

From Quadratic Reciprocity to the Langlands Program — in the pictorial way

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 appendix¹ 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 mark such parts with the sign .

Essentially, our aim is to expose the simplest case (of those [not covered by Class Field Theory](#)) where both representations connected by the Langlands program have “almost” completely elementary counterparts (“almost” since one of them needs Fourier series...²). In fact, we do not mention representations anywhere else in 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.)

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

Contents

Digest: Meetings on Quadratic Reciprocity etc (Grades 1–4, Ilya 2018-05)	
Divisors of polynomial sequences: the simplest cases	4
Wheels	4
Example in $\deg = 2$: pizza numbers	5
Conductor of another sequence of degree 2: “squares + 3”	8
Divisor of sequences of $\deg = 2$: two more cases	10

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

²It is not exactly clear how much of this theory is *needed*. I expect that it would definitely help if one understands that Fourier transform is a *bijection*, and how decay of Fourier coefficients is related to smoothness of the sum of Fourier series.

Improved coloring	13
Euler's formulation	14
Degree 3: a coarse-grained approach (Grades 3–4, Ilya 2018-05)	
From degree 2 to degree 3	16
Example: Divisors of “tetrahedral numbers + 2”	17
Recent developments: the Langlands program	18
The simplest Langlands' patterns in more detail	
Bread crumbs: A very coarse outline of the Langlands' pattern	20
The appetizers for what follows	23
In more detail	27
Fractality laws: the simplified example	27
The zoo of fractality laws	30
Example: the toy fractality law as a symmetry	30
The Cantor set of non-smooth points on the example plot	32
All the fractal transformations together: infinities and regularizations	35
Fractality law for antiderivative	36
The first “real life” case	36
A simpler-to-plot example: $M = 6$	38
Maass fractality laws	41
The transliteration rules	45
Appendix: More patterns, and additional pictorial examples	
Finer points of the transliteration rules	48
“Purification” and Motives	50
Fractional-linear transformations	53
Prime conductors and “Tetrahedral + 2” again	54
The honest fractality law for $F^{(-1)}(t)$	57
Plots for degree 2	58
Denominators in Weil Conjectures	59
Maass fractality laws: decomposable and abelian cases	61
Historical approach: cases that <i>only</i> the Langlands program can explain	66
On Lobachevsky geometry and zones of self-similarity	
The groups of symmetries	68
Lobachevsky-symmetries: the case $c = 1$	69
Enhance the picture: the gray disks	70
The case $c = 5$	71
The gray disks and the “special zones”	72
Covering properties of the zones of horizon-self-similarity	75
More symmetries	77
Adding “sign-flipping” zones	79
All horizon-similar zones	80
Complement to zones	83
Appendix²: Eisenstein series	
Examples of dealing with Eisenstein series	85
Appendix: On verification, — and the future	
More details on the M -family	94
On gamma-factors	95
Examples: degree 4	96
Verification and further examples	108
The bird's eye view and the Grothendieck group of manifolds	110

Appendix: Quadratic reciprocity: Euler vs. Legendre

Euler formulation was future-proof	112
Legendre's notation and top-multiplicativity	113
Euler's formulation implies the case of small $ N $	113
Legendre's $p \leftrightarrow q$ -reciprocity	114
Euler's formulation implies $p \leftrightarrow q$ -reciprocity	114
Legendre's formulation implies bottom-periodicity	115
Legendre's formulation implies palindromicity	115
Legendre's formulation and bottom-multiplicativity	115
Compare Euler's and Legendre's formulations	116

Appendix: A few more words on Quadratic Reciprocity

The case $p = 2$ of $\left(\frac{n}{p}\right)$ and the shortest period	118
Divisors of $P(n)$ with quadratic P	119


Used resources

How to compute	120
----------------	-----

Digest: Meetings on Quadratic Reciprocity etc (Grades 1–4, Ilya 2018-05)

This is a digest of what we did in our Math Circles. Unless you are interested in teaching, it is OK to read this “diagonally”. (Only the results of the last two sections are going to be used in what follows.)

Divisors of polynomial sequences: the simplest cases

The problem we investigate in these notes is describing the divisors of numbers in a given polynomial sequence P_k . In more abstract language, this would be a particular case of solving (systems of) polynomial equations in modular arithmetic.  However, our particular formulation allows us to introduce this problem to the kids quickly—all that is needed is a rudimentary knowledge of multiplication.³ An impatient reader may want to skip the examples and jump to the section on p. 14.

For every number n , we ask: does it divide one of the numbers P_k ? The answer is a function of n with values **Yes** or **No**. (For pedants: above, “one” means “one or more”.)

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 $P_k = ak + b$, then for every n mutually prime with a the answer is **Yes**.⁴

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_k .

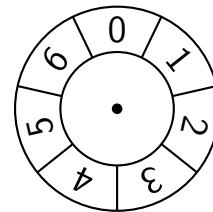
As this shows, even in the simplest cases, allowing a finite number of exceptions leads to significant simplifications of the statements. Moreover, restricting attention to prime divisors may lead to further similar simplifications. For example, one can cover all the cases as in:

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_k ?” does not depend on p .

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


Wheels


 One way to visualize the divisors of numbers in arithmetic progression is to use “wheels”. To check whether, e.g., 7 is a divisor, organize residues mod 7 in a circle, as on the right. An arithmetic progression may be imagined as a sequence of jumps of equal length between the positions on the wheel.



The claims of the preceding section can now be restated as:

This sequence of jumps reaches all the position on the wheel if and only if the length of the jumps is mutually prime with the size of the wheel.

³ We introduce polynomial sequences by investigating the differences of differences of differences (etc.) and eventually finding a sequence of 0s.

⁴ For example, for $n = 10$, if a is odd, then the last digit of P_k would go through all the digits (unless the last digit of a is 5). In particular, 0 appears as the last digit of P_k .

Remark 1: We want to warn the reader that the wheels drawn below, when we discuss sequences of degree 2, are going to be of *very* different nature. The rest of this section is, essentially, a sneak preview of this digest. It is written very cursorily, and it may be wise to skip it at the first reading.

Above, we took a particular number n (the “potential divisor”); to answer the question: “can it be a divisor of numbers of a 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_k . If one of these is $0 \bmod n$, then the answer for the number n is **Yes**.

This may be summarized as the first row in

	Size	Take positions of:	Look at:
n -wheel	n	Numbers P_k	Reaching the position $0 \bmod n$
Conductor wheel	The conductor c	Potential divisors n	The color of the position

On the other hand, for P of degree 2, we are going to work with “the *conductor* wheels”. There is going to be one such wheel per a sequence P_k ,⁵ its size is called the *conductor* c of the sequence.

