi/2+1/x,Pi/2-1/x])`. Here `Si` is defined (after putting `Eps_=1e-18`) in https://oeis.org/w/images/2/21/LiEiRelatedFunctions.txt.

Exercise K24: Consider the shape S defined in Exercise K18, restricted to x in $[0, \pi/2]$. Show that by extending S by the zone between graphs of $\pi/2 \pm 1/(mx)$ for x in $[2\pi/m, \pi/2]$ the resulting shapes are within $4/m$ of the graph of $M_{\leq m}(x)$ on $[0, \pi/2]$. Here $m = 10{,}000$.

[611] As an alternative, one can use the fact $1 - 1/3 + 1/5 - 1/7 + \ldots = \pi/4$. To understand this, note that the antiderivative of $1 - x^2 + x^4 - x^6 + \ldots$ is $\operatorname{atan} x$.

[612] **Question:** how this $\pi/2$ is related to Exercise K23?

Exercise K25: Show that the method of Exercise K24 on p. 189 works in the context of Exercise K20 on p. 189 as well with $\varepsilon = 4/m$.

Exercise K26: Show that Exercise K18 on p. 189 can be amplified: 0.02 may be replaced by 0.0104.

Exercise K27: Show that in Exercise K18 on p. 189 one cannot replace 0.02 by 0.0096.

Exercise K28: Consider the odd 2π-periodic function W which coincides with $\pi - x$ for $x \in (0, 2\pi)$ (the "triangular wave"). Put $W_{\{m\}}(x) := 2(\sin x + {}^1\!/\!_2 \sin 2x + {}^1\!/\!_3 \sin 3x + \ldots + {}^1\!/\!_m \sin mx)$. As in the exercises above, show that $W_{\{m\}}(x) \to W(x)$ as $m \to \infty$ for every x.

Show that the function $W_{\{m\}}(x) - W(x)$ of x behaves very similar to what we found about $M_{\le m}(x) - M(x)$.

Hint: Write a formula for $(W_{\{m\}}(x) + x)'$.

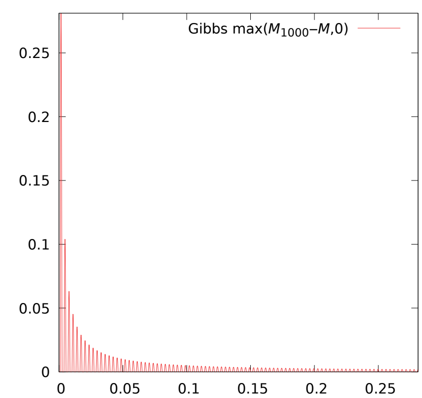
[613]

Exercise K29: Show that $W_{\{m\}}$ is the mth partial Fourier sum $W_{\le m}$ for W.

Show that $W_{\le m}(x) - W_{\le m}(x + \pi) = 2M_{\le m}(x)$. (Hence $M_{\le m}(\pi/2) \approx \pi/2$.)

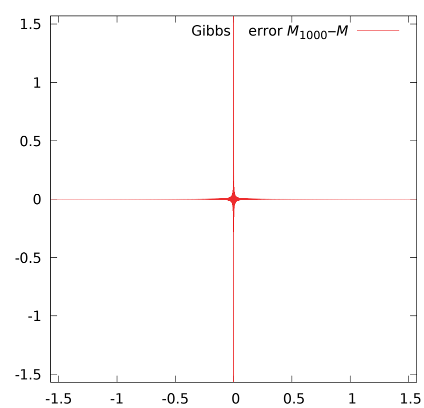
Remark 113: Note that only in Exercises K0 on p. 188 and K23 we refer to the fact that the inverse Fourier transform inverts the summing of Fourier series. If we ignore these exercises, we still gave a rather detailed description of partial sums $M_{\le m}(x)$ of the Fourier series with the coefficients a_n given in Exercise K0 on p. 188.

In particular, we have shown that $M_{\le m}(x)$ is close to a meander wave of a certain height, and we know that this height is given by the integral in Exercise K23 on p. 189. **Conclusion:** the "non-ignored" exercises prove that the Fourier transform of the inverse Fourier transform of M is *proportional to* M. The "ignored" exercises allow us[614] to calculate the integral given that *actually* $M_{\le m} \approx M$.



If one reflects a graph of an oscillating function w.r.t. the bisectrix of the first quadrant, the resulting curve obviously cannot be a graph of a function. However, the following "mathematical jokes" show that if one laces in the Hausdorff distance, this may still happen:

Exercise K30: Given a symmetry σ sending S to σS, say that a shape S is ε-approximately-symmetrical w.r.t. σ if σS is within distance ε from S. Show that the plot of $\max(M_{\le 10{,}000,0}(x) - \pi/2)$ for x in $[0, C - \pi/2]$ is 0.001-approximately-symmetrical w.r.t. reflection σ in the bisectrix of the first quadrant. Here C is defined in Exercise K19 on p. 189 (so $C - \pi/2 \approx 0.28114$).



The plot above illustrates what happens in the last exercise — only for $M_{\le 1{,}000}$ instead of $M_{\le 10{,}000}$. Likewise, the plot on the right illustrates the next one (with the same substitution):

Exercise K31: Is the conclusion of Exercise K30 still applicable to the graph of $M_{\le 10{,}000}(x) - M(x)$ for x on $[-\pi/2, \pi/2]$?

Finally, the following exercises have no relationship to the main body of the text: they only clarify another facet of the notion of Hausdorff distance introduced above.

Exercise K32: Given triangles $\triangle A_1A_2A_3$ and $\triangle B_1B_2B_3$ with $|A_kB_k| \le 1$ for every $1 \le k \le 3$, shows that $\text{Hausdorff\_Distance}(\triangle A_1A_2A_3, \triangle B_1B_2B_3) \le 1$.

Exercise K33: Find triangles $\triangle A_1A_2A_3$ and $\triangle B_1B_2B_3$ with $\text{Hausdorff\_Distance}(\triangle A_1A_2A_3, \triangle B_1B_2B_3) = 1$ such that $|A_kB_l| \ge 2$ for every $1 \le k \le 3$ and $1 \le l \le 3$.

[613] **N.B. (???) See section on Cesàro.**

[614] The alternative way is Exercise K29.

The preceding exercise works with both notions of triangle: “the filled triangle”, and “the closed broken line”.

```
set samples 5000
set size ratio -1
unset autoscale
set autoscale
set autoscale noextend
set output "gibbs-diff.pdf"
set term pdfcairo size 36,35 round fontscale 5
max(a,b)=a>b?a:b
diff(x,N)=-pi/2*sgn(x)+2*(sum [i=1:N] sin((2*i-1)*x)/(2*i-1))
tit(pre,N,post)=sprintf("Gibbs%s{/:Italic M}_{%d}-{/:Italic M}%s",pre,N,post)
N=1000; plot [-pi/2 : pi/2] sample [-pi/2 : -pi/2*2/N] diff(x,N) title tit("    error ",N,"") lc "#ed2d2e", [-pi/2*2/N : pi/2*2/N]
set output "gibbs-above.pdf"
set term pdfcairo size 7.2,7 round fontscale 1
plot [0 : 0.28114] max(0,diff(x,N)) title tit(" max(",N,",0)") lc "#ed2d2e"
unset output
unset term
```

Exercises L: Fourier transform and measures

Here we discuss the language which predates historically the language of generalized functions. The latter one turns out to be much simpler (we handle this in the group H of exercises on p. 178), but it may feel a bit too abstract, and may require a little bit more effort to visualize.

Depending on one’s tastes, it may be easier to do generalized functions first, or this section first. Another approach is to skim through this section to understand the *style* of these exercises, then switch to the group H of exercises on Fourier transform of generalized functions on p. 178 — and only after covering them, return to *solving* exercises here. (Moreover, keep in mind that we do not use the spirit and the results of these exercises directly in the main body of our notes.)

In the group G of exercises on p. 176 and in the preceding section, we investigated the case of Fourier transforms of sequences which were decaying relatively quick. The following exercise just recalls a particular case of Exercise G0 on p. 177:

Exercise L1: Given a sequence (a_n) such that $|a_n| \leq C/|n|^r$ for $n \neq 0$, its Fourier transform F is continuous provided $r > 1$.

Hint: Show that $|F_m(x_1) - F_m(x_2)| \leq 2S_m \sin\min(\pi/2, m(|x_1 - x_2|/2)$ with $S_m := \sum_{k=-m}^{m} |a_k|$.

One cannot extend this to $r = 1$: we already saw such examples where the function F had jumps — but in our examples, F had left and right limits at every point. Unfortunately, in general with $r = 1$ it is not only possible that F has pretty bad behaviour — but it may happen[615] that the series in $F(x) := \sum_n a_n e^{inx}$ is diverging for every x.

Nevertheless, at least in some cases[616] one can discuss “what is F” even if the series above diverges! To explain, we need to allow F to be a new kind of object, which — like a function — may be thought of as “depending on x”, but it may “not have a value anywhere”. Our first aim is to describe objects $F(x)$ of this kind *which still have well-defined average value on intervals.*

The latter condition can be restated as $\int_a^b F(x)\mathrm{d}x$ *making sense.* (Another way to say this is “F has an antiderivative well-defined at every point”.) **Our first aim** is to cover the subcase when $F(x)$

[615] **N.B. (???) Is it happening for our Eisenstein examples? Probably Kolmogorov’s example is not having $r = 1$?**

[616] **N.B. (???) Check!**

may be defined as a number for every x. Such functions F which "may be integrated" are called *L^1-integrable* (locally on every finite part of $\mathbb{R}$).[617] **Example:** Any continuous function $F(x)$ is such.[618]

Remark 114: Already for L^1-integrable functions the notion "of a value at a point" becomes blurry. Indeed, changing the value of F at one particular point does not change $\int_a^b F(x)\mathrm{d}x$, hence would not change the average values on intervals. **Conclusion:** If we are interested *in the averaged values only*,[619] then $F(x)$ is not uniquely determined by these values.

In fact, if F has a limit at x, then one *can require*[620] that $F(x)$ coincides with this limit. Note that if this holds *for every* x, then F can be chosen to be continuous. **Conclusion:** if F is continuous, then it is uniquely determined by its antiderivative[621] $F^{(-1)}$.

Our **next aims** are to consider "the general situation in which an antiderivative makes sense". Suppose that to every (finite) interval $\mathscr{I}$ of $\mathbb{R}$ we assign "its weight" $\mu(\mathscr{I})$, and if $\mathscr{I}$, $\mathscr{I}_1$, $\mathscr{I}_2$ are intervals such that $\mathscr{I} = \mathscr{I}_1 \cup \mathscr{I}_2$ and $\mathscr{I}_1 \cap \mathscr{I}_2 = \varnothing$, then "additivity" $\mu(\mathscr{I}) = \mu(\mathscr{I}_1) + \mu(\mathscr{I}_2)$ holds. Here we allow "an interval" to be closed or open — independently on the left and the right side. We also allow intervals $[b, b] := \{b\}$ made of one point b. (One can think of μ as assigning every interval "its weight w.r.t. a certain distribution of mass".[623]) One calls μ *an additive function of intervals.*

With this language, **start with** generalizing "non-negative functions". A "non-negative measure" should satisfy $\mu(\mathscr{I}) \geq 0$ for every $\mathscr{I}$, and one more "non-pathology" condition:[624] given any x, by a suitable choice of $\varepsilon > 0$

$$\boxed{\text{One can make } \mu((x, x+\varepsilon)) \text{ and } \mu((x-\varepsilon, x)) \text{ arbitrarily close to 0.}}$$

Example: The δ-measure assigns 1 or 0 to $\mathscr{I}$ depending on whether $0 \in \mathscr{I}$. **Example:** given a continuous (or L^1-integrable) function $F(x) \geq 0$, assign to any (closed/open/etc.) interval with ends a and b the weight $\int_a^b F(x)\mathrm{d}x$. As "distributions of masses", they describe a point-mass at $x = 0$ on a massless wire, and a wire with variable cross-section $F(x)$. One calls F *the density* of the corresponding measure μ.

Exercise L2: Show that these examples define non-negative measures.

Remark 115: The example of δ-measure μ_δ may be thought of "having an infinite density at $x = 0$": away from 0 the density is obviously 0, but "the total integral should be[625] 1". So when the density makes sense, it may be thought of as "value at a point" of the measure μ — but there may be points

---

[617] Here we again cheat a bit: to be honest, we need to first describe "in which sense we understand the integral". The notion of L^1 corresponds to the so called *Lebesgue integrability* — the most widely used flavor of integrability nowadays.

In addition to the "fundamental properties of measures" listed below, we use the following properties of integrability: that it is closed under addition, under multiplication by a piecewise-continuous functions, under inf of a sequence of non-negative functions. We also use the facts that antiderivative is linear, and if infinum above is 0, then any definite integral goes to 0.

[618] Moreover, in this case the integral "works in the Riemann sense". — Or: it may be described in the "high-school's language".

[619] Which is the same as being interested only in the "antiderivative $F^{(-1)} := \int_{-C}^{x} f(y)\mathrm{d}y$ of F".

[620] ... without changing the average values of F on the intervals.

[621] In general one F is uniquely defined by $F^{(-1)}$ *away from a "very small" set of points x*. The exceptions make "a set of *measure 0*". The way to reconstruct $F(x)$ "outside of this small set" is to take the derivative of $F^{(-1)}$ — but one can do this only at points x where[622] this derivative makes sense!

[622] If F is continuous, then $F^{(-1)}$ has a derivative everywhere. So this reconstruction of F is possible everywhere — since this derivative coincides with F.

[623] The physicists would prefer to call it "a distribution of charges" — since $\mu(\mathscr{I})$ may be negative. They would reserve the term "distribution of mass" to the case of non-negative measures considered below.

[624] We clarify one aspect of this condition in Footnote 628 on p. 193.

[625] If one replaces the point 0 by an interval $[-\varepsilon/2, \varepsilon/2]$, then to make the total weight 1, the density should be $1/\varepsilon$ on this interval. To "approximate" δ-measure, ε should get closer and closer to 0 — and note how this leads to the density growing to ∞.

where the density "is infinite" — or it may be "just undefined" at some points, as in Footnote 621 on p. 192.

Exercise L3: Show that an additive function μ of intervals is uniquely determined by its values on closed intervals. Same for the open intervals. Same for the values of μ on open-on-the-left and closed-on-the-right intervals, *and* on intervals $[b, b]$ for all b.

Exercise L4: Given a non-negative measure μ and an interval $\mathscr{I}_0$, there is a unique non-negative measure μ_0 such that $\mu_0(\mathscr{I}) = \mu(\mathscr{I})$ if $\mathscr{I} \subset \mathscr{I}_0$ and $\mu_0(\mathscr{I}) = 0$ if $\mathscr{I} \cap \mathscr{I}_0 = \varnothing$.

The same holds for additive functions of intervals.

In fact, the same holds for measures (which we introduce below). One calls μ_0 *the restriction* $\mu|_{\mathscr{I}}$ *of* μ *to* $\mathscr{I}$.

Exercise L5: Given a function U, assign to the interval $(a, b]$ its weight $\mu((a, b]) := U(b) - U(a)$.

(1) Then μ can be extended to an additive function μ of intervals.

Hint: The choice of $\mu([b, b])$ is arbitrary.

(2) Moreover, there is a non-negative-measure choice of such μ iff U is non-decreasing and continuous on the right (and then such μ is unique).

Hint: One must investigate "other kinds" of intervals!

(3) Every non-negative measure μ may be obtained in such a way. Moreover, the corresponding U is defined uniquely up to addition of a constant.

The condition of (2) are not symmetrical. However, there is a way to fix this:

Exercise L6: Given a function $\widetilde{U}$, say that it is an *antiderivative* of a measure μ if $\mu([a, b]) + \mu((a, b]) + \mu([a, b)) + \mu((a, b)) = 4(\widetilde{U}(b) - \widetilde{U}(a))$ for any $a \leq b$. (Here $(a, a) := \varnothing$.)