Assume that P is already fixed. After finding⁶ the corresponding conductor c , the positions on the conductor wheel are residues mod c , and we consider such residues not of the numbers in the sequence, but of “potential divisors” n . Moreover, the positions on the wheel are going to be colored, and the color of the residue of a *prime* “potential divisors” would determine whether the answer is **Yes** or **No**.

Essentially, this means that 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_k ?”
is a periodic function of p .**

Several following sections provide examples clarifying this claim.

Remark 2: Consider this statement for $\deg P \leq 1$. Note that there is no need to exclude a finite number of exceptional values of p . Indeed, if p_1, \dots, p_r are the exceptions, then these prime numbers do not share their positions mod $p_1 \dots p_r$ with other prime numbers. Hence there is a periodic function with period of length $p_1 \dots p_r$ which takes one value (**Yes** or **No**) at these prime numbers, and the other value at the remaining prime numbers.

Remark 3: However, if we allow a finite number of exceptions, then the laws of the preceding section 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.⁷ However, for irreducible polynomials of $\deg = 2$, the “reduced” conductor is always greater than 1.

Example in $\deg = 2$: pizza numbers

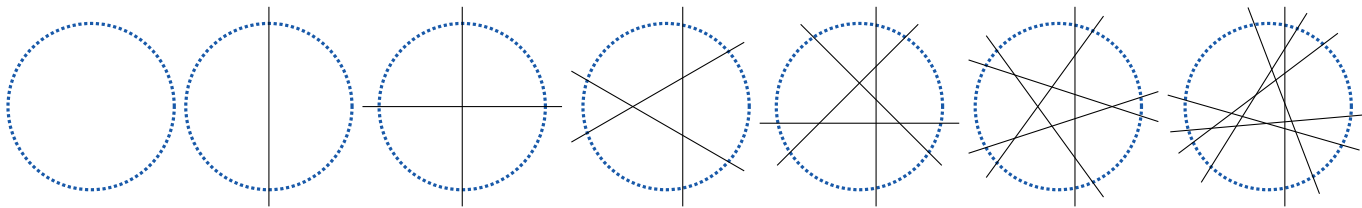
What we are going to discuss here is one of the most dramatic discoveries in arithmetic. The typical expositions try to play this dramatic aspect down; while we cannot have the Chorus singing “Something is going to happen! Just you wight, ’enry ’iggins, just you wight!”, we want to start with a sequence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

⁵Unless we are going to distinguish the “conductor” and the “level”, as we do on p.9. Compare with Footnote 14.

⁶While 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 degree 3), the calculation of these conductors may be very involved.

⁷In 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 14.)

Ⓜ Observe the largest number of pieces of pizza one can make with k 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
		$\xrightarrow{+1}$	$\xrightarrow{+2}$	$\xrightarrow{+3}$	$\xrightarrow{+4}$	$\xrightarrow{+5}$	$\xrightarrow{+6}$	$\xrightarrow{+7}$	$\xrightarrow{+8}$	$\xrightarrow{+9}$	$\xrightarrow{+10}$	$\xrightarrow{+11}$	$\xrightarrow{+12}$

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

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 k th cut meets the old cuts in at most $k - 1$ points. — However, there are only $k - 1$ old cuts!

(In fact, one also needs a bound on the other side: that $k - 1$ 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 $P_k - P_{k-1} = k - 1$, which leads to the formula $P_k = k(k + 1)/2 + 1$. 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_k ?

Ⓜ 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 numbers! 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, they form a palindromic sequence: $P_{-1-k} = P_k$. Because of this, the sequence $P_k \bmod n$ has a palindromic period.⁸ Hence if numbers P_0, P_1, \dots, P_m are not divisible by n , then numbers $P_{-1}, P_{-2}, \dots, P_{-1-m}$ are also not divisible by n . If $2 + 2m$ is at least as long as the period of $P_n \bmod n$, we can also conclude that no number P_k would be divisible by n .

⁸In particular, this means that the group of symmetries of this sequence is larger than translations by $n\mathbb{Z}$. It is the infinite dihedral group.

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 ...

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 simpler⁹ than this one.)

Ⓜ It turns out that a noticeable proportion of people cannot see the pattern in the table below. 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.

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

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

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			

What does it mean for a function of a prime number to be periodic?!

The pattern above suggests the answer: 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 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 $\bullet\bullet\bullet\bullet\bullet\bullet\bullet$ (of length 7):

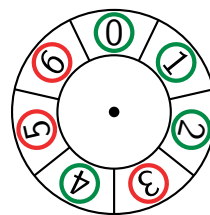
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 ...
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 ...

⁹Here the measure of simplicity is the number of necessary columns in the tables below.

As you can see, these sequences match at prime numbers (marked bold)!¹⁰

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 $0 \bmod p$ ”), it is enough to know $p \bmod 7$ (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. 5:


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

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.

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. 4).

Recall that for pizza numbers, after ignoring non-prime numbers, the red/green color pattern is fully controlled 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 $n^2 + 3$.  We use a handout with a table (attached at the end) representing the numbers in this sequence as products (similar to the table on p. 6) up to $n = 60$. Observing the divisors of $n^2 + 3$ in this table, one gets

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 above, one can show that no other number up to 60 can divide numbers $n^2 + 3$.

Note that with pizza numbers, we used different notations: we marked the divisors (as the numbers in list above are) in green, and the rest in red. This way, one gets the coloring:

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 ...

The former presentation is more compact, so reflects more data than the colored row.

Recall that the pattern we expect to see is:

- Select the prime numbers from the first list.
- Choose *an appropriate* size of a wheel (*the conductor*), and write numbers $1, \dots, 60$ in that many columns.¹¹
- 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.

¹⁰This 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 **2** below, on p. 11.

¹¹Now each column matches one of the positions on the wheel.

(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!

The prime numbers in the 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 ...

hence we arrive at the list of prime divisors

2 3 7 13 19 31 37 43 ...

What remains is to compare this with arrangements of natural numbers into several columns.

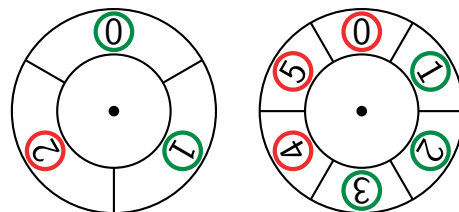
Ⓜ Below, instead of using green for divisors, and red for non-divisors, we use “circled” for (prime) divisors, and “uncircled” for (prime) non-divisors:¹²

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

One can immediately see which of the tables matches: the table with 6 columns. The columns 1, 2, 3 of this table have exactly the same bold numbers as the primes in 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 columns 1 and 3 match all the prime numbers in the list above—with one exception **2**.

This leads to these conductor wheels:

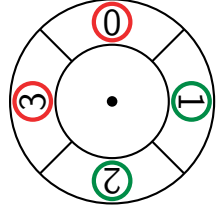


¹²Ⓜ This change of notation allows us to make this into a problem for kids: we give the students the tables (with nothing circled yet), and ask them to circle the numbers from the list above, then find which of the tables lead to “observable patterns” (similar to those discussed above).

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.¹⁵ (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.)

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

✗, 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 ✗. With “squares + 1”, we do not need to exclude anything.

Our next example is to investigate is “squares − 2”. (M) As above, we start with a handout representing these numbers as products (up to $60^2 - 2$). This gives the list of divisors:

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

Marking the prime numbers gives

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

leading to

2 **7** **17** **23** **31** **41** **47** ...

¹⁵However, these observations are just guesses: they do not show that the observed patterns would continue forever.

With small number of columns, this leads to

[illegible]

None of these table looks like what we want! With more columns, one can see

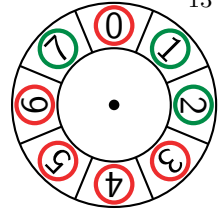
1	②	3	4	5	6	⑦
8	9	10	11	12	13	14
15	16	①⑦	18	19	20	21
22	②③	24	25	26	27	28
29	30	③①	32	33	34	35
36	③⑦	38	39	40	④①	42
43	44	45	46	④⑦	48	49
50	51	52	⑤③	54	55	56
57	58	⑤⑨	60	⑥①	62	63

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

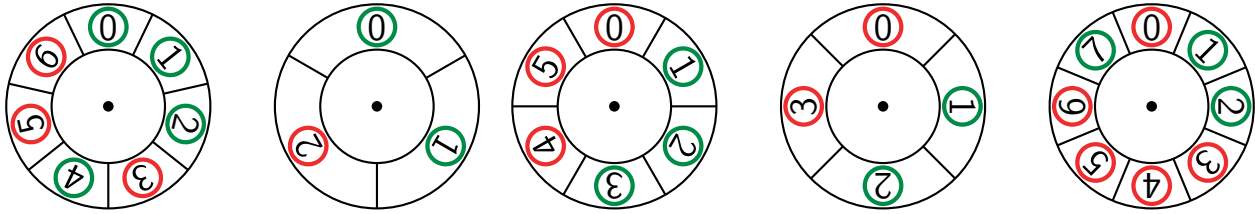
For 3, 4, 5, 6, 7, 9 and 10 columns, we see that a lot of columns contain both circled and non-circled prime numbers. This is not what we want to see.¹⁶ On the other hand, with 8 columns we see exactly the same pattern as expected (and we do not even need “a small number of exceptions”): columns of 1, 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, the 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.)

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 color¹⁸ 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!

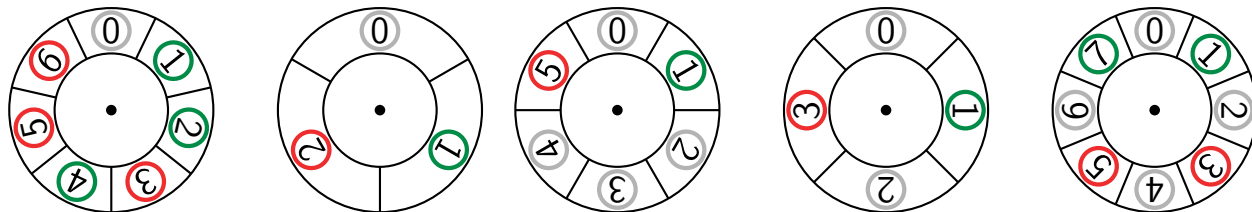
¹⁶However, these tables show the (relative) dearth of our data. Observe how the table with 5 columns contains no circled numbers in the column of 4. If this continues forever, this would be at least a *partial match* with the pattern we expect (the *full match* would be all columns having “all bold numbers circled”, or “all bold numbers uncircled”; but already one column with this property is something “very interesting”¹⁷).

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!

¹⁷In fact, for sequence of degree 3, such “partial matches” do actually appear (see Footnote 31). 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.

¹⁸Well, recall that in some of the tables, instead of coloring prime numbers, we circled “the green ones”, and left “the red ones” uncircled.  This allowed a conversion of these tables into problems for kids to solve.

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 $(\bullet)\bullet\bullet\bullet\bullet\bullet$, while the other colorings lead to anti-palindromic periods $(\bullet)\bullet\bullet\bullet\bullet\bullet$, or $(\bullet)\bullet\bullet$, or $(\bullet)\bullet\bullet\bullet\bullet$, or $(\bullet)\bullet\bullet$. Here we assume that $-\bullet = \bullet$, and $-\bullet = \bullet$, and put 0 in parentheses to make this symmetry of the wheels more visible in this “linear” rendition.

Euler’s formulation

To see what is common and what is different for our colorings of these wheels, note that the leftmost wheel (for pizza numbers) corresponds (see Remark 5 below) to prime divisors of “odd squares + 7” (which turn out to be the same as for “squares + 7”); the remaining wheels correspond to prime divisors of “squares + N ”, with $N = 3, 3, 1, -2$. Observe that the size of these wheels is $|N|$, or $|2N|$ or $|4N|$, while positive N lead to anti-palindromic colorings, and negative N to palindromic ones.

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 $\bmod |4N|$ is palindromic.
- (3) For $N > 0$ the pattern of answers $\bmod |4N|$ is anti-palindromic.

Here we use a special answer (“gray”) for residues not mutually prime with $|4N|$.¹⁹

The first two of these three conditions²⁰ constitute what is now known as “Euler’s formulation of Quadratic Reciprocity”, invented by Swiss/Russian/Prussian mathematician Euler. At his time, the proofs were known for $-5 \leq N \leq 4$ (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. (Note that 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 5: Similar questions about arbitrary polynomials of $\deg = 2$ can be reduced to questions about squares + N . For example, note that pizza numbers $P_n := n(n+1)/2 + 1$ satisfy $8P_n = (2n+1)^2 + 7$; hence the question about divisors of P_n can be rewritten as the question about divisors of $m^2 + 7$ for odd m . It is quite elementary that the latter questions has the same answer as “divisors of $m^2 + 7$ ”, and the Euler formulation implies that the colors have a period of length 28.