(1) Every non-negative measure μ has an antiderivative. This antiderivative is defined uniquely up to addition of a constant.

(2) Every right- and left-continuous non-decreasing function $\widetilde{U}$ such that $\widetilde{U}(x) = ½(L(x) + R(x))$ is an antiderivative of a uniquely determined non-negative measure μ. Here $L(x)$ is the left limit of $\widetilde{U}$ at x, and $R(x)$ is the right limit.

(3) Moreover, then $\mu((a, b)) = L(b) - R(a)$, and $\mu([b, b]) = R(b) - L(b)$.

One calls a non-negative measure μ as in the last Exercise *the formal derivative* of the function $\widetilde{U}$. The following exercise shows that this is a direct generalization of the "usual" notion of derivative.

Exercise L7: If $\widetilde{U}$ has a continuous derivative F, then[626] μ from Exercise L6 is obtained from F as in Exercise L2 on p. 192.

For **the last step** towards definition of "measures", say that an additive function μ of intervals *is a measure* if $\mu = \mu^+ - \mu^-$ with non-negative measures μ^+, μ^-. For measures, the notion of antiderivative still makes a perfect sense. However, characterising which functions are antiderivatives of measures is much more complicated than what we do[628] in Exercise L6(2).

---

[626] In fact, the same holds if μ has a density. — In other words, $\widetilde{U}$ has a derivative F "almost everywhere", F is L^1-integrable, *and* $\widetilde{U}$ *coincides with the antiderivative of* F *constructed via integration*[627]. (The latter condition is not void!) In analysis, such functions U are called "*absolutely continuous*".

[627] **N.B. (???) The classical definition of antiderivative requires derivative to exist in every point. Correct in other places too?**

[628] *This* is the reason why we start with non-negative measures first. The blue-framed property on p. 192 is significantly simpler than what is required to characterise the general case.

The corresponding general property of antiderivatives is called having "*bounded variation*".[629] The corresponding *property of measures* is the so-called *σ-additivity*: it is similar to additivity, but we allow splitting an interval into a disjoint union of not 2 intervals, but of countably many intervals.

[629] Another way to say this is that "functions U of bounded variations have a formal derivative, this derivative is a measure, and U is one of the antiderivatives $U + \mathrm{const}$ of this measure.

To clarify the notion of a measure, say that $\mu \geq \mu_0$ (with functions of intervals μ and μ_0) iff $\mu(\mathscr{I}) \geq \mu_0(\mathscr{I})$ for every interval $\mathscr{I}$. One of the fundamental facts about non-negative measures is that given two such measures μ_1 and μ_2, there is a measure $\min(\mu_1, \mu_2)$ which satisfies $\min(\mu_1, \mu_2) \leq \mu_k$ for $k = 1, 2$, and

$$\boxed{\min(\mu_1, \mu_2) \geq \mu \text{ for any measure } \mu \text{ such that } \mu \leq \mu_1 \text{ and } \mu \leq \mu_2.}$$

(so it is *a largest* measure which is below μ_1 and μ_2).

Exercise L8: Given a measure μ there are unique non-negative measures μ_+ and μ_- s.t. $\mu = \mu_+ - \mu_-$ and $\min(\mu_+, \mu_-) = 0$. A sum or difference of measures is a measure. A measure multiplied by a number is a measure.

In the conditions of this exercise, $|\mu|$ is defined as $\mu_+ + \mu_-$.

Exercise L9: If μ has density F, then $|\mu|$ has density $|F|$. If $|\mu|$ has density D, then[630] μ has a density.

Define $\max(\mu_1, \mu_2) = \mu_1 + \mu_2 - \min(\mu_1, \mu_2)$. Show that $|\mu| = \max(\mu, -\mu)$.

Exercise L10: Show that there is no measure with the antiderivative U coinciding with $x \sin(1/x)$ for $x \neq 0$.

Hint: Consider the *total* increment of U on a suitable collection of intervals.

Show that if the antiderivative of a measure coincides with $xG(1/x)$ for $x \neq 0$ and G is periodic, then $G = \text{const}$.

Hint: Consider oscillations of $G(x)/x$ for $|x| \gg 0$.

Exercise L11: Say that an additive function of intervals is *T-periodic* if $\mu(\mathscr{I}) = \mu(\mathscr{I}_T)$ for any $\mathscr{I}$. Here $\mathscr{I}_T$ is the parallel translation of $\mathscr{I}$ by the distance T. If μ is T-periodic, then $\mu((x, x+T]) = \mu([x, x+T)) = \mu([0, T))$ for every x.

Define the *norm* $\|\mu\|$ of a T-periodic[631] measure μ as $|\mu|([0, T))$. Consider $\|\mu - \mu_0\|$ as the *distance*[632] between two T-periodic measures μ and μ_0.

Exercise L12: Given three T-periodic measures μ_k, $k = 1, 2, 3$, show that $\|\mu_1 - \mu_3\| \leq \|\mu_1 - \mu_2\| + \|\mu_2 - \mu_3\|$. Show that $\mu_1 = \mu_2$ iff $\|\mu_1 - \mu_2\| = 0$.

Since we can measure distances, we can consider infinite sums: we say that $\mu_1 + \mu_2 + \dots$ *ultra-converges* to μ if $\|\mu - (\mu_1 + \mu_2 + \dots + \mu_n)\|$ decreases to 0 as n grows. If measures μ_k have densities F_k, we can also write down the LHS as[633] $F_1 + F_2 + \dots$.

Exercise L13: Show that if a T-periodic measure μ has a density, then $\|\delta_{\text{per}} - \mu\| \geq 1$. Here $\delta_{\text{per}}(x)$ is the δ-measure at 0 extended outside of $(-2\pi, 2\pi)$ by 2π-periodicity.

In particular, $\cos 0x + 2\cos 1x + 2\cos 2x + \dots$ is not ultra-converging to $2\pi\delta_{\text{per}}(x)$.

This means that to have Fourier transform of 2π-periodic measures, we need *another* notion of convergence. One convenient approach is based on the following fundamental property of measures; however, before we introduce it, we need to twist our language a bit.

Given a function F, we say that "F' *makes sense* as a measure" if there is a measure μ such that F is an antiderivative of this measure. Note that in this phrase, F' is not a function, but just a symbol. On the other hand, one can (and we would) identify F' with the measure μ. The property claims:

$$\boxed{\text{If } U \text{ is } 2\pi\text{-periodic and } U' \text{ makes sense as a measure, the Fourier series of } U \text{ converges at every } x.}$$

---

[630] **N.B. (???) Check!**

[631] Likewise, one can consider measures which vanish outside of the given interval.

[632] **N.B. (???) Examples where such convergence happens? Multiplication by a continuous function? (In general can multiply by an L^∞-function on $(\mathscr{M}, \mu) := \coprod_\alpha (\mathbb{R}, \mu_\alpha)$; here $\{\mu_\alpha\}$ is the maximal collection of pairwise singular non-negative non-0 measures on $\mathbb{R}$. Note that a measure on $\mathbb{R}$ can be identified with an L^1-function on (a not σ-finite) $\mathscr{M}$. Compare with the desciption of the dual space to measures with "ultra" topology.)**

[633] **N.B. (???) WRONG: too strong convergence! Need to consider a much weaker distance... Should not we discuss σ-additivity first? Infinite sums?**

In fact, one can amplify this: there is a *uniform bound* for all functions $U_k(x)$: a number $C > 0$ such that $|U_k(x)| \leq C$ for every k and every x.

(The Fourier coefficients "make sense" due to Exercise L16.)

Exercise L14: Consider a sequence $V_k(x) \to V(x)$ convergent for every x. Assume that all functions $V_k(x)$ are uniformly bounded and have antiderivatives $V_k^{(-1)}$. Then the sequence of functions $V_k^{(-1)}$ *converges uniformly*[634] provided $V_k^{(-1)}(0) = 0$. In other words:

$$\boxed{\text{Taking antiderivative improves pointwise-uniformly-bounded convergence to uniform convergence.}}$$

Hint: It is enough to consider the case when V_k monotonically increases/decreases (put $W_{K,L}(x) := \max_{0\leq k<L} V_{K+k}(x)$ and $W_K(x) := W_{K,\infty}(x)$.)

Our next aim is to proceed "one more step to the left": we want *to force* the antiderivative to $\boxed{\text{improve "convergence of measures" to pointwise-uniformly-bounded convergence}}$. With the "brute force approach", assuming that μ and μ_k, $k \geq 1$ have antiderivatives U and U_k, say that[635] $\mu_1+\mu_2+\dots$ *converges* to μ if F_k are uniformly bounded on every finite interval, and $F - (F_1 + F_2 + \ldots + F_n)$ goes to 0 at every point when n grows.[636] Likewise for convergence of sequences of measures. (Since the antiderivative may be changed by a constant, we assume that this property holds for *a suitable* choice of antiderivatives.[637])

Say that a measure is *continuous* (or *smooth*, etc.) if it has continuous (or smooth, etc.) density F. In other words, we more or less identify μ with its density.[638 639] Observe that this identification *is compatible* with our notation for derivatives and antiderivatives: if μ has a continuous density F, then $\mu^{(-1)}$ coincides with $F^{(-1)} + \text{const}$; and if G has a continuous derivative F, then G' makes sense as a measure, and the measure G' has density H.

Exercise L15: Show that δ-measure is a limit of a sequence of smooth measures. Show that $\mu = \sum_k \mu_k$ does not imply $\mu(\mathscr{I}) = \sum_k \mu_k(\mathscr{I})$.

[640]

Exercise L16: Given a function U such that U' makes sense as a measure, the integral $\int_{-\infty}^{\infty} U(x)G(x)\mathrm{d}x$ makes sense if G is continuous and vanishes outside of $[-B, B]$ for some B.

Hint: One may assume that U is non-decreasing. Then if $|G| < C$, subdividing $[-B, B]$ into intervals of length ε allows to find $\int_{-\infty}^{\infty} U(x)G(x)\mathrm{d}x$ up to precision $\varepsilon C(U(B) - U(-B)) + 2B\zeta \max(|U(-B)|, |U(B)|)$. Here ζ is the maximum of variation of G on these intervals.

Exercise L17: Given a function U such that U' makes sense as a measure μ, and a continuously differentiable function Φ such that[641] $\Phi(x) = 0$ for $x \notin [-B, B]$, define $I := \int_{-\infty}^{\infty} \Phi(x)\mu(x)$ (here $\mu(x)$ is just a symbol)

---

[634] This means that the pointwise limit $W(x)$ exists for every x, and the "speed of convergence" may be chosen independently of x: for every $\varepsilon > 0$ there is $K > 0$ "which works" for every x — or $|V_k^{(-1)}(x) - W(x)| < \varepsilon$ provided $k > K$.

[635] **N.B. (???) Why is this important?**

[636] **N.B. (???) Check convergence on intervals $\mathscr{I}$.**

[637] In fact, one may just require that $F(0) = F_k(0) = 0$ for every k.

[638] Note that this "identification" loses a bit of information about the density. For example, changing density in one point does not change the measure.

Note the similarity of this loss of information with the definiton of a vector in the vector space L^1: it is an absolutely integrable function *up to a change which does not affect the corresponding measure*. Hence L^1 *coincides* with the set of measures which have integrable densities.

[639] To preserve "dimension", sometimes μ is identified with the expression $F(x)\mathrm{d}x$.

[640] **N.B. (???) Our notion of convergence does not cover convergence of $\mu([x, x])$ — but it covers convergence if one "averages between closed and open intervals"! In other words, convergence of measures implies convergence of the weight of "blurred intervals" — on which the boundary points are considered "with multiplicity ½". Such weight of the interval between a and b is $½\mu(\{a\}) + \mu((a, b)) + ½\mu(\{b\})$. (This is hardwired into our definition of the antiderivative of measure.)**

[641] **N.B. (???) Is not $U(x)\Phi(x) \to 0$ when $x \to \pm\infty$ enough?**

as $I = -\int_{-\infty}^{\infty} \Phi'(x)U(x)\mathrm{d}x$. Show that if U' is continuous, then $I = \int_{-\infty}^{\infty} \Phi(x)U'(x)\mathrm{d}x$. (Same for U' integrable.[642])

The third fundamental property of measures is that

$$\boxed{\text{Given } \mu\text{, the integral } I := \int_{-\infty}^{\infty} \Phi(x)\mu(x) \text{ can be generalized to the case of continuous } \Phi \text{ too.}}$$

Moreover, if $\Phi(x) = 0$ for $x \notin [-B, B]$, then $|I| \leq C \max |\Phi(x)|$ for a certain C depending on B. Furthermore, if μ and Φ are non-negative, then $I \geq 0$.

Exercise L18: Consider an interval $\mathscr{I}_0$; assume that $\mu(\mathscr{I}) = 0$ if $\mathscr{I} \cap \mathscr{I}_0 = \varnothing$. Then $\int_{-\infty}^{\infty} \Psi(x)\mu(x) = C\mu(\mathscr{I}_0)$ if Ψ is continuosly differentiable and $\Psi|_{\mathscr{I}_0} \equiv C$.

Same if Ψ is only continuous.

Hint: Squeeze Ψ between two continuously differentiable functions which are constant on $\mathscr{I}_0$.

Hence for μ as in this exercise and $\mathscr{I}_0 = [a, b]$, one can define $\int_a^b \Phi(x)\mu(x)$ for a continuous function Φ on $[a, b]$.

Exercise L19: If $\Phi(x) = 0$ for $x \notin [-B, B]$, then $\left|\int_{-\infty}^{\infty} \Phi(x)\mu(x)\right| \leq |\mu|((-B, B))M$; here $M := \max_x |\Phi(x)|$.

Hint: One can assume $|\mu|((-\infty, B]) = |\mu|([B, \infty)) = 0$; then if $\mu, \Phi \geq 0$, then $\int_{-B}^{B}(M - \Phi(x))\mu(x) \geq 0$.

Exercise L20: Suppose that $\mu_k \to \mu$ and a continuously differentiable function Φ vanishes outside of $(-B, B)$. Then $\int_{-\infty}^{\infty} \Phi(x)\mu_k(x) \to \int_{-\infty}^{\infty} \Phi(x)\mu(x)$.

In fact, with practically the same proof, one can allow Φ to be a function of bounded variation. However, this does not work for general continuous Φ:

Exercise L21: Put $\Phi(x) := x \sin \pi/4x$ for $x \neq 0$ and 0 otherwise. Find measures $\mu_k \to 0$ with $\int_{-\infty}^{\infty} \Phi(x)\mu_k(x) \to +\infty$.

Hint: Consider linear combinations of measures $\nu_n := \delta(x - {}^1\!/\!_n)$.

Exercise L22: Given a 2π-periodic measure μ, one can find a number m_0 such that $\mu - m_0\mu_1$ has a 2π-periodic antiderivative F. (Here μ_1 is the measure with constant density 1. — Sometimes it is denoted as $\mu_1 = \mathrm{d}x$.)

In the latter case, say that m_0 is the 0th *Fourier coefficient* of μ, and $m_k := -\mathrm{i}F_k/k$ is the kth *Fourier coefficient*.

Exercise L23: For a 2π-periodic measure μ with Fourier coefficients m_k, consider the measure μ_K with the (smooth) density $\sum_{k=-K}^{K} m_k \mathrm{e}^{\mathrm{i}kx}$. Show that $\mu_K \to \mu$.

Exercise L24: Fourier coefficients of a 2π-periodic μ coincide with suitable Φ-weighted densities of μ.

Hint: Take Φ vanishing outside $[-2\pi, 2\pi]$, and choose a suitable function $\Phi(x - 2\pi) + \Phi(x)$ on $[0, 2\pi]$.