Remark 6: However, we saw above that the *observed* period in the case above has length 7. How to use the 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 (call it a *check-prime*), one knows the color of its position $p \bmod 28$ on the 28-wheel. This means that to find the period of length 28 (predicted by Euler’s formulation), it is enough to

¹⁹Note that a gray column may contain at most one prime number, so the first condition is trivial for such primes. Moreover, such prime $p \neq 2$ is a divisor of $0^2 + N$, so is of the “can divide” type. Likewise, $p = 2$ divides either $0^2 + N$ or $1^2 + N$.

²⁰The third condition turns out to be an immediate corollary of the periodicity for $N = 1$ and of top-multiplicativity (discussed below on p. 113). See Remark 71.

find check-primes for every (non-gray) position on 28-wheel. 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!

It turns out that the largest check-primes are 71 for the position $15 \bmod 28$ and 83 for the position $-1 \bmod 28$; 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. 7), 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. 7 would continue forever.

(We describe how to find the length of the shortest period in the general case in the section on p. 118.)

Remark 7: 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* the Legendre's formulation.²¹

On the other hand, as the (amazing!) progress in number theory in 20th century has shown, it is the Euler's formulation which has far-fetched generalizations.²² Moreover, nowadays we know that either one of these formulations is an almost immediate corollary of the other!²³

Summing up: in what follows, Legendre's formulation does not play any role. For people who are already fluent with the Legendre's formulation, to make it easier to switch the gear to the Euler's one we discuss interconnections about these formulations in the Appendix on p. 112. (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.)

²¹It 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 the Euler's formulation.

²²The 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 the 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.

²³Nowadays, the most useful application of Legendre's formulation is to prove the the Product Formula for Hilbert symbol — compare with the section on p. 122 and Footnote 231 on 84. However, this deduction is not immediate, so using Euler's formulation instead would not make the proof much harder.)


Degree 3: a coarse-grained approach (Grades 3–4, Ilya 2018-05)

This part completes the digest of what we did in our Math Circles (this time, just in Grades 3–4). It is still OK to read this “diagonally”. Here we only *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

In the investigation of divisors of polynomials of degree 2, one could restrict attention to polynomials $x^2 + N$ (see Remark 5). 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 case, “when N crosses the boundary $N = 0$ ”, there is a “phase transition”: the symmetry of the answer makes a sudden change.

Likewise, the cases of larger degree break into two similar “regions”, and “crossing a boundary” between these regions leads to a dramatic change in the answer. 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 indecomposable polynomials with 3 real roots, and indecomposable polynomials with 1 real root.


This immediately rejects the “obvious first choice” of the family “cubes + N ”.²⁴  For our students, triangular numbers are as natural as squares, and, with 3D shapes, tetrahedral numbers are as natural as cubes. For them, using the sequence “tetrahedral numbers + N ” with $N = 1$ is a natural counterpart of “triangular numbers + 1”. Unfortunately, this polynomial is decomposable (it has a root when side = -3); hence every prime number is going to be a divisor. So one should²⁵ use a different value of N , for example, $N = 2$.²⁶

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 above²⁷), 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. Among the latter, several have an abelian Galois group.²⁸ This gives a rich enough zoo of polynomials of degree 3, which is quite sufficient for our purposes.²⁹

²⁴This 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.

²⁵While 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. 60.

²⁶ 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).

²⁷For example, in Footnote 13 on p. 10.

²⁸Compare with Footnote 100 on p. 45.

²⁹See also the section on p. 122.

Example: Divisors of “tetrahedral numbers + 2”

Proceeding as 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 $1, 2, 3, 4, 6, 8, 9, 11, 12, \dots$. 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).³⁰ 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.³¹

Since for about 20 years now we know the actual pattern of the colors, 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!³²

³⁰For pizza numbers, we found (see Observation on p. 6) 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.

³¹While 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! Compare with Remark 42).

With 971 columns, about half of them would contain only green primes. However, the remaining columns would not be “all red”, even for primes—every such column would contain a mix of red and green primes (the mix happens to be in “proportion” 2-to-1).

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

Describe the pattern of colors *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. 66). We discuss this in more details in Remark 43 on p. 52.

(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 12 on p. 23.)

³²In 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.


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 what we do on p. 45.


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

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 recently, there was no clue whether the color sequences 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 and tables (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 symmetries we saw in red/green coloring of prime numbers in the case of a polynomial of degree 2 — *periodicity* and *mirror (anti)symmetry*³³ — are replaced by “hidden symmetries”.

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

- Wheels;
- symmetric tessellations (or “tilings”);
- curved geometries,
- working with infinities,
- fractals,
- harmonies, harmonics and waves (“Harmonic Analysis”),
- heat propagation.

Similar to the case of degree 2, there is a particular size of the wheel (the *conductor*) which is related to a particular sequence. However, it controls the sequence not directly, but by selecting a particular “size”³⁴ of a tessellation of a curved geometry (as mentioned above) in which we consider the heat propagation.³⁵

³³In other words: (anti)palindromicity of the period.

³⁴Note 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.

³⁵An 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 14 on p. 24.)

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!³⁶

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 4.³⁷ In particular, this leads to a proof of Langlands’ pattern for our sequence of colors for “tetrahedral numbers + 2”.³⁸

Ⓜ 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.

³⁶After finding the conductor, the Langlands program leads to a recipe describing certain integrals (see Remark 9 on p. 22). 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. 45).

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.

³⁷They also cover polynomials of degree 5 if the discriminant is a perfect square.

³⁸When 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 “ $20 \times \text{tetrahedral numbers} + 1$.” See Remark 12 on p. 23.

The simplest Langlands' patterns in more detail

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

We did not discuss what follows at Math Circles.

On p. 18 we gave very vague hints about what one should be fluent with to be able to understand the Langlands' patterns for our sequence of colors encoding divisors of “tetrahedral numbers + 2” (on p. 17). 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 12 on p. 23). 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.³⁹) Moreover, a few prime numbers⁴⁰ 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 .⁴¹ We discuss this in more detail in the section “Transliteration rules” (starting on p. 45).

- Denote the resulting sequence of numbers by N_n . The second step is to take the Fourier transform of this sequence. This is, automatically, a periodic function $F(t) = \sum_n N_n \cos nt$.⁴²

³⁹So, 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 34 on p. 48 for details.

⁴⁰Divisors of the discriminant, of the denominators of coefficients, and of the numerator of the leading coefficient.

⁴¹For 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. 7.