In fact, Φ can be chosen to be smooth.

Exercise L25: Define μ_m to be the measure with (smooth) density $\cos 0x + 2\cos 1x + 2\cos 2x + \ldots + \cos mx$. Show that $\mu_k \to 2\pi\delta_{\text{per}}$. Here δ_{per} is a 2π-periodic measure such that $\delta_{\text{per}}(\mathscr{I}) = \delta(\mathscr{I})$ if $\mathscr{I} \subset (-2\pi, 2\pi)$.

Let's sum up what we obtained. Given a continuous function Φ_1 vanishing outside of a finite interval and such that $\int_{-\infty}^{\infty} \Phi_1(x)\mathrm{d}x = 1$, it is natural to call the number $\int_{-\infty}^{\infty} \Phi_1(x)\mu(x)$ *the Φ-weighted averaged density of the measure μ*. (At least, this is compatible with the case when the density is L^1-integrable.) However, fixing one continuous function Φ_1 as above, given any continuous function Φ vanishing outside of a finite interval, a suitable expression $\Phi - C\Phi_1$ is again of the form considered above. **Conclusion:** dropping the condition $\int_{-\infty}^{\infty} \Phi_1(x)\mathrm{d}x = 1$ results in an equivalent collection of data about μ. However, we need to change the name: use "*Φ_1-weighted value*" in the general case.

So:

- the measures do not have values at points, but

[642] **N.B. (???) If U' is integrable, then the latter integral makes sense. And???**

- they have “Φ-weighted average value” for suitable Φ.
- If μ has a density F, and Φ vanishes outside a small interval containing x, the “Φ-weighted average value” approximates the value $F(x)$ at the point x.
- Moreover, convergence of measures results in convergence of Φ-weighted values — but:
- ... this only holds provided Φ is smooth enough.

Remark 116: In fact, we cover measures also in another group of exercises on p. 186.

Exercises M: Other starting points for formal derivatives

(What we discuss in this section is for people fluent with the language of measures. — If not, one may try to skim through the group L of exercises on p. 191 before reading what follows.)

In the definitions of the preceding groups of exercises, the phrase “a continuous function” played a very special role. However, we used *only* the following properties of continuous functions:

- The set C^0 of continuous functions is a vector space. (See Footnote 584 on p. 179.)
- Certain continuous functions are called “continuously differentiable functions”. They form a vector subspace C^1.
- There is linear operation $\mathrm{d}/\mathrm{d}t$ sending $C^1 \to C^0$.
- Every continuous function is in the image of $\mathrm{d}/\mathrm{d}t$. If $\mathrm{d}/\mathrm{d}t$ sends a function f (from C^1) to 0, then f is proportional to a particular function (“a constant”).

Using these, we can define the class $C^n \subset C^1$, $n \geq 1$, inductively as $(\mathrm{d}/\mathrm{d}t)^{-1}C^{n-1}$ (together with the mapping $(\mathrm{d}/\mathrm{d}t)^n : C^n \to C^0$); one can also define polynomials of degree $< n$ as functions in C^n which are sent by $(\mathrm{d}/\mathrm{d}t)^n$ to 0. (One can interpret C^0 as a vector space with an *operation of antidifferentiation* — but the result of this operation is defined only “up to an additive constant”.[643])

What is important to us is that after replacing C^0 by another class $\mathfrak{C}^0$ of “functions” the construction of formal derivatives up to undistinguishability can be translated word-by-word (together with all its corollaries) — provided $\mathfrak{C}^0$ satisfies the properties in the preceding list.[644] So one obtains the set (a vector space!) of formal derivatives of elements of $\mathfrak{C}^0$ up to undistinguishability. Denote it as $\mathfrak{C}^{\mathrm{gen}}$.

For example, put $\mathfrak{C}^0 := \{\text{smooth functions}\}$ (so then $\mathfrak{C}^1 = \mathfrak{C}^0$). — But then the result is trivial:

Exercise M1: If $\mathfrak{C}^1 = \mathfrak{C}^0$, then any element of $\mathfrak{C}^{\mathrm{gen}}$ coincides with a uniquely defined element of $\mathfrak{C}^0$.

However, if one puts $\mathfrak{C}^0$ to be absolutely continuous functions, or L^1-functions,[645] or measures,[646] then these give “meaningful theories” $\mathfrak{C}^{\mathrm{gen}}$ — but, as the following exercises shows, they do not give anything new. The results coincide with “the generalized functions as based on continuous functions”.

Exercise M2: Consider a class $\mathfrak{C}$ as above with the corresponding subspaces $\mathfrak{C}^n$ and maps $\mathrm{d}/\mathrm{d}t : \mathfrak{C}^n \to \mathfrak{C}^{n-1}$ given for $n \geq 1$. Suppose that for a suitable $N \geq 0$ we can identify $\mathfrak{C}^N$ with a subspace $V \subset C^k \subset C^0$ for a suitable $k \geq 0$. Suppose that the subspace $\mathfrak{C}^{N+1} \subset \mathfrak{C}^N$ is identified with $W \subset V$. Assume that either $k \geq 1$, or $k = 0$ and $W \subset C^1$; then there is a well-defined mapping $\mathrm{d}/\mathrm{d}t : W \to C^0$. Finally, suppose that the identification above sends the mapping $\mathrm{d}/\mathrm{d}t : \mathfrak{C}^{N+1} \to \mathfrak{C}^N$ to the mapping $\mathrm{d}/\mathrm{d}t : W \to C^0$.

Note that given $F \in \mathfrak{C}^0$, one can write $F = (\mathrm{d}/\mathrm{d}t)^N G$ with $G \in \mathfrak{C}^N$; denote by g the element of C^k corresponding to G. Show that then there is a well-defined linear map sending F to $g^{(N)}$ (considered as a generalized function) — its result does not depend on the choice of G. Likewise, if one considers a formal

---

[643] This may be formalized by fixing a 1-dimensional subspace $\langle\text{const}\rangle \subset V$ (here $V := C^0$) and an injective mapping $\alpha : V \to V/\langle\text{const}\rangle$. Then C^1 is the preimage in V of $\operatorname{Im}\alpha$.

[644] In particular, one can define subspaces $\mathfrak{C}^n$ inside $\mathfrak{C}^0$ and the (surjective) maps $\mathrm{d}/\mathrm{d}t : \mathfrak{C}^{n+1} \to \mathfrak{C}^n$.

[645] Here it is important that we consider not “L^1-integrable functions”, but such functions “up to a change on a subset of measure 0”. Otherwise there is no mapping $\mathrm{d}/\mathrm{d}t : \mathfrak{C}^1 \to \mathfrak{C}^0$ (in other words: the mapping of taking the antiderivative has a kernel).

[646] We consider “L^1-integrability and “measures” in the group L of exercises on p. 191.

derivative $F^{[M]}$ as an element of $\mathfrak{C}^{\text{gen}}$, show that sending $F^{[M]}$ to $g^{(M+N)}$ gives an operation sending $\mathfrak{C}^{\text{gen}}$ to the space of generalized functions.

Hint: Show that undistinguishable formal derivatives are sent to the same image.

Suppose now that $V \supset C^K$ for a suitable number $K \geq k$. Show that the map above is a bijection from $\mathfrak{C}^{\text{gen}}$ to generalized functions. Show that this map is compatible with taking derivatives, and sends polynomials[647] to polynomials preserving the degree.

Exercise M3: The assumptions of the preceding exercise are satisfied if $\mathfrak{C}$ consists of absolutely continuous functions, or L^1-functions, or measures. In the latter case, δ-measure is identified with the δ-function.

Conclusion: We introduced generalized functions (essentially) by saying that a continuous function *is* a generalized function, and a "derivative" of a generalized function is also a generalized function. For any kind of mathematical object for which taking repeated antiderivatives makes sense — and one of them is continuous, any object of this kind may be identified with a generalized function.

For example, measures allow antiderivatives, and the second antiderivative of any measure is continuous — hence any measure *is* a generalized function. Moreover, any continuous function is a density of a measure — hence any generalized function may be written as a (formal) repeated derivative of a measure. (This is what Exercise M3 claims!)[648]

In particular, the description above is good for functions with the second continuous derivative. However, the main body of our notes deals with "functions" $F(t)$ such that only the second *antiderivative* of F is a well-defined continuous function.[649]

One stumbling stone is that they *remain valid* in more complicated cases *too* — but then time to time one needs to mix in the word "generalized".

[650] as "usual functions".[651]

For example, if the sequence is constant ("no decay at all") then the periodic function is proportional to a δ-function: it is "essentially" generalized function! It cannot be described by "sending every numeric argument to a numeric value"; it is a more complicated object. However, the antiderivative (or "the derivative of order -1) of the δ-function takes the value 0 to the left of 0, and the value 1 to the right of 0 — so it *may* be described "by sending arguments to values". Hence (ignoring the value of the antiderivative at 0 — which is typically irrelevant) one may say that "no decay at all" is *matched* to "smoothness of degree[654] -1".

---

[647] Recall that we define polynomials of degree m as the kernel of $(\mathrm{d}/\mathrm{d}t)^{m+1}$.

[648] **N.B. (???) Can we illustrate the corresponding filtrations C^m and $\mathscr{M}^m$?**

[649] **N.B. (???) Moreover, for our example of Eisenstein series . . . (Only left- and right-continuous — and only conjecturally!).**

[650] **N.B. (???) Fix this!**

[651] Here we implicitly refer to the fact that *generalized* functions have all the derivatives — when these derivatives are taken "in the sense of generalized functions". (Same for *anti*derivatives.)

However, any generalized function is either continuous (so we may ask: how many derivatives make sense as usual functions?), or it may be described as nth derivative of a "usual" function (e.g., of a continuous function).[652] (Then we can say that its degree of smoothness is $-n$.)

The finer points of the "degree of smoothness" are whether the "topmost" of these "usual" derivatives is continuous (or left-/right-continuous). Compare with the plot on p. 62 — it is already a plot of an antiderivative, but one needs to take one more antiderivative to make it continuous.

[652] For example, δ-function is a derivative of a "usual"[653] step-function, and is the second derivative (in the sense of generalized functions) of a continuous function $\frac{1}{2}|x|$. So its "degree of smoothness" is -1 — in the sense that its -1st derivative is left-and-right-continuous.

[653] What is a "usual" function? There are many different sub-flavors. The most useful are "continuous", or "left-and-right-continuous", or L^1, or L^2, or "measures".

[654] . . . with the flavor of -1st derivative "being left-and-right-continuous".

Similarly, if (a_n) decays at least as quick as $\operatorname{const}/|n|^2$, then its Fourier transform is continuous.[655]

It is also useful to understand that the rate of decay of Fourier coefficients corresponds to the smoothness of the sum of Fourier series. (In particular, taking derivative — which makes a C^k-function "less smooth" — corresponds to multiplication by n — which makes the Fourier coefficients decay slower. Likewise for integration: it makes a function "smoother", and makes the Fourier coefficients to decay quicker.)

Finally, it may help to know some particular cases of the preceding connection. In particular, if coefficents are in ℓ_1 (hence "do not decay too slow") then the sum of the series is a continuous function. (In the opposite direction one can get only a much weaker estimate: continuity implies that the coefficients are bounded.[656])

Solutions to Fourier "Meander wave" exercises

The main tool are the estimates of $F(y)$ provided $F'(y) = G(y)\sin y$ with $G \geq 0$ a decreasing function of y (the case of an increasing $G > 0$ can be handled very similarly). (Here y is a linear coordinate change of the variable x from the exercises.) Note that the local maximums N_{2k-1} of F are achieved at the points $y_{2k-1} = (2k-1)\pi$, and local minimums n_{2k} of F are achieved at the points $y_{2k} = 2k\pi$. Moreover,

$$N_{2k-1} - N_{2k+1} = -\int_{y_{2k-1}}^{y_{2k+1}} G(y)\sin y \,\mathrm{d}y = \int_{y_{2k}}^{y_{2k+1}} (G(y-\pi) - G(y))\sin y \,\mathrm{d}y \geq 0,$$

hence local maxima decrease. Likewise, local minima increase. Furthermore,

$$N_{2k-1} - n_{2k} = -\int_{y_{2k-1}}^{y_{2k}} G(y)\sin y \,\mathrm{d}y \leq 2 \max_{[y_{2k-1},y_{2k}]} G(y) = 2G(y_{2k-1}).$$

Similarly, $N_{2k+1} - n_{2k} \leq 2G(y_{2k})$.

Moreover, for $k \leq l$

$$\begin{aligned} &N_{2k-1} - N_{2l+1} \\ &= \int_{y_{2k}}^{y_{2k+1}} ((G(y-\pi) - G(y)) + (G(y+\pi) - G(y+2\pi)) + \ldots + (G(y+(l-k)\pi-\pi) - G(y+(l-k)\pi))) \sin y \,\mathrm{d}y \\ &\qquad \leq \int_{y_{2k}}^{y_{2k+1}} ((G(y-\pi) - G(y+(l-k)\pi))) \sin y \,\mathrm{d}y \leq 2(G(y_{2k-1}) - G(y_{2l+1})). \end{aligned}$$

Therefore $F(y)$ with y between y_{2k-1} and y_{2l+1}

- oscillates, with
- every increasing/decreasing half-wave "crossing" the all the levels in the interval $[B, C] := [n_{2l}, N_{2l+1}]$.
- The interval $[B, C]$ has length at most $2G(y_{2l})$.
- The maximum of $F(y)$ between y_{2k-1} and y_{2l+1} is at most $C + 2(G(y_{2k-1}) - G(y_{2l+1}))$.
- The minimum of $F(y)$ between y_{2k-1} and y_{2l+1} is at least $B - 2(G(y_{2k}) - G(y_{2l}))$.

As the result, we can squeeze the values of $F(y)$ on $[y_{2k-1}, y_{2l+1}]$ into an interval of length at most[657]

$$2(G(y_{2l}) + G(y_{2k-1}) - G(y_{2l+1}) + G(y_{2k}) - G(y_{2l})) = 2(G(y_{2k-1}) - G(y_{2l+1}) + G(y_{2k})) \leq 4G(y_{2k-1}).$$

(If $l \gg k$ and G decreases quickly enough, then two estimates on the right are close to each other.)

One can get similar estimates if one replaces one (or both) y_{2k-1} and y_{2l+1} by y_{2k} and y_{2l}.

---

[655] For example, if $a_0 = 0$ and $a_n := 1/n^2$ for $\mathbb{Z} \ni n \neq 0$ then its Fourier transform (the sum of the corresponding Fourier series $\sum_{n>0} 2\cos(nx)/n^2$) is a 2π-periodic "parabolic sawtooth" function; it is $(|x|-\pi)^2/2 - \pi^2/6$ on $[-2\pi, 2\pi]$.

[656] Recall that it is possible to get a 1-to-1 match between "degrees of the growth of coefficients" and "degrees of smoothness of the sum" — but one needs a bit more complicated gauges of these degrees, such as *the Sobolev classes*

[657] Since we do not know the integration constant, we cannot estimate the ends of this interval, only its length!

In the "meander wave" exercises, a partial Fourier sum $M_{\le 2r}$ is an antiderivative of $2(\cos x + \cos 3x + \ldots + \cos(2r-1)x) = \sin(2rx)/\sin x$. So put $y = 2rx$, and $G(y) := 1/(2r\sin(y/2r))$. Note that if $y \le \pi r$, then $\sin(y/2r) \ge y/(\pi r)$, hence $G(y) \approx 1/y$ for small y, while $G(y) \le \pi/(2y)$. The analysis above is applicable on $(0, \pi r]$ where G decreases.

This shows that $M_{\le 2r}(x)$ takes values inside an interval of length $\le \pi/(rX)$ for x on $[X, \pi/2]$ if $2rX \in \pi\mathbb{Z}$. (Indeed, the right end of the interval of $y_{\lrcorner}$s is $r\pi$, so has the required form of y_r.) **Conclusion:** if $X > 0$, then $M_{\le m}(x) - M_{\le m}(\pi/2)$ is uniformly going to 0 on $[X, \pi/2]$ as $m \to \infty$. Moreover, one can replace $[X, \pi/2]$ by $[X_m, \pi/2]$ provided $mX_m \to \infty$.

Likewise, for $F(Y) := \operatorname{Si}(Y) = \int_0^Y \sin y \mathrm{d}y/y$ with $Y \ge 0$ we can use $G(y) = 1/y$. The estimates above show that F has a certain limit as $Y \to \infty$; denote it by $\Pi/2$. Moreover, although G is not smooth at 0, still the integral makes perfect sense — and hence putting $y_0 = 0$ does not break the fact that the sequence $F(y_{2m})$ increases. Therefore $F(Y) > 0$ for $y > 0$, and the limit of F is positive.

For Exercise K22 on p. 189, put $g := \operatorname{Si}$. Then put $\Delta_m(y) := M_{\le m}(y/m) - g(y)$, hence $m\Delta'_m(y) = \widetilde{G}(y/m)\sin y$ with $\widetilde{G}(y) := 1/\sin(y) - 1/y \ge 0$. Since $\widetilde{G}$ increases[658]on $[0, \pi/2]$, we can repeat the analysis above with y changed to $-y$. Hence one can estimate $m\Delta_m(y)$ on $[0, Y]$ as being confined to an interval of length $\le 4\max_{[0,Y/m]} \widetilde{G}/m$ (provided $y \in \pi\mathbb{Z}$, $0 < Y/m < \pi/2$ and m is even).

On the other hand, this interval contains $\Delta_m(0) = 0$; hence we can conclude that $|\Delta_m(Y)| \le cY/m^2$ for a suitable constant c. Therefore $|M_{\le m}(x) - \operatorname{Si}(mx)| \le cx/m$. Hence if mx_m goes to ∞, then $M_{\le m}(x_m) \approx \operatorname{Si}(mx_m) \to \Pi/2$.

Conclusion: $M_{\le m}$ is close to $\Pi/2$ on the interval $[X_m.\pi - X_m]$ if $mX_m \to \infty$. In other words, $\pi M_{\le m} \to \Pi M$ as $m \to \infty$.

Next, $W_{\le m}(x) + x$ is an antiderivative of $1 + 2(\cos x + \cos 2x + \cos 3x + \ldots + \cos mx) = \sin((m + ½)x)/\sin(½x)$. So we can put $y := (m + ½)x$ and $G(y) := 1/((m + ½)\sin(y/(2m+1)))$ which is positive and decreases on $(0, (m + ½)\pi)$. (Since the right end is not in $\pi\mathbb{Z}$, one needs a minor modification to the arguments above. We are going to ignore this complication.)

Since G is of very similar form to what we considered above (only twice larger), we can show that $W_{\le m}(x) + x$ is odd, and is approximately constant on any interval of the form $(X_m, 2\pi - X_m)$ with $X_m \to 0$ and $mX_m \to \infty$. Moreover, this shows that[659] $W_{\le m}(x) + x \approx 2\operatorname{Si}((m + ½)x)$ on $(-2\pi + X_m, 2\pi - X_m)$.

Finally, since $W_{\le m}(x)$ is odd 2π-periodic, $W_{\le m}(\pi) = 0$. Therefore $W_{\le m}(x) + x \approx W_{\le m}(\pi) + \pi = \pi$. **Conclusion:** $\operatorname{Si}(x) \approx \pi/2$ for $x \gg 0$; hence $\Pi = \pi$. This concludes the proof that the inverse Fourier transform is actually inverse to the Fourier transform on the examples of rectangular and triangular waves M and W.

For the Hausdorff distance approximation: consider the broken line connecting the local maxima and a similar line for minima. For $x \le \pi/2$ above the first local minimum both graphs make sense. The description of oscillation of the function F above shows that any point in the region between these graphs may be shifted horizontally by at most "the period of oscillation" to get on the graph of F. Moreover, a point with x below the first local minimum can be shifted to the y-axis by a similar shift.

Therefore the graph of F may be approximated by the region described above together with the projection of the graph to the y-axis. The latter is described by the absolute maximum of F. Finally,

---

[658] In fact, all derivatives of $\widetilde{G}$ are positive (and increase) on $[0, \pi)$. Indeed, note that $-1/t$ has all derivatives positive (and growing) for $t < 0$; hence the same holds for $\varphi_a(t) := 1/(t-a) - 1/t$ with $a > 0$. Therefore all derivatives of $\varphi_a(t) - \varphi_a(-b-t)$ are positive on $(-b/2, 0)$ if $b > 0$. (Indeed, the odd-order derivatives of two terms have the same sign; for even-ordered, it is an increment of the corresponding repeated derivative between $-b - t < t$.)

Now use the fact that the Mittag-Leffler's decomposition of $1/\sin x - 1/x$ is a convergent sum of terms $\varphi_\pi(x - (2k-1)\pi) - \varphi_\pi(-x - (2k-1)\pi)$, and put $t := x - (2k-1)\pi$, $b := (4k-2)\pi$.

[659] In fact, the same argument as above shows that $|W_{\le m}(t) + t - 2\operatorname{Si}((m + ½)t)| \le ct/(m + ½)$ for $0 \le t \le \pi$.

the curve connecting the maxima may be approximated by a hyperbola — as required. Likewise for minima.

i. ii. iii. iiii. iiiii. iiiiii. iiiiiii. iiiiiiii. iiiiiiiii. iiiiiiiiii. iiiiiiiiiii. iiiiiiiiiiii. iiiiiiiiiiiii. iiiiiiiiiiiiii. iiiiiiiiiiiiiii. iiiiiiiiiiiiiiii. iiiiiiiiiiiiiiiii.

i. ii. iii. iiii. iiiii. iiiiii. iiiiiii. iiiiiiii. iiiiiiiii. iiiiiiiiii. iiiiiiiiiii. iiiiiiiiiiii. iiiiiiiiiiiii. iiiiiiiiiiiiii. iiiiiiiiiiiiiii. iiiiiiiiiiiiiiii. iiiiiiiiiiiiiiiii.

Exercises N: Cesàro-like summation

For the following argument, we need to enhance arguments in the beginning of the section on p. 199 to handle estimates of $F(y+2\pi)-F(y)=\int_y^{t+2\pi}G(y)\sin y\,dy$. Since this may be written as $\int_y^{t+\pi}(G(y)-G(y+\pi))\sin y\,dy$, its magnitude may be bounded[660] by $2\pi\max_{Y\in[y,y+\pi]}|G'(Y)|$.

Heuristically speaking, this means that for $|F(y+T)-F(y)|$ to be small (for all y, assuming T fixed), T must be close to a multiple of 2π — provided $|G'|$ is much smaller than $|G(y)|$. And the latter condition matches the choice $G(x) := g(x/m)$ with $m\to\infty$ in our applications of estimates of F. Moreover, in our cases the oscillations of F are *parasitic* — they are due to "us cutting the Fourier series early" — so, generally speaking, our preference is to lower the effect of these oscillation (when there is such a possibility).

In other words, we consider $F(y) = (G(y)\sin y)^{(-1)}$ as "having two parts": the "parasitic" oscillating part (which is essentially determined by the behaviour of $G(x)$ for x "not very far" from y), and the slowly changing "*secular*" part F_{sec} made by contribution of remaining x␣s. Considering $F(x+2\pi)-F(x)$ has a chance to suppress the "parasitic" oscillating part — which can make the "secular" part more visible.

Conclusion: Define the *approximate derivative* $\Delta_T F(y) := (F(y+T/2)-F(y-T/2))/T$ of F; then for contribution of parasitic oscillations into $\Delta_T F$ to be small, one should choose T to be close to a multiple of 2π. Moreover, when acting on the "secular" part, the approximate derivative has a "discretization error" which goes down as T becomes smaller; so the best candidates seem to be $T\approx 0$ or $T\approx 2\pi$. In the former case, it turns out that T in denominator of $\Delta_T F$ kills all the advantages of "cancellation of oscillations". Hence "the best choice" of the approximate derivative to estimate the derivative of the "secular" part of F is

$$F'_{\text{sec}} \approx \Delta_{2\pi}F.$$

Recall the conclusion of Exercise G2 on p. 178: to calculate $F(t)$ given the Fourier coefficients F_k of F, one cannot just use partial sums of *this* Fourier series (they may diverge!). One should instead proceed in steps:

- Consider *another* Fourier series with coefficients $-F_k/k^2$ (for $k\neq 0$).
- Calculate its sum (converging in a very strong sense!) $S(\tau)$ at points $\tau\approx t$.
- Take the second derivative of this sum $S(\tau)$ at $\tau = t$.

Exercise N1: Show that exchanging the last two steps (in the sense of taking the second derivative of a series term-wise) brings one back to the Fourier series of F (so it may diverge).

If F is continuous, then S has two continous derivatives, hence the second derivative may be approximated by the *approximate second derivative* $(S(t+\Delta t)-2S(t)+S(t-\Delta t))/(\Delta t)^2$ when $\Delta t\to 0$. **Conclusion:** $F(t)$ may be approximated by $(S_{\le m}(t+\Delta t)-2S_{\le m}(t)+S_{\le m}(t-\Delta t))/(\Delta t)^2$ if one

- First takes the limit $m\to\infty$.
- Second, lets $\Delta t\to 0$.

Recall that $S_{\le m}(t)$ is the partial sum of the Fourier series for S (with the coefficients $-F_k/k^2$).

[660] **N.B. (???) Find local extrema?**

Exercise N2: Show that for a continuous F with a 2π-periodic second antiderivative S one can approximate $F(t) \approx (S_{\leq m}(t+\Delta t) - 2S_{\leq m}(t) + S_{\leq m}(t-\Delta t))/(\Delta t)^2$ with arbitrary precision by choosing Δt sufficiently small, and m depending on Δt.

Hint: For a fixed Δt, it is enough to approximate S by $S_{\leq m}$ only in C^0 sense.

This shows that to approximate F, one can take $\Delta t \to 0$ and choose m as $m \geq \mathbf{m}(\Delta t)$ in the formula of the exercise. Here the choice of $\mathbf{m}$ may depend on F; moreover (for certain functions F) the function $\mathbf{m}$ may grow very quick when $\Delta t \to 0$. Below we investigate *another approach*: it turns out that by fine-tuning, m may be choosen relatively small, and be independent of F. However, instead of an infinite interval $[\mathbf{m}, \infty)$ of possible m, every Δt matches a quite narrow zone of "fine-tuned" ms.

As we saw in our investigation of the meander and triangular waves in the group K of exercises their partial Fourier sums are contaminated by "parasitic oscillations". The following exercise shows that these oscillations may be quite large.

Exercise N3: Define the periodic measure ω_m as the derivative of $W_{\leq m}(t) + t$ with W the triangular wave from the section on p. 187. Recall that we expect ω_m to "approximate" the measure $(W(t)+t)' = 2\pi\delta_{\text{per}}$ (sum of δ-measures at points of $2\pi\mathbb{Z}$). The latter measure vanishes on the interval $[\varepsilon, 2\pi-\varepsilon]$ (with $0 < \varepsilon < \pi$); so denote by $\omega_{m,\varepsilon}$ the restriction of ω_m to such an interval.

Show that the norm $\|\omega_{m,\varepsilon}\|$ does not go to 0 when $m \to \infty$.

Hint: Estimate the average value (or "amplitude") of $|W'_{\leq m} + 1|$.

Show that the norm $\|\omega_m\|$ is not bounded when $m \to \infty$.

Hint: Harmonic series.

Compare this with Exercise L13 on p. 194 which shows another breakage of a somewhat similar property. In fact, our next aim is to show that the conditions of Exercise N3 are in fact much more important than the property from Exercise L13 on p. 194. But before we do this, we want to highlight a situation where the properties of the last exercise are *actually satisfied*.

To reach this situation, we "mollify"[661] the oscillations above by a suitable choice of the step of the approximate derivative.[662]

Exercise N4: Define the periodic measure Ω_m as the approximate derivative $\Delta_{2\pi/m}$ of $W_{\leq m}(t) + t$ with W as above. Recall that we expect Ω_m to be "close" to ω_m, so the measures Ω_m have a chance to "approximate" the measure $2\pi\delta_{\text{per}}$. Again, denote by $\Omega_{m,\varepsilon}$ the restriction of Ω_m to the interval $[\varepsilon, 2\pi-\varepsilon]$ (with $0 < \varepsilon < \pi$).

Show that the norm $\|\Omega_{m,\varepsilon}\| \to 0$ when $m \to \infty$ (call this "*condition (A)*"). Show that the norm $\|\Omega_m\|$ is bounded when $m \to \infty$ (call this "*condition (B)*"). Show that the average value of Ω_m is 2π.

In other words, taking "the best choice" of the step of the approximate derivative (same as in the beginning of this section), we replace "the pathological measures" ω_m by the measures Ω_m which are "warm and cozy" in the sense of the exercises above. Our next step is to explain why being "warm and cozy" is so useful.

Say that a sequence μ_k of 2π-periodic measures *specially converges* to $c\delta_{\text{per}}$ if $\mu_k([0, 2\pi)) \to c$, and $\|\mu_k|_{[\varepsilon.2\pi-\varepsilon]}\| \to 0$ for any $0 < \varepsilon < \pi$, as well as there is C such that $\|\mu_k\| < C$ for any k. (We are going to use this for $c = 0, 1, 2\pi$.)

Recall that if f and g are 2π-periodic, *the convolution* $f \star g$ of f and g is defined as $(f \star g)(t) := \int_0^{2\pi} f(t-s)g(s)\mathrm{d}s$. Note that if f and g are continuous, then $(f \star g)(t)$ is a well-defined number. Moreover, if g is a measure and f is continuous, then putting $f_t(s) := f(t-s)$, the integral may be understood as the value of the measure $f_t g$ on the interval $[0, 2\pi)$. Since the change of variable

[661] **N.B. (???) Use this word before?**

[662] This corresponds to making Δt depend on m in Exercise N2 on p. 202. However, instead of the second approximate derivative (of the second antiderivative) in that exercise, below we use the first derivative (of the first antiderivative). (Although a very similar argument works for the literal situation of the latter exercise.)

Keep in mind that the step of approximate derivative is $1/m$th of what we discuss above since y is t rescaled m times.

$s_1 = t - s$ shows that the integral is actually symmetric in f and g, one can proceed similarly in the case when f is a measure, and g is continuous.[663]

Exercise N5: If g is a trigonometric polynomial $\sum_{k=-K}^{K} a_k e^{ikt}$, then[664] $(F \star g)(t) = 2\pi \sum_{k=-K}^{K} a_k F_k e^{ikt}$. Here F_k are Fourier coefficients of a continuous function F.

Hint: It is enough to consider $g(t) := e^{ikt}$.

Hence $F_{\leq m}(t) = F \star (W'_{\leq m}(t) + 1)/2\pi$.

Remark 117: This explains the attention we pay to the properties of $W_{\leq m}$: the convolution with the derivative of this (smooth!) function describes the process of taking the $\underline{m}$th partial Fourier sum of any 2π-periodic function F. Hence using the formula for convolution, the information about $W_{\leq m}$ may be translated to the information about the properties of partial Fourier summation.

In other words: although W is not continuous, the details of convergence of $W_{\leq m}$ to W reflect *the worst case* of convergence of partial Fourier sums for continuous functions.