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 (but it is more complicated for N_n for composite n). 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” \rightarrow Numbers $N_n \rightarrow$ “colors”. It turns out that it replaces our row of colors by a periodic row of colors (see p. 7). 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. 113.)

⁴²A 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. 68 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 $F(t)$:⁴³

The graph of $F(t)$ 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.⁴⁵ Here is an example of such a fractal behavior of a graph from an [about-15-years-old paper](#):

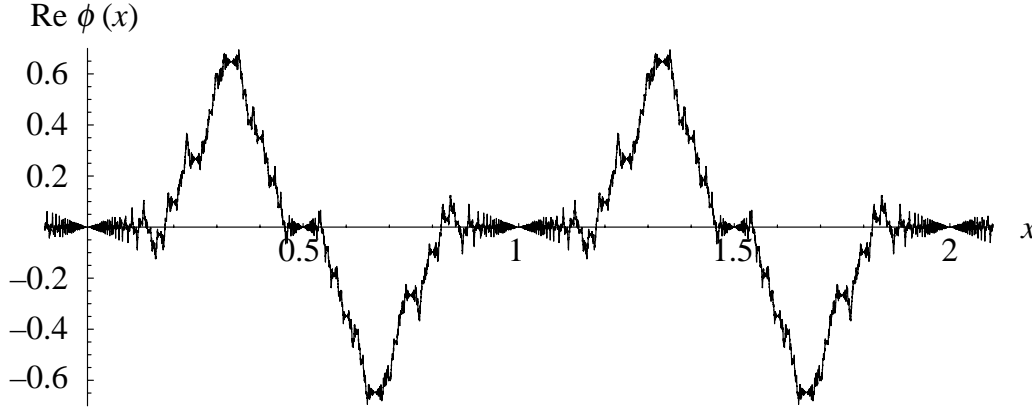


Figure 1. The real part of the antiderivative $\phi(x)$ of the automorphic distribution corresponding to the Mass form for $\mathrm{SL}(2, \mathbb{Z})$ with $\lambda \approx 27.56 i$.

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 8: 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

The fractality laws and the first few numbers N_n determine all the numbers N_n .

⁴³This 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.⁴⁴

⁴⁴This started with the [circle method](#) of Hardy–Littlewood.

⁴⁵The corresponding law is the red-framed formula in [Footnote 65](#) on p. 29.

And this is what actually happens!⁴⁶ Moreover, this is exactly what one expects from “a sequence having a pattern”: knowing “the type of the pattern”⁴⁷ 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 starting on p. 71.

Remark 9: 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.⁴⁸

Remark 10: 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”.⁴⁹ Remark 9 provides one of the most striking examples of such a unity.

Remark 11: 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 interconnecting⁵⁰ the passes by cross-references.

⁴⁶After 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!

⁴⁷For 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.

⁴⁸Moreover, the famous circle method of Hardy–Littlewood is based on a very similar observation. Compare with Footnote 43 on p. 21.

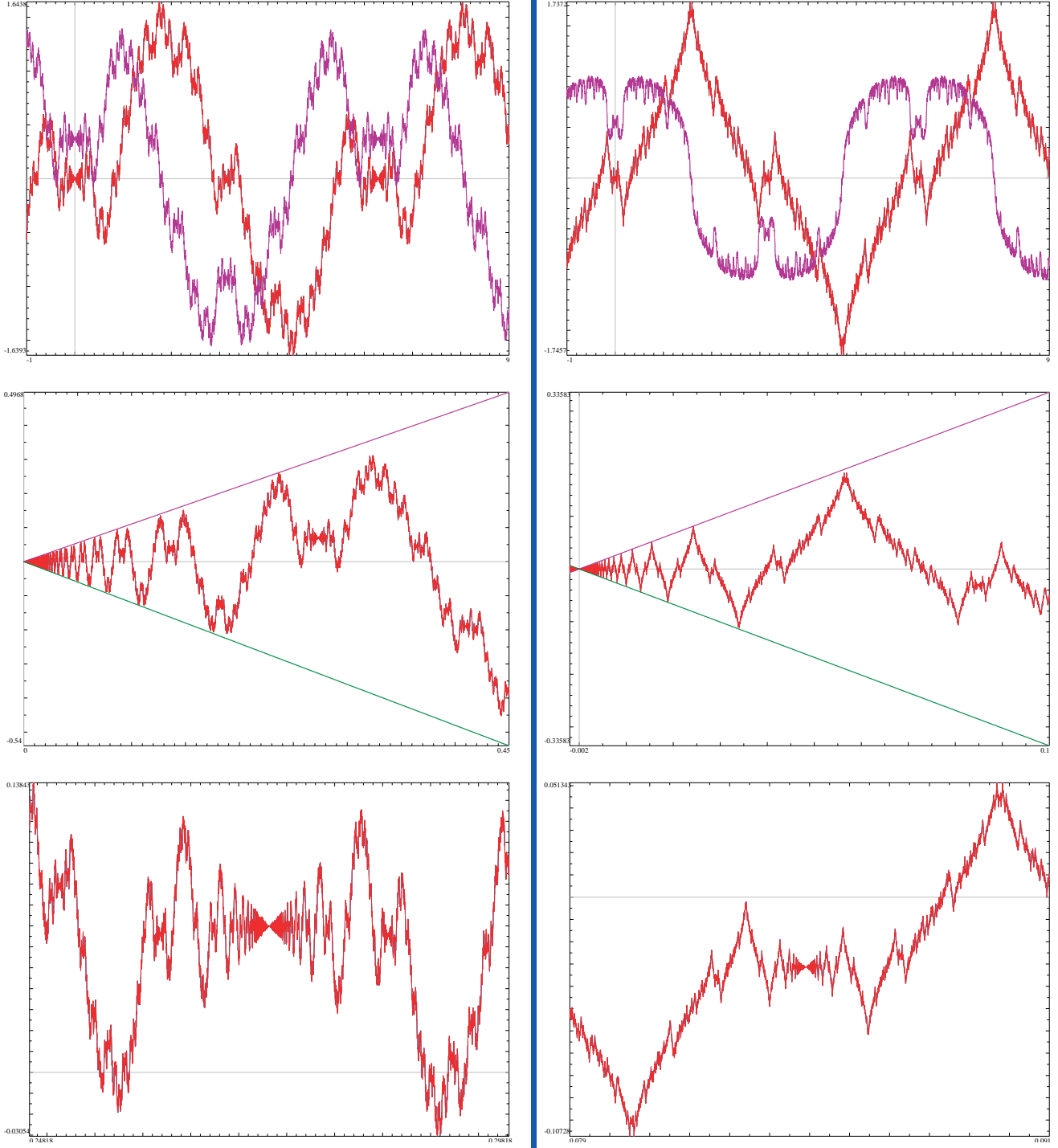
⁴⁹... although for most mathematicians, maturing to this idea takes much longer than it took I. M. Gelfand in an epigraph to these notes!