Exercise N6: Consider a sequence of trigonometric polynomials (g_m) having Fourier coefficients $g_{m,k}$. Suppose that this sequence specially converges to $2\pi\delta_{\text{per}}$ and $g_{m,k}$ vanish for $|k| > K_m$ for a certain sequence (K_m). Then given a continuous function F with Fourier coefficients F_k, the sequence of trigonometric polynomials $\Phi_m(t) := \sum_{k=-K_m}^{K_m} g_{m,k} F_k e^{ikt}$ converges to $F(t)$ at every point t.

Hint: It is enough to assume that $t = 0$ and $F(0) = 0$, $|F(s)| < \delta$ if $|s| < \varepsilon$.

Show that this convergence is actually uniform in t.

Hint: Use uniform continuity to find suitable δ and ε.

Exercise N7: Show that failure of Exercise N3 is directly related to partial sums of Fourier series of continuous functions F not converging to F.

Hint: Remark 117. (We do not require a complete proof here. Handwaving arguments are enough.)

Given a sequence of *weights* $(\widetilde{w}_k)$ which vanishes for $|k| \gg 0$, call the (finite!) sum $\sum_{k=-K_N}^{K_N} \widetilde{w}_k F_k e^{ikt}$ the $\widetilde{w}$*-weighted partial sum* of the infinite sum $\sum_{k=-\infty}^{\infty} F_k e^{ikt}$. Likewise, if $w(r)$ is a function on[665] $[0,1]$, a w*-weighted partial sum* is $\sum_{k=-K}^{K} w(|k|/K) F_k e^{ikt}$. In particular, the "usual" partial sum is 1-weighted (it corresponds to $w \equiv 1$).

Exercise N6 gives a certain recipe for checking that a sequence of weighted partial sums of Fourier series converges to the function F in question. Moreover, by Exercise N4 on p. 202 this recipe is not void. This allows to apply the conditions (A) and (B) of Exercise N4 on p. 202 to a given weight function w.

Exercise N8: Show that the weighted partial sums of Exercise N6 with $g_m = \Omega_m$ (with the sequence Ω_m from Exercise N4 on p. 202) match the function $w(r) = (e^{i\pi r} - e^{-i\pi r})/(2\pi i r) = \sin(\pi r)/\pi r$.

Hint: One should treat $k = 0$ specially.

Show that the formula of Exercise N2 on p. 202 matches

$$w(r) = (e^{ik\Delta t} + e^{-ik\Delta t} - 2)/(-k\Delta t)^2 = 2(1 - \cos(\alpha r))/(\alpha r)^2 \qquad \text{with } \alpha := m\Delta t.$$

Recall that we know that the former function w leads to convergence of w-weighted partial Fourier sums — but we do not know the conditions on $m\Delta t$ which make the latter w satisfy this. The following exercise should clarify what to expect:

---

[663] In fact, when both f and g are measures, one can define $f \star g$ as a measure. (Although this uses quite delicate properties of measures.)

[664] **N.B. (???) Check 2π.**

[665] One can allow w to be defined on $[-1,1]$ by replacing $|k|$ by k in the following formula (then it describes the case of even w). Likewise, one can replace the upper limit 1 by u if one replaces K in the limits of summation by uK.

Finally, one may extend w by 0 outside of its interval of definition; then one can change K in the limits of summation to ∞.

Exercise N9: Define the periodic measure $\Omega_{m,\beta}$ as the approximate derivative $\Delta_{2\pi\beta/m}$ of $W_{\leq m}(t)+t$ with W as above; here $\beta>0$. Show that the average value of $\Omega_{m,\beta}$ is 2π. Show that the corresponding function w (see the preceding exercise) is $\sin(\beta\pi r)/\beta\pi r$.

Show that the norm $\|\Omega_{m,\beta,\varepsilon}\|\to 0$ when $m\to\infty$ iff $\beta\in\mathbb{Z}_{>0}$. (Here $\Omega_{m,\beta,\varepsilon}$ is the restriction of $\Omega_{m,\beta}$ to $[\varepsilon, 2\pi-\varepsilon]$.) Show that the norm $\|\Omega_{m,\beta}\|$ is bounded when $m\to\infty$ iff $\beta\in\mathbb{Z}_{>0}$.

This brings the question: which functions w (extended to $\mathbb{R}$ by 0 as in Footnote 665 on p. 203) satisfy the conditions of the preceding exercise? So far, we can only answer this for w in the family $\sin(\beta\pi r)/\beta\pi r$. One obvious condition is:

Exercise N10: One should have $w(0)=1$.

Exercise N11: Consider a trigonometric polynomial $g(t) := \sum_{k=-K}^{K} g_k e^{ikt}$. Suppose that g is real and $[-K,K]$ can be broken into I subintervals of monotonicity of g_k. Then $|g(t)|\leq L(t)$ with $L(t) := \max_k |g_k|\cdot \min(2K+1, (I+2)/2|\sin{}^{t}\!/\!_{2}|)$.

If the sequence g_k has J intervals of convexity/concavity and[666] $g_K=g_{-K}=0$, then one can replace $L(t)$ by $\min(L(t), \max_{-K\leq k<K}|g_k-g_{k+1}|\cdot J/|\sin{}^{t}\!/\!_{2}|^2)$.

The following exercise "explains" the conditions of Exercise N9 on p. 204.

Exercise N12: If the interval $[-u,u]$ of definition of a weight function w can be split into a finite number of subintervals where w is convex or concave, and w is continuous when extended to $\mathbb{R}$ by 0, then w satisfies the condition (A) of Exercise N4 on p. 202.

If in addition there are $\alpha, C>0$ such that $|w(r_1)-w(r_2)|\leq C\cdot|r_1-r_2|^\alpha$ on $[-u,u]$, then w satisfies the condition (B) as well. (This covers functions having a continuous derivative away from a finite collection of points R_l near which $w-w(R_l)$ has a power-law asymptotic.)

As a corollary, one can use any such w as a way to find good approximations of a continuous function F by its w-weighted partial Fourier sums. This covers not only Exercise N9 on p. 204, but also shows that one should use $m\Delta t\in 2\pi\mathbb{Z}_{>0}$ in Exercise N2 on p. 202.

In fact, there are two approaches to understand the significance of weighted partial Fourier sums. In both approaches, one considers "putting *extra* coefficients" $\widetilde{w}_k$ into a linear combination $\sum\widetilde{w}_k F_k e^{ikt}$ of sin-or-cos-waves as a "frequency-dependent[667] filtering of the signal" $\sum F_k e^{ikt}$. (So the "components $F_k e^{ikt}$ of different frequencies k" *are attenuated* at different rates.)

In the first approach, one considers "the usual" partial Fourier sums $F_{\leq m}$ as the primary objects. Only may say that "if they do not converge to F, then this is due to 'parasitic oscillations' in $F_{\leq m}$". So one can consider a w-weighted partial sum as "an *honest* partial sum which is filtered to dump the parasitic oscillations". Then one says that

After a suitable filtering, the partial Fourier sums give excellent approximations for F.

In the other approach, one considers "the usual" partial Fourier sums $F_{\leq m}$ as "just the w_{step}-weighted partial Fourier sums" with $w_{\text{step}}(r)$ being the step-function (which is 1 or 0 depending on $|r|\leq 1$). This makes "the usual" sums just *one of a collections of equally legitimate approaches*. In other words: here the primary object is *the infinite Fourier series as a whole* which is considered to be "undistinghuisable from F itself".[668] Taking "the usual" partial Fourier sum $F_{\leq m}$ is considered as "this infinite series filtered with 'a very abrupt' filter" — the filter which passes through the frequencies up to m unchanged, and kills the high frequencies completely. Then one can say:

A filter with an "abrupt" spectral response w "produces *ringing*".

---

[666] Of course, this condition can be achieved by replacing K by $K+1$ — but note that this increases the range of k in the max below.

[667] To be honest, this covers a case of an even function w, for which the coefficients do not depend on "the difference of phase" of sin and cos.

[668] Compare with Exercise I2 on p. 181.

(In other words, an interval of t where F is non-smooth has "non-local effects" after filtering: changing F on this interval can change the result of filtering significantly even far away from this interval.[669])

With this approach, it is the "w-weighted partial Fourier sums with a 'good' w" (or, more generally, a 'good' $\widetilde{w}$) which are the cornerstones of approximating "the infinite Fourier sum". Taking a limit $w_{[k]} \to w_{\text{step}}$ (with "good" $w_{[k]}$) allows calculation of "the usual" partial sums in terms of "the good weighted sums" — but one should accept that such an extra limit may break convergence to F.

One can call such "good" weighted partial Fourier sums *the Cesàro-like approximations*. Among possible functions w, the choice $w(r) := (1-|r|)^N$ is called "*the Nth order Cesàro sum*" or "*the Hölder sum*" (with $N = 1$ giving "the Cesàro sum" or "*the Fejér sum*").

As with the "usual" partial Fourier sums (see Exercise I2 on p. 181), one has:

Exercise N13: If $w(0) = 1$ and w is continuous at 0 then the w-weighted partial Fourier sums of any generalized function F converge to F in the sense of generalized functions.

Hint: Since such a partial sum of F' is the derivative of the sum for F, it is enough to consider $F \in C^2$.

However, the "better" is a filter w, the stronger is the sense in which "ringing disappears". The following exercise gives a quantitative description (although in most contexts, the qualitative description is enough):

Exercise N14: Suppose that the extension of w to $\mathbb{R}$ is of class C^s with sth derivative having a finite number of intervals of monotonicity and $0 < \varepsilon < \pi$. Then the mth w-weighted partial Fourier sums $W_{[w,\leq m]}$ of the triangular wave W go to W uniformly on $[\varepsilon, 2\pi - \varepsilon]$ together with their derivatives up to the order s. The error in the derivative of order $K := s + 1 - l$ with $l \geq 0$ can be estimated as $|(W - W_{[w,\leq m]})^{(K)}| \leq \text{const}/m^l$.

Hint: The derivatives of $w(r)$-weighted partial Fourier sums match $r^k w(r)$-weighted partial Fourier sums. Multiplication by r^k does not *break much* the condition on intervals of monotinicity since one can split everything into a sum of mononotic terms. Compare with Exercise N11 on p. 204.

Exercise N15: Suppose that F is a 2π-periodic generalized function of class C^{-k} and the extension of w to $\mathbb{R}$ is of class C^{k+K+2}. If $a < b < c < d$ and F vanishes on $[a, d]$, then the w-weighted partial Fourier sums of F go to 0 uniformly on $[b, c]$ together with their derivatives up to order K.

Hint: It is enough to consider $k = 0$. To estimate Kth derivative of a partial sum, one convolves with $K+1$st derivative of a partial sum for W. Note that $w^{(k+K+1)} \in C^1$ hence may be represented as sum of two continuous functions with two intervals of monotonicity.

Likewise if (instead of vanishing) F has K continuous derivatives on (a, d). One can also allow F to be a kth (formal) derivative of a piecewise-continuous function with simple jumps (away from (a, d)), or even of a measure.[670] Similarly, the hint shows that as an extra generalization one may only require that $w^{(k+K+2)}$ is a measure and $w^{(k+K+1)}$ is continuous.

---

[669] This is somewhat similar to "resonances", when a momentary disturbance of a system leads to long-persisting oscillations.

[670] This boils down to the statement that given a measure μ vanishing outside of $[-L, L]$ the integral $\int_{-\infty}^{\infty} G(t)e^{imt}\mu(t)$ can be estimated *uniformly in* m given $\max_{[-L,L]} |G(t)|$. However, a similar property holds if instead of μ we use a generalized function which is smooth outside of 0, vanishes outside of $(-L, L)$, and equals $1/t$ near 0 (however, one needs to also know $\max_{[-L,L]} |G'(t)|$).

Here the generalized function $1/t$ is defined in Exercise J19 on p. 186 as $(\log|t|)' = (t\log|t| - t)''$. (This makes sense if one understands log as an L^1-function, or understands $t\log|t|$ as a continuous function.)

Indeed, assume that $H(t)$ vanishes outside of $(-L, L)$. Then $\int H(t)dt/t$ vanishes if H is even, hence one can consider instead $\int_0^{\infty}(H(t) - H(-t))dt/t$. This allows reducing to the case $G \equiv 1$, which in turn is reduced to estimating $\int_0^K \sin mt\,dt/t = \int_0^{mK} \sin t\,dt/t$. However, by calculations in the section on p. 199 this function of mK has a limit at ∞, hence is bounded.

Therefore we may also require that F is kth derivative of a linear combination of a measure and functions $1/(t - T_k)$ (or of boundary traces of functions $1/(z - T_k)$).

The most important application is the case $K = 0$: to get a convergence of the Fourier series at a suitable point, it is better to have w of class C^{k+2}. This is a very important particular case of Abel summation.[671]

Exercise N16: The w-weighted partial sums of $1 - 2 + 3 - 4 + 5 - \dots$ converge if $w \in C^3(\mathbb{R}_{\geq 0})$, and the result depends only on $w(0)$. (Here one defines the w-weighted partial sums the same way as we defined the w-weighted partial Fourier sums.)
Hint: $1 \cdot \cos t + 2 \cdot \cos 2t + 3 \cdot \cos 3t + 4 \cdot \cos 4t + 5 \cdot \cos 5t + \dots$ is related to W'' which is a formal derivative of a measure which is smooth on $(0, 2\pi)$. Hence one can calculate its density at $t = \pi$ as above, with $k = 1$ and $K = 0$.

Likewise, to regularise the sum $1^v - 2^v + 3^v - 4^v + 5^v - \dots$ one should use[672] $w \in C^{v+2}(\mathbb{R})$ with $w(0) = 1$.

As above, we can weaken the condition on w as $w^{(v+2)}$ being a measure and $w^{(v+1)}$ being continuous. Moreover, replacing $1 \cdot \cos t + 2 \cdot \cos 2t + 3 \cdot \cos 3t + 4 \cdot \cos 4t + 5 \cdot \cos 5t + \dots$ by $1 \cdot \mathrm{e}^{\mathrm{i}t} + 2 \cdot \mathrm{e}^{2\mathrm{i}t} + 3 \cdot \mathrm{e}^{3\mathrm{i}t} + 4 \cdot \mathrm{e}^{4\mathrm{i}t} + 5 \cdot \mathrm{e}^{5\mathrm{i}t} + \dots$ make only the arguments $r \geq 0$ of $w(r)$ matter.[673] Hence it is enough to require that the restriction of w to $\mathbb{R}_{\geq 0}$ are "sufficiently smooth".[674]

The importance of this is due to the identity

$$(1 - 2 \cdot 2^v)(1^v + 2^v + 3^v + 4^v + 5^v + \dots) = 1^v - 2^v + 3^v - 4^v + 5^v - \dots$$

(valid when both sides make sense). This means that one may *define* $1^v + 2^v + 3^v + 4^v + 5^v + \dots$ as $(1^v - 2^v + 3^v - 4^v + 5^v - \dots)/(1 - 2^{1+r})$, and *define* $1^v - 2^v + 3^v - 4^v + 5^v - \dots$ as the common limit of "good" w-weighted sums. This allows one to define this sum with the only (pole) singularity at $v = -1$.

Exercise N17: Show that in the case $v = 0$ the Cesàro summation gives $1 - 1 + 1 - 1 + 1 - \dots = ½$. Likewise, for $v = 1$ one needs the 2nd order Cesàro summation leading to $1 - 2 + 3 - 4 + 5 - \dots = ¼$.

Using the definition above, this leads to[675] $\beta_0 := 1+1+1+\dots = -½$ and $\beta_1 := 1+2+3+4+5+\dots = -1/12$.

Hint: Construct a 2π-periodic generalized function F with a suitable Fourier series such that F is smooth on $(0, 2\pi)$ and calculate $F(\pi)$.

---

[671] **N.B. (???) Relation to k.**

[672] **N.B. (???) Do we need $v + 2$? Is the result dependent on $w'(0)$?**

[673] Note that as in Remark 111 on p. 186, this makes a function "a bit more singular". However, this does not matter due to the argument in Footnote 670.