⁵⁰... 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!

Remark 12: 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.⁵¹ Each plot in the top row shows two graphs: about $1\frac{1}{2}$ periods for the real and the imaginary part of the function $F_{\mathbb{C}}^{(-1)}(t)$ (see Footnote 42 on p. 20). The

⁵¹Mathematically, 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”).⁵²

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.⁵³ We examine another approach to these two cases in Remark 18. See also Remark 30, and the section on p. 66.

Remark 13: In the outline above, we needed to cheat to circumvent certain delicate points. Note that the graph above, on p. 21, plots not the function $F(t)$, but its antiderivative $F^{(-1)}(t)$. (Same for plots of Remark 12.)

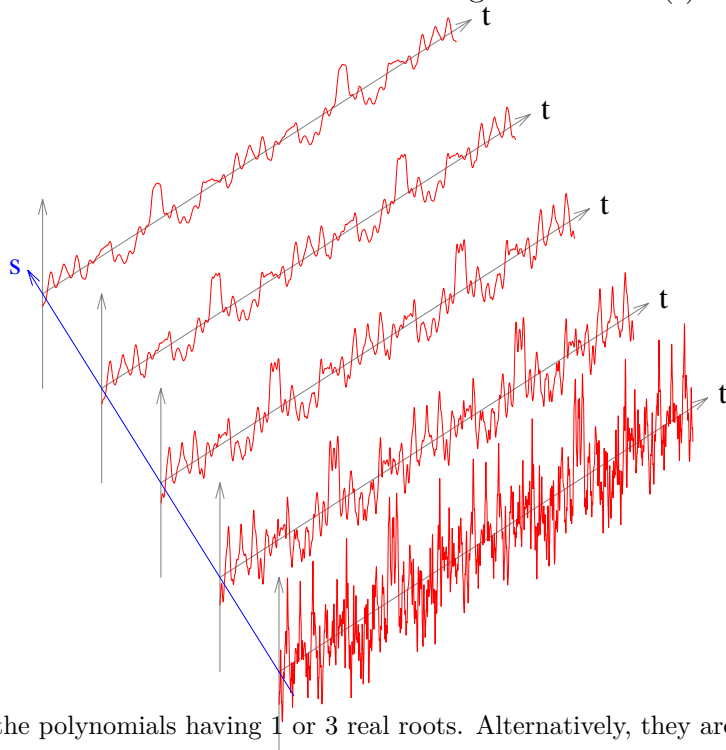
The reason for this is that, in a naive sense, the function $F(t)$ has no value anywhere: the Fourier series defining $F(t)$ diverges for every value of t . In particular, the graph of $F(t)$ itself does not make a lot of sense. However, the antiderivative of $F(t)$ has “a much milder” Fourier series; and it has an honestly defined graph. (Note how this is similar to the relation of “white noise” and “Brownian motion” — see Remark 26 on p. 35.)

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 14: 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 8. (See the section on p. 35 for details.)

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

Remark 15: The other approach “regularizes” the infinities away altogether. Here “regularization” means a particular way to



⁵²In 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.

⁵³This 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 Section on p. 68.

(In fact, these pictures⁵⁴ 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 26 on p. 35.)

Remark 16: 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

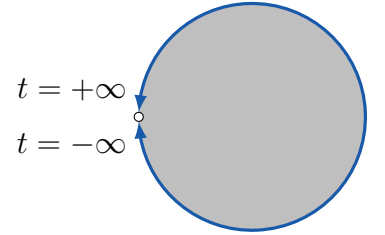
- If one chooses a “suitable” way to regularize, and
- if one chooses a “suitable” curved 2-dimensional geometry,

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:

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

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 starting on p. 68.)

Remark 17: In the second approach, the domain of definition of the function $F(t)$ becomes “the absolute”, or the “horizon” of the curved geometry.⁵⁵ 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 $(-\infty, \infty)$).⁵⁶ In Remark 16, we worked with a point of Lobachevsky geometry writing it as (t, s) with $s > 0$. This is the half-plane model of this geometry; however, there is another, equally useful model



⁵⁴For 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 $[0.015 \dots 0.095]$, while t changes in $[0 \dots 14]$.

In fact, the scales of variables s and t are closely interconnected (see Remark 16). This means that we scaled s up about 200 times. So the plots show what happens in a *very narrow strip* near the line $\{s = 0\}$.

⁵⁵A 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.

⁵⁶People familiar with the *Stereographic Projection* may recognize the significance of this formula.

in a disk,⁵⁷ 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 16.⁵⁸ Moreover,

Regularization $F(t) \rightarrow f(t, s)$ 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. Hence

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 16.

Remark 18: 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.⁵⁹

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 12. 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.⁶⁰ We illustrate this in section “Maass fractality laws” starting on p. 41.

⁵⁷There is no best way to visualize this curved geometry. Sometimes the half-plane model $((t, s)$ with $s > 0$) used in Remark 16 is more convenient; sometimes the disk model.

⁵⁸We illustrate these coordinates on p. 69.

⁵⁹This 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 appearance 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.

⁶⁰In fact, this is how these forms were first discovered: not on the absolute, but on the Lobachevsky plane.

Remark 19: 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 20: For example, the plot above (on p.21) *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).⁶¹ 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.

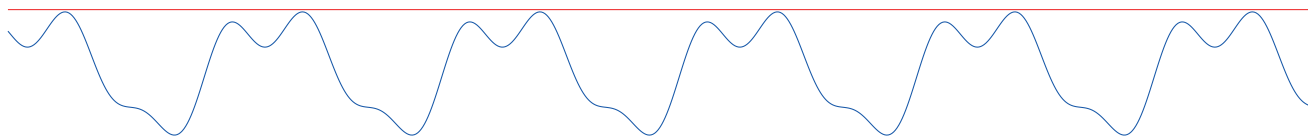
Unfortunately, the usual way the Langlands program is stated is extremely technical and very far removed from the simplified point of view discussed above. Translation to down-to-earth terms is error-prone if one is not a specialist; on the other hand, there are very few published attempts to do this—and 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 12.)

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.