[674] This helps since we may want to consider $w(r) := (1 - |r|)^L$ on $[-1, 1]$. Then $L \geq r + 1$ works.

[675] Here it is *very important* how we index the summands. For the identities without alternating signs to work, we must consider the leading term 1 as a_1, the next term (1 or 2) as a_2 etc. (Considering the leading term as a_0 etc. leads to *different* answers! For example, the analogue of β_1 becomes $1 + \beta_1 + \beta_0 = 5/12$ instead of $\beta_1 = -1/12$. Indeed, denoting the position by the lower index, $1_0 + 2_1 + 3_2 + \dots = 1_0 + (1+1)_1 + (2+1)_2 + \dots = 1 + (1_1 + 2_2 + \dots) + (1_1 + 1_2 + \dots)$.)

Appendix: Quadratic reciprocity: Euler vs. Legendre

In context of these notes, the principal aim of this appendix is to try to reorient people who are already fluent with the Legendre's formulation of Quadratic Reciprocity, but who are bewildered by our use of (more important!) Euler's formulation. However, this appendix is self-contained, so may be also used by anybody who wants to find more about Quadratic Reciprocity than the basics of Euler formulation discussed so far (see p. 15). In the rest of these notes, we do not rely on the results of this appendix.

Note that here we do not discuss proofs of Quadratic Reciprocity — just what is common and what is different for its two important formulations.

Essentially, what we want to highlight here are the features which have a sporting chance to survive generalizations to the case of polynomials of higher degree. In this respect, the Euler's approach is much better than the Lagrange's one.

Euler formulation was future-proof

After Legendre discovered much more structure in the patterns considered at the beginning of this paper, the Euler's formulation have been mostly shadowed by the Legendre's one. It took more than a century for mathematicians to realize that *in the context of direct generalizations of Quadratic Reciprocity* the Euler formulation (see p. 15) is way more natural (compare with Footnote 32).

To a large extent, the aim of the simplest generalization ("the Class Field Theory") can be stated the same way as above: find "possible prime divisors of $P(n)$"; this theory describes the answer under the condition that P "leads to an abelian field extension".[676] It turns out that Euler's formulation extends almost literally to this case!

So nowadays in the context of number theory "at large" the Euler's formulation would be considered at least on equal footing with the (more elaborate) Legendre's one.

Recall that Euler's formulation describes the *symmetries* of the answers to the question of "possible prime divisors of a quadratic sequence": one can color $\mathbb{Z}$ red and green so that the coloring is periodic, (anti)symmetric, and a prime p appears as a divisor if and only if p is colored green.[677]

In particular, the collection of possible (anti)symmetries (the "*group of symmetries*") is an infinite dihedral group. Moreover, the Euler formulation says *how large* this group is comparing to the whole group of symmetries of $\mathbb{Z}$ (which is also an infinite dihedral group): its index is (a divisor of) $|4D|$, where D is the discriminant of the quadratic sequence.[679] [680]

In fact, the (anti-)palindromicity is a particular case of top-multiplicativity (considered in the following section).

---

[676] Any irreducible quadratic P satisfies this condition. An irreducible cubic P satisfies it if and only if its discriminant is a perfect square.

[677] This is a 2-colors variant of Euler's formulation. Above, on p. 15, we discussed a coloring into 3 colors, when the residues not mutually prime with $|4N|$ were colored gray.[678] On the other hand, given such a residue r, two columns $\pm r \bmod N$ contain at most one prime number (even in the exceptional case $N = 2r$, when these two columns degenerate into one). Because of this, it is easy to convert gray to red or green as required above.

[678] In fact, we described "gray" differently: as "this residue has only a finite number of prime number representatives". However, a residue $\bmod c$ not mutually prime with c cannot contain more than 1 prime number. Moreover, by Dirichlet theorem on arithmetic progressions, the other residues *are* represented by infinitely many prime numbers.

[679] It turns out that such a focus on symmetry survives even the widest possible generalization of our naive questions about prime divisors, given by the *Langlands* program. In fact, the usual formulations of the Langlands program are written completely in terms of describing particular flavors of symmetries.

[680] For pizza numbers, $D = -\frac{7}{4}$, which leads to the length 7 of the period. For polynomial with integer coefficients, one can replace $|4D|$ by $|D|$.

Similar to the Euler's formulation, the answers given by the Class Field Theory are encoded by a periodic coloring of $\mathbb{Z}$ into several colors. This coloring also has a suitable palindromicity property, as well as top-multiplicativity (discussed in the next section). The only difference is that colors match not the numbers $\{\pm 1\}$, but complex roots of 1 of a suitable degree d.[681]

In contrast, the generalizations of Lagrange's formulation turn out to be much more esoteric.

On the other hand, almost simultaneously with the discovery of Class Field Theory in the beginning of 20th century, another development took place: Quadratic Reciprocity entered the realm of "popular mathematics". And, as expected, what was popularized was offset by decades w.r.t. the frontier of math; it was the Legendre's formulation which entered the math pop-culture!

So, in the last century, a curious situation arised: the major textbooks on number theory as well as "capsule expositions" of Quadratic Reciprocity by the leading number theorists would highlight the Euler's approach. — And, at the same time, what most mathematicans *know* is the Legendre's one, since they "learned Quadratic Reciprocity too early", when they were more focused on the pop-math!

Legendre's notation and top-multiplicativity

Half a century after Euler, Legendre found a different way to write down the patterns of colors we observed above. He would use 1, -1 and 0 instead of our 🟢, 🔴 and ⚪ (see p. 15; this is compatible with our rules $-$⚪ $=$ ⚪ and $-$🟢 $=$ 🔴). At least, this convention allows a concise way to write down the property which was known long before Legendre: consider three sequences: "squares $+\,N$", "squares $+\,M$", and "squares $-\,MN$"; every prime number p acquires 3 colors each depending on whether it is a divisor of numbers in the first, and/or the second, and/or the third sequence. Then

$$\boxed{\text{The third color is “the product” of two other colors.}}$$

(Here the "product" is calculated according to the assignments of numbers $0, \pm 1$ to the colors).

Using the Legendre's notation (Legendre symbol) $\left(\frac{-N}{p}\right)$ for "the color" of prime p (taking values in $\{0, \pm 1\}$, with 0 meaning "p divides N") for "squares $+\,N$", this may be written as

$$\left(\frac{-N}{p}\right) \cdot \left(\frac{-M}{p}\right) = \left(\frac{NM}{p}\right)$$

(and this is a much simpler fact that it looks: it is an almost immediate corollary of non-0 residues $\operatorname{mod} p$ being invertible).

For example, from the list on p. 10 we can see that 23 cannot be a divisor of "squares $+\,3$", and the list on 12 shows that 23 is not a possible divisor of "squares $+\,1$". Using the rule above with $N = 3$ and $M = 1$, we can conclude that 23 must be a divisor of the sequence "squares $-\,3$". And indeed, $7^2 - 3 = 23 \times 2$.

This may be called *multiplicativity in the top argument*: when the top argument $NM = (-N)(-M)$ is a product, one can replace the symbol by a product of symbols.[682]

Euler's formulation implies the case of small $|N|$

Essentially, top-multiplicativity reduces calculation of $\left(\frac{n}{p}\right)$ to the cases when $n = -1$, or n is prime. Obviously, since a square $+\,n$ can be even for any n, the number 2 is going to be always green. Hence one can focus on odd primes p only.

Note that if $|n| = |N|$ is fixed and small, the first statement (the periodicity) in Euler's formulation reduces finding $\left(\frac{n}{p}\right)$ for all odd primes p to a check of a finitely many values of (odd) primes p. For

[681] In other words, the roots $e^{2\pi i n/d}$ of $z^d - 1 = 0$. The corresponding prime number p can be a divisor if and only if $z = 1$; the other possible values of z may be thought of "as different reasons for p not to be a divisor".

[682] This property explains why using $-N$ in the definition of Legendre's symbol simplifies working with this notation.

example, for $n = -1$ it is enough to check $p = 3, 5$; likewise, for $n = 2$ it is enough to check[683] $p = 3, 5, 7, 17$: these prime numbers cover all the odd residues mod $|4n|$ — and the even residues are going to be gray anyway.

Remark 118: For what follows, the values $n = -1, 2$ are of particular importance. Sometimes it is convenient to describe $\left(\frac{-1}{p}\right)$ and $\left(\frac{2}{p}\right)$ by a compact formula; the customary way is to condense the red/green colors given above into $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$ and $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$. For odd residues p (mod 4 or mod 8 correspondingly) these formulas match the colors found above.

However, the particular way the right-hand sides of these formulas are written down carries absolutely no significance. (There is a lot of other expressions which would give the same results!)[684]

Legendre's $p \leftrightarrow q$-reciprocity

Top-multiplicativity and two cases of Remark 118 reduce the calculation of Legendre symbols $\left(\frac{n}{p}\right)$ to the case $\left(\frac{q}{p}\right)$ where both p and q are different (positive) odd primes. By definition, $\left(\frac{n}{p}\right) = \left(\frac{n'}{p}\right)$ if $n \equiv_p n'$ ("top-periodicity"); hence one can further reduce the calculation to the case $q < p$.

The final nail to get a recipe for a quick calculation is the Legendre's $p \leftrightarrow q$-law ("reciprocity"):

$$\boxed{\text{The sign in } \left(\frac{q}{p}\right) = \pm\left(\frac{p}{q}\right) \text{ is "}-\text{" if } p \equiv_4 q \equiv_4 -1\text{, otherwise it is "}+\text{".}}$$

This assumes that p and q are distinct odd positive primes.

Remark 119: This helps in calculations since if $q < p$, the right-hand side has a smaller number at the bottom, so $p \leftrightarrow q$-law may be used in recursive algorithms. For example, to deal with the right-hand side, one can reduce $p \bmod q$ (by top-periodicity), factor the resulting residue $p' < q$ as $p' = p_1 \dots p_r$ and use top-multiplicativity — so now one needs just to calculate $\left(\frac{p_1}{q}\right)$, …, $\left(\frac{p_r}{q}\right)$ (and now p_1, …, p_r are much smaller than p). To these symbols, one can apply the $p \leftrightarrow q$-reciprocity again; etc.

It turns out that this gives a very quick algorithm for calculation of $\left(\frac{n}{p}\right)$. This algorithm is the principal reason for interest in $p \leftrightarrow q$-reciprocity.

Euler's formulation implies $p \leftrightarrow q$-reciprocity

We already saw that two special cases of $n = -1, 2$ for $\left(\frac{n}{p}\right)$ are *immediate* corollaries of the Euler' formulation. What may be yet more surprising is that the $p \leftrightarrow q$-reciprocity is *also* an immediate corollary!

Apparently, this fact was not discovered until 20th century: A. Scholz published this argument in his *Einführung in die Zahlentheorie* in 1939 as a part of his proof of Quadratic Reciprocity. (Baumgart–Lemmermeyer enumerate this as "Proof No. 175" in their list of 314 proofs.[685])

If $p \equiv_4 q$, this argument does not even need palindromicity, just periodicity! Indeed, write $q = p + 4n$; then

$$\left(\frac{q}{p}\right) \overset{\circ}{=} \left(\frac{q-p}{p}\right) = \left(\frac{4n}{p}\right) \overset{*}{=} \left(\frac{n}{p}\right) \underset{\circ}{=} \left(\frac{n}{p+4n}\right) = \left(\frac{n}{q}\right) \overset{*}{=} \left(\frac{4n}{q}\right) \overset{\circ}{=} \left(\frac{4n-q}{q}\right) = \left(\frac{-p}{q}\right) \overset{*}{=} \left(\frac{-1}{q}\right)\left(\frac{p}{q}\right).$$

---

683 In fact, this was already checked in the section on p. 12. Moreover, using Euler's palindromicity, the latter check can be reduced to $p = 3, 7$; in other words, this follows from $m^2 - 2$ being divisible by 7 for $m = 3$, and from $m^2 - 2$ being only $\pm 1 \bmod 3$ (enough to check $m = 0$ and $m = \pm 1$) which shows that $m^2 - 2$ cannot be divisible by 3.

684 Well, having p^2 in the formula for $\left(\frac{2}{p}\right)$ has an advantage: it makes palindromicity explicit.

685 In fact, this is one of only two proofs in their list which they mark as first proving the Euler's formulation, then deducing the rest from this formulation. The second such proof is Proof No. 243 by D. M. Goldschmidt of 1981.

(observe also that $\left(\frac{-1}{q}\right) \underset{\circ}{=} \left(\frac{-1}{p}\right)$). Here we mark $=$-signs which use periodicity in top/bottom arguments with $\circ$ above/below, and mark them with $*$ if they use top-multiplicativity. (Only the step marked with $\underset{\circ}{=}$ is non-obvious!)

Likewise, if $p \equiv_4 -q$, write $p + q = 4n$ (with $n > 0$). Then

$$\left(\frac{q}{p}\right) = \left(\frac{4n-p}{p}\right) \overset{\circ}{=} \left(\frac{4n}{p}\right) \overset{*}{=} \left(\frac{n}{p}\right) \overset{!}{=} \left(\frac{n}{q}\right),$$

likewise $\left(\frac{p}{q}\right) = \left(\frac{n}{q}\right)$, hence $\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$. The equality marked with "!" uses the palindromicity — and this is the only non-trivial step.

Legendre's formulation implies bottom-periodicity

The "Legendre's formulation" consists of 3 statements: the answers for $\left(\frac{n}{p}\right)$ with $n = -1, 2$ found in Remark 118, and the $p \leftrightarrow q$-reciprocity.

To deduce periodicity of $\left(\frac{n}{p}\right)$ in p from Legendre's formulation, we need to find (for a given n) a $|4n|$-periodic function $f(m)$ such that for prime $m = p$ it coincides with $\left(\frac{n}{p}\right)$. By top-multiplicativity, it is enough to consider the cases when $n = -1$, $n = 2$, or $n = q$ is an odd prime. In the first two cases the Legendre's formulation explicitly implies bottom-periodicity with a period of length $|4n|$.

However, in the last case $\left(\frac{q}{p}\right) = g(p)\left(\frac{p}{q}\right)$ for a certain 4-periodic function g. Now $\left(\frac{m}{q}\right)$ is explicitly q-periodic in m, which immediately implies that the right-hand side is $4q$-periodic.

Legendre's formulation implies palindromicity

To deduce the palindromicity from Legendre's formulation is trickier. When showing periodicity, we found a periodic function $f(m)$ such that for prime $m = p$ it coincides with $\left(\frac{n}{p}\right)$; this function takes values $0, \pm 1$, and it was constructed as a product over factors of n. Since palindromicity means $f(-m) = f(m)$ (provided $n > 0$), it is enough to show palindromicity for the case $n = p$ with a positive prime p. Since $\left(\frac{2}{p}\right)$ is an even function of $p \bmod 8$ (we already checked this — see the wheels above on p. 15!), we may assume that q is odd.

So what we need to show is $\left(\frac{q}{p}\right) = \left(\frac{q}{p'}\right)$ for distinct odd primes q, p and p' such that $4q | p + p'$. If $q \equiv_4 1$, then $q \leftrightarrow p$-reciprocity reduces this to $\left(\frac{p}{q}\right) = \left(\frac{p'}{q}\right)$, which follows from top-periodicity, top-multiplicativity, and from $\left(\frac{-1}{q}\right) = 1$. If $q \equiv_4 3$, then $q \leftrightarrow p$-reciprocity may be rewritten as $\left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{p}{q}\right)$. Therefore palindromicity is reduced to $\left(\frac{-1}{p}\right)\left(\frac{-1}{p'}\right)\left(\frac{-1}{q}\right) = 1$, or $\left(\frac{-1}{p}\right)\left(\frac{-1}{p'}\right) = -1$, which follows from $4 | p + p'$.