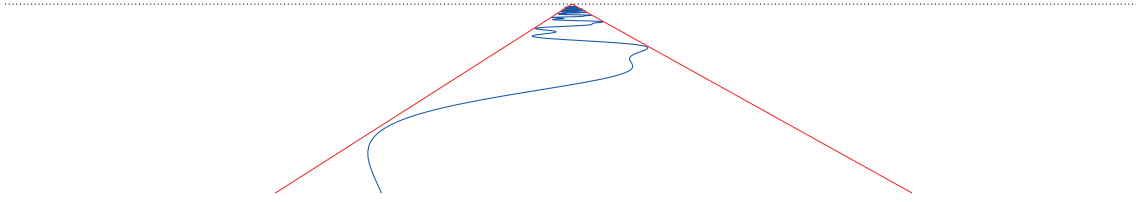
Take a graph of a periodic function $g(t)$:



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

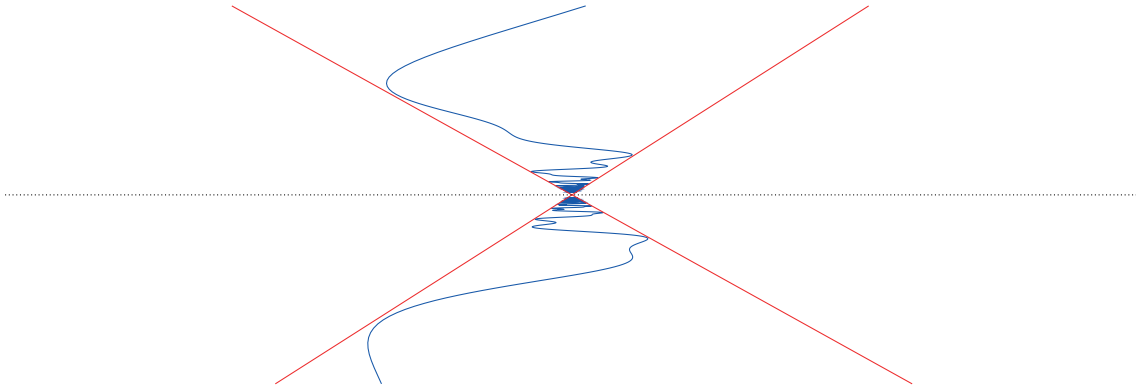
⁶¹This number is related to the eigenvalue for this eigenfunction of the heat transfer operator.

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.

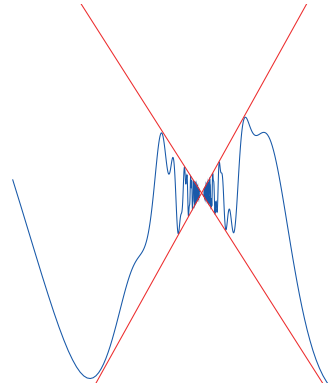
As it is customary done in “Projective Geometry”, above the horizon we put the reflected picture of what is “behind us”:



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.)⁶² 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 $tF(-1/t)$).⁶³



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 $g_P(t)$.⁶⁴

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

⁶²Moving our “observation point”, one can also get functions $tG(-1/t)$ with $G(t) = Ag(Bt + C)$.

⁶³What 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 74 on p. 34.)

⁶⁴In 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*”.⁶⁵

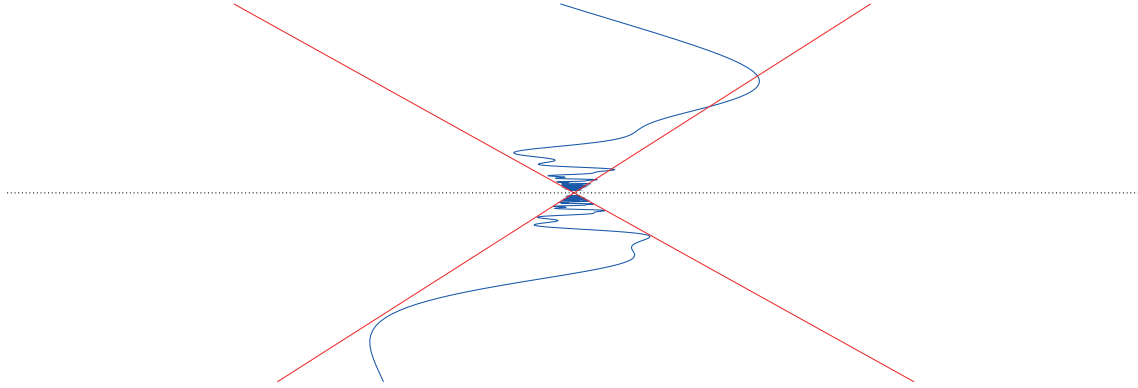
Remark 21: Above, what we did “above the horizon” looks very logical—provided one knows projective geometry. Indeed, when we look in some direction, our gaze “hits” everything on the half-line starting at our pupil, and going in the direction we look at. Now, half-lines are not very natural geometric 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 12). 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 12). The contrast between having t and $|t|$ leads to the difference of the graphs in Remark 12: the right one needs $|t|$.

⁶⁵Using the formula from Footnote 63, horizon-similarity “to $G(t)$ ” at $t = 0$ means:

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

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

$$tF(t + 2\pi^R/s) = \varepsilon \cdot F(\zeta - 1/\gamma t).$$

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. 53.)

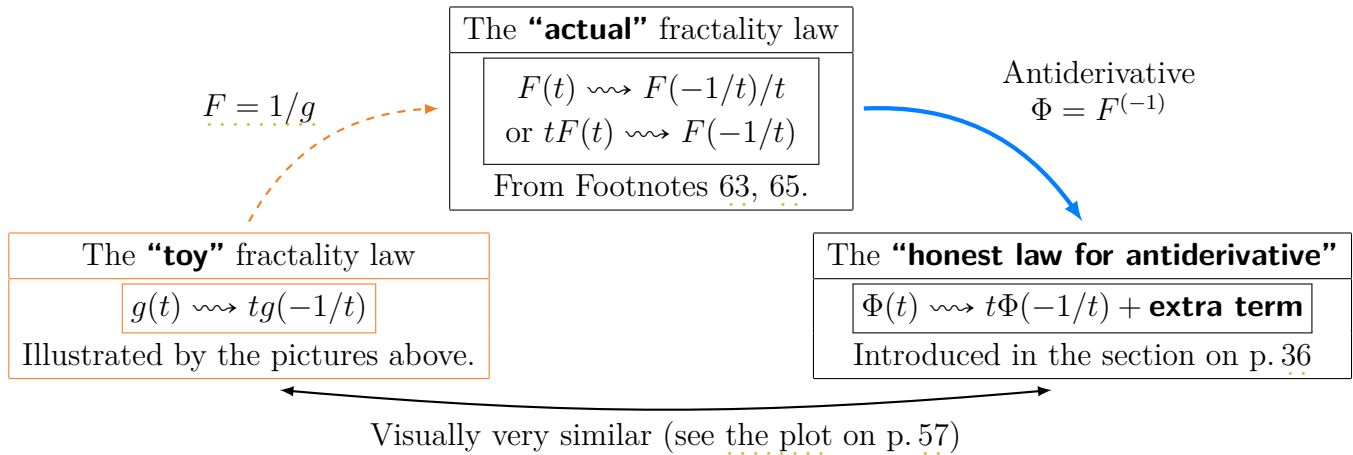
Below, we illustrate the notions of “horizon-similar” and “horizon-self-similar” with many plots.⁶⁶