Remark 120: If we want to prove anti-palindromicity for $n = -N < 0$, then by multiplicativity, it is enough to consider the case $n = -1$. What we need to show is that $\left(\frac{-1}{p}\right)$ coincides with a $4N$-periodic odd function for any $N > 0$. However, we already know this for $N = 1$ — and this implies the general case.

Legendre's formulation and bottom-multiplicativity

There is another very important aspect of Quadratic Reciprocity which becomes much more conceptual in the Euler's formulation. A certain crucial feature, *bottom-multiplicativity*, is "hidden inside a definition" when one uses a Legendre's formulation.

Recall that the bottom-periodicity allows considering the argument p of $\left(\frac{n}{p}\right)$ as a residue mod $|4n|$. Now the *bottom-multiplicativity* can be stated parallelly to top-multiplicativity:

$$\left(\frac{n}{a}\right)\left(\frac{n}{b}\right) = \left(\frac{n}{ab}\right) \qquad \text{with } a,\ b \text{ residues mod } |4n|.$$

However, the *meaning* of this is very different from the top-multiplicativity, since we defined $\left(\frac{n}{p}\right)$ just for prime values of p: what this equality says is that if three (positive) prime numbers p, p', p'' satisfy $p'p'' \equiv_{|4n|} p$, then $\left(\frac{n}{p'}\right)\left(\frac{n}{p''}\right) = \left(\frac{n}{p}\right)$.

This was the Euler-styled approach to the bottom-multiplicativity. In Legendre's approach, it is kind of hidden behind a trick: so far, we defined $\left(\frac{n}{p}\right)$ just in the case of prime p (see p. 208). In fact, Jacobi *defined*[686] his generalization of Legendre symbol for any odd $m > 0$ by multiplicativity: $\left(\frac{n}{p_1 \dots p_r}\right) \stackrel{\text{def}}{=} \left(\frac{n}{p_1}\right) \dots \left(\frac{n}{p_r}\right)$ with prime $p_1, \dots, p_r$.[687] Note that what was surprising in Euler's formulation becomes a *definition* in the Legendre's (Jacobi's) approach.

However, when Legendre's symbol $\left(\frac{n}{m}\right)$ is defined for any[688] odd $m > 0$, the bottom-periodicity can be stated in a much more straightforward way: $\left(\frac{n}{m}\right) = \left(\frac{n}{m+4n}\right)$ if $m > 0$, $m + 4n > 0$ are odd. This is completely parallel to the top-periodicity (which preserves its form with a composite m as well): $\left(\frac{n}{m}\right) = \left(\frac{n+m}{m}\right)$.

Conclusion: Before the observation above, to color a residue m mod $|4n|$ on the conductor wheel we needed to find a prime number $p \equiv_{|4n|} m$, and use $\left(\frac{n}{p}\right)$ as the color. Now one can factor $m = p_1 \dots p_r$ instead, and use $\left(\frac{n}{p_1}\right) \dots \left(\frac{n}{p_r}\right)$. (This is using the bottom-periodicity vs. the bottom-multiplicativity.)

Remark 121: Likewise, now the palindromicity may be rewritten as $\left(\frac{n}{m}\right) = \left(\frac{n}{4n-m}\right)$ if $n > 0$, $0 < m < 4n$, similarly for anti-palindromicity (for $n < 0$). In fact, the found above formulas for $n = -1$ and $n = 2$ preserve their form for a composite m as well; same for the top-multiplicativity. In particular, $\left(\frac{-n}{m}\right) = \left(\frac{-1}{m}\right)\left(\frac{n}{m}\right)$.

Using these rules, one can change n to make $|n| \le ½m$, or change m to make $m \le 2|n|$, or factor m. Doing these steps in this order, one can reduce calculation of $\left(\frac{n}{m}\right)$ to the case of prime $m < |2n|$; then one can repeat this round again (etc). The process stops if $m = 1$ (when $\left(\frac{n}{m}\right) = 1$), or if $n = 0$ (when $\left(\frac{n}{m}\right) = 0$ if $m \ne 0$).

This gives an algorithm for recursive calculation of $\left(\frac{n}{m}\right)$ which does not use $p \leftrightarrow q$-reciprocity. In fact, it terminates very quickly even without using any "fancy" factorization methods.[689]

Compare Euler's and Legendre's formulations

One can conclude:

- It is as easy to deduce the Legendre's formulation from the Euler's one as in the other direction.

---

[686] About half a century after Legendre.

[687] We want to emphasize it: $\left(\frac{n}{m}\right)$ for a non-prime m is *not* defined as the "color" of m for the sequence squares $-\, n$. Instead, it "combines" the colors of the prime factors of m.

This distinctness is highlighted in Remark 123.

[688] There are several convenient ways to extend this to $m \le 0$, but different contexts benefit from different extensions. So it is reasonable to restrict attention to $m > 0$.

[689] Indeed, the only case when the first two steps do not decrease $|n| + m$ a lot is when $m = 2n - a$ and $a \ll n$; then they reduce n, m to become $n' = n - a$, $m' = 2n' - a = m - 2a$. Choosing a small odd prime $p \nmid a$, a few such steps would ensure $p|m'$, and one will be able to decrease m' a lot by factoring out p.

- The Legendre's $p \leftrightarrow q$-symmetry shows interrelation of prime divisors of two *different* quadratic sequences.
- The Euler's formulation shows an infinite-dihedral symmetry of prime divisors of any *particular* quadratic sequence.
- Both approaches lead to quick algorithms for calculation of $\left(\frac{n}{p}\right)$.

In short: the Euler's formulation is "much more fundamental" — and contemporary ("*adelic*") — in its approach: it chooses a particular equation $x^2 = n$, and examines existence of its solution $\bmod p$ for different prime numbers p.

Therefore, while our "colorings of conductor wheels" do not look "very mathematical" if one is fluent with just the Legendre's approach, in fact they bring us much closer to the high-voltage wire in the guts of Quadratic Reciprocity.[690]

[690] I paraphrase Jordan Ellenberg's "You feel you've reached into the universe's guts and put your hand on the wire" on the nature of mathematical understanding, from *How Not to Be Wrong*.

Appendix: A few more words on Quadratic Reciprocity

The case $p = 2$ of $\left(\frac{n}{p}\right)$ and the shortest period

Above, on p. 208, we argued why the case $p = 2$ is simpler than other primes: for any n, the sequence "squares $- n$" contains an even number, hence the number 2 is going to be always green (which Legendre encodes as 1). However, this does not still address the question about the color of the residue of 2 mod c on the conductor wheel! First, the color of 2 may be an exception comparing to other prime numbers $p \equiv_c 2$ (as we saw for $n = -3$ on the 3-wheel on p. 15); second, this residue may be colored gray (as we saw for the same $n = -3$ on the 6-wheel on p. 15)!

In particular, the answer depends on our choice of c, the size of the conductor wheel.[691] While the Euler's formulation implies that $c = |4n|$ *will work*, it does not claim that shorter periods are not possible.

So we need to know what is the *shortest* period in Euler's formulation "with exceptions", and what are the possible exceptions. It turns out[692] that the answer is[694] $C = 2^M|n_0|$, depending only on the square-free part[695] n_0 of n, here M depends only on n_0 mod 4: if $n_0 \equiv_4 1$, then $M = 0$; otherwise $M = 2$. Moreover:[696]

$$\boxed{\text{On the } C\text{-wheel, there is no exception unless } n_0 \equiv_8 5, \text{ when 2 } \textit{is} \text{ the only exception.}}$$

(We already saw such an exception happening for $n = -3$ on p. 13.) Hence 2 mod C is going to be colored gray unless $n_0 \equiv_4 1$, when it is colored as $\left(\frac{2}{n_0}\right)$ (which coincides with the RHS of $\frac{n_0+1}{2} \equiv_4 \pm 1$). Obviously,[697] an odd prime p is gray if and only if it divides n_0.

Conclusion: put $C' := C$ unless $n_0 \equiv_8 5$, when $C' = 2C$; then C'-wheel is the smallest conductor-wheel with no exceptional primes.

---

[691] For example, if $2|c$, then the color of 2 mod c must be gray, which would lead to $\left(\frac{n}{2}\right) = 0$.

[692] Indeed, when we deduced the bottom-periodicity from Legendre formulation (see p. 210), we already saw that the color $\left(\frac{n}{p}\right)$ of p may be rewritten as $\left(\frac{p}{q_1}\right) \dots \left(\frac{p}{q_k}\right)$ (here q_l are odd prime divisors of n which enter the prime decomposition of n with odd exponents), possibly multiplied by $\left(\frac{-1}{p}\right)$ and/or $\left(\frac{2}{p}\right)$ (which we know to be 4-periodic in p, and 8-periodic in p). (This assumes that p is odd and mutually prime with n.)

For a fixed odd prime q, note that $\left(\frac{p}{q}\right)$ takes *both* values ± 1 for infinitely many primes p. Moreover, the collection of numbers $\left(\frac{p}{q_1}\right)$, …, $\left(\frac{p}{q_k}\right)$ is a combinations of ± 1, and *any* such a combination appears for infinitely many odd primes p (and here one can also require $p \equiv_8 k$ for any odd k).[693] Since $\left(\frac{p}{q_1}\right)$ is q_1-periodic, this implies that the shortest period which works with finitely many exceptional primes p has length $C := 2^m q_1 \dots q_k$ with $m = 0, 2, 3$ depending on whether $\left(\frac{-1}{p}\right)$ and/or $\left(\frac{2}{p}\right)$ appears above. (Moreover, assume $p \neq 2$; then p mutually prime with n cannot be an exception, and the remaining $s_\lrcorner$p are gray — as the only prime number in its column.) Therefore, even if one allows exceptions, the shortest period has length C.

A more detailed examination of the argument above shows that C depends just on the square-free part n_0 of n, and m depends just on n_0 mod 4: if n_0 is even, then $m = 3$; if $n_0 \equiv_4 3$, then $m = 2$; otherwise $m = 0$. Hence 2 mod C is going to be colored gray unless $n_0 \equiv_4 1$.

In the latter case, C is odd, and one can also find out when $p = 2$ is going to be an exception. Indeed, the argument above shows that all odd p with $p \equiv_C 2$ have the same color $\left(\frac{2}{C}\right)$ — while 2 is always green.

[693] All these statements follow from the Chinese remainder theorem, and from Dirichlet theorem on arithmetic progressions.

[694] One can recognize this as the discriminant of the field $\mathbb{Q}[\sqrt{n}\,]$.

[695] Here we write $n = n_0 K^2$ with the maximal possible K.

[696] Indeed, all prime divisors of C appear as different different residues mod C and do not share their residues with other primes. Hence their color *is* determined by their residues mod C. **Conclusion:** the only exception may be $p = 2$, and just if $m = 0$.

[697] See Footnote 678.

Remark 122: It took almost a century after Legendre (and 50 years after Jacobi) to realize the importance of treating $p = 2$ like this! Kronecker noticed that defining $\left(\frac{n}{p}\right)$ for $p = 2$ using the C-wheel above allows extension to $\left(\frac{n}{m}\right)$ (with any m) by bottom-multiplicativity. What is crucial is that this extension is simultaneously bottom-periodic, bottom-multiplicative, is defined for every $m > 0$, and has close relation to our problem of divisors of numbers in a quadratic sequence.[698]

Remark 123: For example, observe two colored rows we matched on p. 8. Now one can recognize the bottom row as colored according to $\left(\frac{-7}{m}\right)$ in the sense of Kronecker.[699] (The "complete" match between these rows is due to the discriminant for our sequence being $D = -7 \not\equiv_8 5$, hence $p = 2$ is not an exception.)

Divisors of $P(n)$ with quadratic P

The considerations above describe more or less completely the prime divisors of numbers in any polynomial sequence $P(n)$ of degree 2 (compare with Remark 6). Indeed, if P is decomposable, then as we saw on p. 5, already the divisors of one linear factor "would cover" all prime numbers (with just a finite number of exceptions).

If P is irreducible, then

- Prime divisors p of the numbers in the sequence coincide with p such that $P(x) = 0$ has solutions $\bmod p$. (Here $p = 2$ may be an exception since it is possible that P takes integer values, and coefficients of P have 2 as a denominator: triangular numbers!)
- The "quadratic formula" $\frac{-b \pm \sqrt{D}}{2a}$, $D := b^2 - 4ac$, shows that existence of solutions $\bmod p$ is equivalent to existence of solutions of $x^2 - D \equiv_p 0$, provided $p \nmid 2a$ (again, this is a finite number of exceptional primes p␣s).

Conclusion: with a finite number of exceptions, prime divisors of numbers $P(x)$ are the same as prime divisors of numbers $x^2 - D$.

---

[698] In fact, there are several flavors of the definition of Kronecker symbol. Our flavor is compatible with them where they all agree.

The reason for discrepancies is that our n and C are in a certain way interchangeable (since $\left(\frac{n}{m}\right) = \left(\frac{C}{m}\right)$ for any m), so it is not clear "whether we are calculating a function of n, or a function of $C =: C(n)$". Since any value of n makes sense, if we accept that our description gives $\left(\frac{n}{m}\right)$, then there would be no possible discrepancy.

However, it turns out that the approach "that 'matching the colors' describes not $\left(\frac{n}{m}\right)$ but $\left(\frac{C(n)}{m}\right)$" is more fruitful. But not every number is a possible value of $C(n)$, since $C(n)$ is a *fundamental discriminant*! Hence if we consider $\left(\frac{n}{m}\right) = \left(\frac{C}{m}\right)$ as a function of $C = C(n)$, then our description does not define it on every number C. In particular, one may imagine several different extensions with the arbitrary upper argument (not necessarily a fundamental discriminant)—and different extensions are useful in different situtations.

[699] This is not exactly true since we were using slightly different notations on p. 8. We have not introduced "the gray color" yet, so m with $7|m$ was colored green, not gray. (Recall that $\left(\frac{n}{p}\right) = 0$ if $p|n$.)

Used resources

Most of the references we used in these notes are accompanied by a PDF crosslink to the corresponding resource. The notable exceptions are the collection edited by Cassels and Fröhlich (from which I found out that what is important about quadratic reciprocity is not the $p \leftrightarrow q$-law, but the periodicity — or, as we call it here, the Euler's formulation), Lang's book *Elliptic functions* (which teached me the relation of the tower of congruence-groups with the adelic approach), the first half[700] of Jared Weinstein's review of reciprocity laws, Example 4.7.5 of which led me to Gelbart's paper with quite a detailed exposition,[701] Lemmermeyer's book on higher reciprocity laws (however, we mentioned Baumgart–Lemmermeyer's compendium of proofs of quadratic reciprocity).

Another text which the readers may find useful is Keith Conrad' notes on history of Class Field Theory, as well as Roquette's book on the related subject.

For the simplest example of how modularity may be related to cubic equations of negative discriminant see Part 6 of Jerry Shurman's notes "Toward Modularity: the Simplest Non-Abelian Example".

For me, the Apanasov, Krushkal and Gusevski's book *Kleinian Groups and Uniformization in Examples and Problems* was very inspiring as a compendium of tricks (and treats!) about groups of symmetries in non-Euclidean geometries. (In fact, Harvey's review of this book highlights many objectives and difficulties equally applicable to the design of our notes!)

The plot on p. 34 is from series of papers by J. Bernstein, F. Chamizo, S. Miller, A. Reznikov, Wil. Schmidt of 90s and 00s. (My interest in these topics stemmed a bit later from answering some questions of Don Zagier using a similar approach.)

For guidance in these labyrinths, I'm indebtful to hints from T. Barnet-Lamb, N. Gurevich and A. Reznikov. (This lists only what happened in the last decade; to clear my earlier misunderstandings in these topics, it probably took whole divisions of people — and it is really sad that now I cannot list them all!)

To continue further, probably the best starting points are the discussion *The Langlands program for beginners* on **StackExchange** and slides by Sury. One can continue by following the discussion *Zeta Functions: Dedekind V* in **n-Cat Café** (as well as following the links mentioned in these discussions).

Another very convenient resource is the online tables of number fields. For example, a query with

`Degree=3, r`$_1$`=3*, |D|=1..1000, sort`$_1$`=Gal, sort`$_2$`=|D|, sort`$_3$`=h`

would result in a list of 27 real cubic fields of small discriminant (first cyclic, then non-cyclic ones).[702]

Note that when searching for polynomials of degree 4, the Galois groups are enumerated (by "`T-num`") according to their size (and, for the first two, rank), with $\mathbb{Z}/(4)$, $\mathbb{Z}/(2)\times\mathbb{Z}/(2)$, D_4, A_4 and S_4 getting `T-num␣`s from 1 to 5. (The first two are named C_4 and V_4 in the reports.)

How to compute

As we said, the recent updates to GP/PARI math-calculator made a lot of tedious calculations much simpler to perform. Here we want to collect tidbits about these calculations. First, below we

[700] The second half of this review is dedicated to Scholtzefication, which looks unrelated to what we discuss here.

[701] See Footnote 552 on p. 172.

[702] One can check that all these cyclic fields, and the non-cyclic ones with 2 smallest values $D = 2^2 \cdot 27, 229$ of discriminant (as well as 4 more of 20 remaining non-cyclic fields) appear in our family $M \cdot$ "Tetrahedral numbers" $+ N$ for relatively small values of M and N.

Likewise, from 10 complex cubic fields with discriminant up to -110, the family includes all but three, with $D = -31, -2^2 \cdot 19, -3 \cdot 29$. In particular, it includes one with the smallest magnitude 23 of discriminant, which we investigated in Section on p. 51.

assume that our polynomial `P` takes integer values, and is in the variable `X`; then one can get the array of "exceptional primes" for `P` as

```
my(Den = denominator(content(P))); factor(abs(poldisc(PP=P*Den))*Den*polcoeff(PP,poldegree(PP)))[,1].
```

One can check that `P` is irreducible by `1==factor(P)[,2]`.

For a non-exceptional prime `p` and an irreducible `P` one can find $\widetilde{N}_p^{\mathrm{res}}$ as

```
# select(x -> x == 1, factormod(P,p,1)[,1]).
```

The alternative way is `poldegree(gcd(P+Mod(0,p),X^p-X))`, but since GP/PARI has no "sparse polynomials", the "calculation" of `X^p` for a large `p` may take too much stack space (and/or time). Both methods may be generalized to finding $\widetilde{N}_{p^k}^{\mathrm{Gal}}$: in the first expression, one should replace `p` by `[ffinit(p,k,varlower("PP")),p]`; for the second one, replace `X^p-X` by `X^{p^k}-X`.

For the following discussion, assume we initialized a few pieces of data with

```
NN = lfunan(lf = lfuninit(lfuncreate(nf = nfinit(P,3)[1]),[0]),1000);.
```

Here one can replace `1000` by a larger number, and get a longer array `NN`. Note that `NN[p]` $= N_p + 1$ for a prime p, likewise $N_{p^k} =$ `NN[p^k]` $-$ `NN[p^(k-1)]` — *including* the exceptional values of p.

Since for Artin's L-function of a field the conductor is the discriminant of the field, one can find the conductor as `nf.disc`. Moreover, if one wants to calculate N_{p^k} "by hand", to choose the correct sequence of 5 listed in Items (c) and (d) on p. 60 it is enough to know the pair $[N_p, N_{p^2}]$.

To see the prime decomposition of `p` in the field `nf`, inspect

```
Mat(apply(x -> ["base-prime",x[1][1],"ramification",x[1][3],"ff-degree",x[1][4],"multiplicity",x[2]], Col(idealfactor(nf,p))))
```

The first three cases (those which may appear for "non-exceptional" primes) correspond to 0, 1, or 3 factors with `"ff-degree"` being 1 (while no factors have `"ramification"` larger than 1). The last two cases correspond to presence of factors with `"ramification"` being 2 and 3 correspondingly.

For these 5 cases, the p-local factor of the denominator of L-function is $1-p^3$ (no points over $\mathbb{F}_p$ means that there is one point over $\mathbb{F}_{p^3}$), or $(1-p)(1-p^2)$ (one point over $\mathbb{F}_p$, unramified, means that there is one other point over $\mathbb{F}_{p^2}$), or $(1-p)^3$ (three points over $\mathbb{F}_p$), or $(1-p)^2$ (two points over $\mathbb{F}_p$, one ramified), or $1-p$ (one triple-ramified point over $\mathbb{F}_p$). Since disjoint union of manifolds corresponds to (product of their equations and to) a product of L-functions, and the L-function of a point (which is a solution to $X=0$) has local factor of the denominator being $1-p$ (so it is the Riemann ζ-function), the process of "distillation" (which proceeds "as if it removes" a point) would *divide* these local factors by $1-p$.

Conclusion: in these 5 cases, after distillation one gets $1+p+p^2$, or $1-p^2$, or $(1-p)^2$, or $1-p$, or 1. Replacing p by a formal variable $\boldsymbol{p}$ and inverting, one gets 5 series in $\boldsymbol{p}$, and the coefficients at $\boldsymbol{p}^k$, $k>0$, are exactly as described (above???).

So N_p is `#idealfactorBase(nf,p)-1` (here `idealfactorBase()` is like `idealfactor()`, but returns only the vector of factors defined over the base field $\mathbb{F}_p$; see the definition below), and to distinguish the second and fifth cases (when $N_p = 0$) one can check `idealfactor(nf,p)[1][1][3]>1` (which detects ramification). This means that the function

```
Ntype(p,nf)=my(f=idealfactorBase(nf,p));if(#f!=1,return([#f-1,0]);[0,f[1][1][3]>1];
```

allows to determine the type of the sequence for every prime ("exceptional" or not):

```
coeff3Npow(k,t)=if(t[1]==2,k+1,t[1]==1,1,t[1]==-1,(k+2)%3-1,t[2],0,!(k%2));
```

here we use the **case**-like extended `if()` introduced in recent GP/PARI.[703]

[703] One can cut-and-paste the code below (*including* the intervening text) into `gp`.

```
NtypeNonSpec(p,P)=my(F);[# select(x -> x == 1, (F=factormod(P,p,1)[,1])) - 1,#F,poldegree(P),0];
idealfactorBase2(nf,p)=my(F);[select(x -> x[1][4] == 1, F=Col(idealfactor(nf,p))),F];
Ntype(p,nf)=if(type(nf)=="t_POL",nf=nfinit(nf,3)[1]);my([f,F]=idealfactorBase2(nf,p));my(d=sum(k=1,#F,F[k][1][4]));\
           return([#f-1,#F,d,sum(k=1,#F,F[k][1][3]-1)]);   \\ 2:total # of factors 3:deg skeleton; 4:"extra" ramific,unused;
coeffNpow(k,t)=if(t[3]==3,coeff3Npow(k,t),t[3]>3,coeff4__5Npow(k,t),coeff1__2Npow(k,t));
   \\ good for -1,0,2; and -2 (for Artin only; [0,-1] for Artin too)
coeff3Npow(k,t)=if(t[1]==2,k+1,!t[1],if(t[2]<0,(-1)^(k\2),1)*!(k%2),t[1]==-2,(-1)^k*(k+1),(k+2)%3-1);
coeff1__2Npow(k,t)=t[1]^k;                                   \\ good for 0,±1
ppFactor(x)=["base-prime",x[1][1],"ramification",x[1][3],"ff-degree",x[1][4],"multiplicity",x[2]];
specPrimes(P)=my(Den = denominator(content(P))); factor(abs(poldisc(P=P*Den))*Den*polcoeff(P,poldegree(P)))[,1];
reportSpecFactors(P,nf=nfinit(P,3)[1])=my(ps=specPrimes(P));\
                                for(n=1,#ps,printp(Mat(apply(x -> ppFactor(x), Col(idealfactor(nf,ps[n]))))));
    \\ Artinization by massaging: good for deg=4, codiscriminant=-3; except 0,-1,0,1,0,-1,...
MSGart(t,p)=my(fix=centerlift(Mod(p,3)));if(fix,[t[1]-fix,-(t[3]==4&&t[1]==-1&&t[2]<=1),t[3]-1],t);
MSGid(t,p)=t;  MSG=MSGid;
    \\ The following operate on a global array N_n.  We do not overwrite known elements of N_n[]!
N_n_preINIT(LIM)= N_n=vector(LIM,i,"");0;
N_n_fill_p(p,t)= my(Lim=floor(log(#N_n)/log(p))); for(POW=1,Lim,if(N_n[p^POW]=="",N_n[p^POW]=coeffNpow(POW,MSG(t,p))));
N_n_INIT_LST(LST)= for(n=1,#LST,N_n_fill_p(LST[n][1],LST[n][2]));
N_n_INIT_SPEC_ps(P,LST=0)= if(LST,N_n_INIT_LST(LST);return); \
                          my(ps=specPrimes(P),nf=nfinit(P,3)[1]);for(n=1,#ps,N_n_fill_p(ps[n],Ntype(ps[n],nf)));
    \\ Check avoids calling NtypeNonSpec() in presence of denominators
N_n_INITpsNONSPEC(P)= forprime(p=2,#N_n,if(N_n[p]=="",N_n_fill_p(p,NtypeNonSpec(p,P))));
N_n_fill_N(n)=if(""!=N_n[n],return); my(d=factor(n),D=1); for(i=1,#d[,2],D*=N_n[d[i,1]^d[i,2]]);N_n[n]=D;
    \\ The last 2 statements compactify (arrays with edited entries are not memory-efficient)
N_n_INIT(LIM,P,LST=0)= N_n_preINIT(LIM);N_n_INIT_SPEC_ps(P,LST);N_n_INITpsNONSPEC(P);for(n=1,#N_n,N_n_fill_N(n));N_n=N_n;0;
```

Now doing `N_n_INIT(1000,X*(X^2-1)+12)` initializes the array `N_n` with 1000 first numbers N_k.

```
\\ Intermediate data to calc the Fourier transform: in global array of poly PN_n
PN_nINIT(LIM,P,LST=0) = PN_n=0; N_n_INIT(LIM,P,LST); PN_n=apply(f -> Polrev(vector(#N_n\f,n,1.*N_n[n]/n)), [1,2]);0;
N_n_cFt(X,f=1)=my(v=exp(I*X));v*subst(PN_n[f],x,v);
N_n_Ft(X,f=1)=imag(N_n_cFt(X,f));
```

After `PN_nINIT(LIM,P)`,[704] one can plot with `ploth(X=-0.1,7,[N_n_Ft(X,2),N_n_Ft(X)])`. This would draw the Fourier transform of half the array `N_n`, and the whole array — so that one can see whether one needs to calculate more elements of `N_n`. (To get pictures of this report, we needed `LIM` of order of magnitude of million(s).)

(Plotting with `my(c);ploth(X=-0.1,7,[N_n_Ft(X,2),imag(c=N_n_cFt(X)),real(c)])` would show the real and the imaginary part.)

If one wants to cover the case of degree 4, one should add this code:

```
coeff4_0Npow(k,t)=if(t[2]>1,(1+k\2)*(1-2*(k%2)),[1,-1,0,0][1+k%4]);
coeff4__5Npow(k,t)=if(t[3]>4,coeff5Npow(k,t),t[1]==3,(k+1)*(k+2)/2,t[1]==1,1+k\2,t[1]==-1,coeff4_0Npow(k,t),!(k%3));
\\ deg=5: Roots: 5 → (1-u)^4 ⇒ (d+1)*(d+2)*(d+3)/6; 3 → (1-u)^2(1-u^2) ⇒ (d+2)^2/4 if 2|d, (d+1)(d+3)/4 otherw;
\\ 2 → (1-u)(1-u^3) ⇒ 1+d\3; ??? 1 → {1}{2}{2} (1-u^2)^2 1+d/2 if 2|d, 0 otherwise or {1}{4} 1-u^4 ⇒ 1 if 4|d, 0 otherwise
\\ 0 {2}{3} (1+u)(1-u^3) 1,-1,1,0,0,0 repeated; {5} 1,-1,0,0,0 repeated
coeff5_01Npow(k,t)=if(t[1]==-1,if(t[2]>1,[1,-1,1,0,0,0][1+k%6],[1,-1,0,0,0][1+k%5]),t[2]>2,(!bitand(k,1))*(1+k/2),!bitand(k,3));
coeff5Npow(k,t)=if(t[3]>5,error("power"),t[1]==4,(k+1)*(k+2)*(k+3)/6,t[1]==2,(k+2)^2\4,t[1]==1,1+k\3,coeff5_01Npow(k,t));
```

Above, we compute the array `N_n` "manually" following the rules of Items (c) and (d) on p. 60; however, `PARI` has a primitive `lfunan()` which calculates essentially the same data. To compare these two results, use:

```
ckPrN(p,n)=my(a,b,c);if((a=N_n[p^n])==(b=NN[p^n])-(c=NN[p^(n-1)]),,print(p"^"n":\t"a"\t"b" - "c));
ckPr(p,L)=for(n=1,floor(log(L)/log(p)),ckPrN(p,n));
ck(L=#NN)=forprime(p=2,L,ckPr(p,L));
```

---

[704] This usage assumes that `P` is irreducible. Otherwise one needs to specify the parameter `LST` explicitly. For example, for the case $M = 16$ considered on 63,

`LST = [[2,[0,0,3]],[3,[2,0,3]],[13,[1,0,3]]]`

One can find the primes to include by `specPrimes(P)`. To find the suitable arrays, follow the explanations above and the section on p. 115. (Here `[2,0,3]` means "N_{p^k} behaves same as for a cubic polynomial with $N_p = 2$".)

```
ckP(P,LIM=1000000)=N_n_INIT(LIM,P);NN = lfunan(lf = lfuninit(lfuncreate(nf = nfinit(P,3)[1]),[0]),LIM);ck();
repP(p)=[[N_n[p^k]|k<-[0..floor(log(#NN)/log(p))]],[NN[p^k]|k<-[0..floor(log(#NN)/log(p))]]];
```

A few more tidbits: one can find $\left(\frac{a}{b}\right)$ by `kronecker(a,b)`. One can find ℓ_s (defined on p. 136) as `lfun(13,-s)`. In the case of modular forms, if one knows N_p for a few values of p, one can use `mfeigensearch([[1..LIMc],1], [[`p_1`,`N_{p_1}`],...,[`p_k`,`N_{p_k}`]])` to list all cases for sequences N_m with these particular values,[705] [706] up to $c =$ `LIMc`.

The code to find "runs" (see Footnote 492 on p. 157), via `findSW(20*10^9,1)` (which may count wrongly divisors of the discriminant, since we use `NtypeNonSpec`):

```
findSW(M,rep,N=1,stp,pr)=my(i=0,j=0,t,prev);\
  forprime(p=1,M,my(v=NtypeNonSpec(p,P));\
    if(v[1],\
      if(v[1]>1,i++;j++;if(p>=N,t=rep*(4*j-2.5-i)>0;if(pr||t,print([i,4*j-2,p,v,prev,j,i-4*j+1]));if(t,if(stp,break,rep*=-1)));prev=p
      i++;prev=p));\
  1.*(i+1)/j
```

[705] For example, one can run `my(p);mf=mfeigensearch([[1..800],1], vector(50,k,[p=prime(k),N_n[p]]));#mf`, increasing the limit `800` until at least one modular form is found, and increasing the count `50` until the number of found forms decreases to 1. Then `mf[1]` contais the found form; one can inspect it via `mfdescribe(mf[1])`.

[706] **N.B. (???) The only way I know to find the conductor is to decrease `LIMc` until the found form disappears.**