We quantify the notion of “many such points P ” in the sections on p. 54, p. 74, p. 80. Moreover, the possible values of R and S —and the corresponding ζ and γ —are described in Footnote 132 on p. 54.

⁶⁶However, 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. 29.)

The zoo of fractality laws

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



(The **extra term** comes from integration by parts. See the calculation on p. 57 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.

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 63 on p. 28).⁶⁷

Recall (see the section on p. 20) 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. 35), 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 “The honest fractality law...” on p. 57 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. 28 before Remark 21 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.)

In other words:

The graph is a rescaled toy transformation of itself.

⁶⁷Recall 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 74 on p. 34.)

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:

Force the graph to be preserved by toy transform, and force 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.⁶⁸ 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.⁶⁹ 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 $G(t) := g_n(t) - ig_{n+1}(t)$ with $n \gg 0$ gives a function

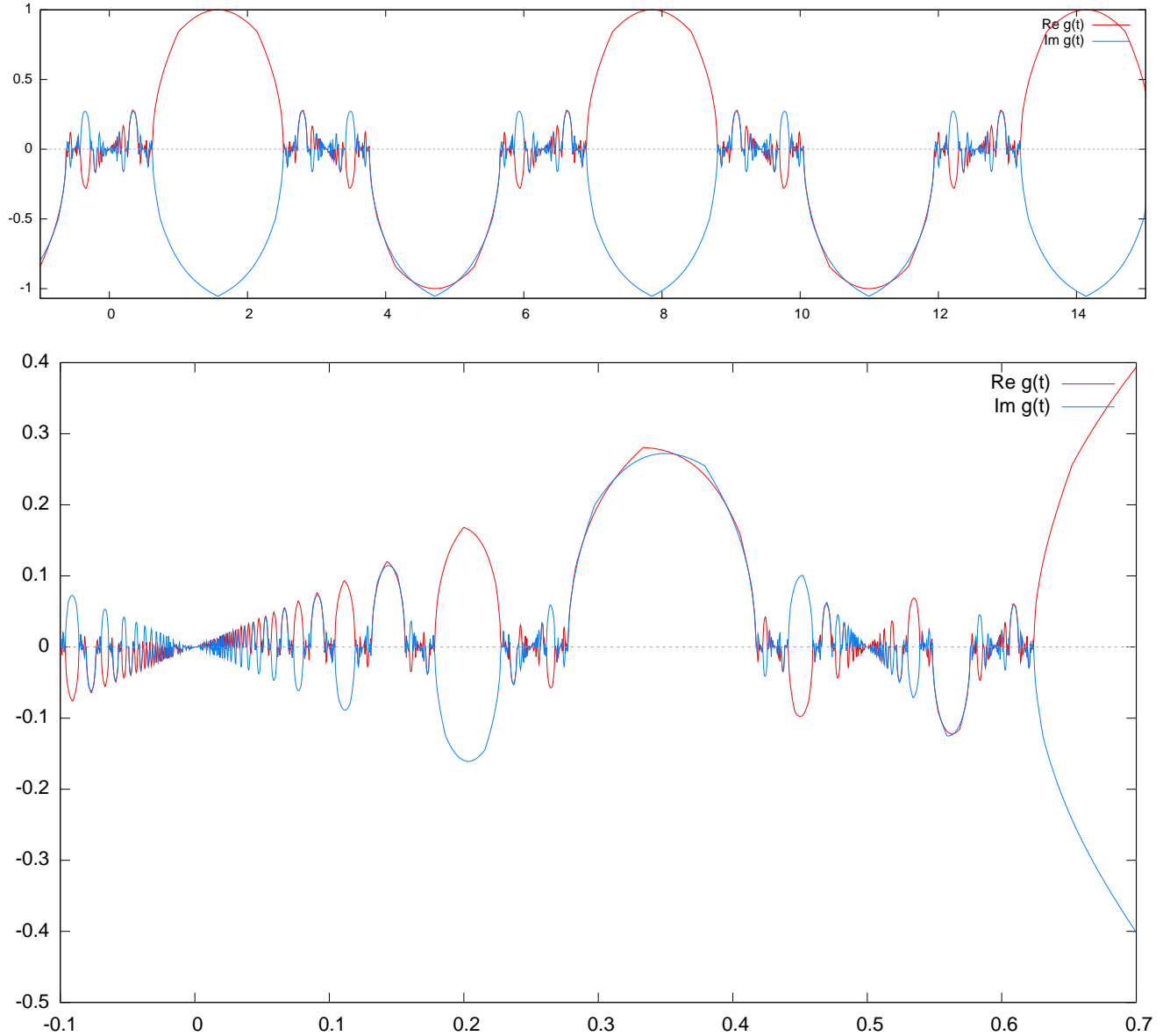
⁶⁸For continuity, it is better to start with $[-\pi/2, \pi/2]$, then extend it to $[-\pi/2, 3\pi/2]$ so that $f(t + \pi/2)$ is even. Then one can extend from $[-\pi/2, 3\pi/2]$ by 2π -periodicity.

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.)

⁶⁹However, 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 $iG(t)$; 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.⁷⁰

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.28 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”

⁷⁰It 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, appearance of i is inevitable since the toy transform chained with itself sends $G(t)$ to $-G(t)$.)

with oscillating behavior near the origin.⁷¹ 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*.⁷²

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 22: 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

⁷¹This is yet more pronounced when the “toy transformation” $tg(-1/t)$ is replaced by the “actual transformation” $g(-1/t)/t$.

⁷²The 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. 29) can move these points off the t -axis.

Remark 23: Since the non-smooth points of the graph form a fractal, “for most⁷³ 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.⁷⁴

Remark 24: 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. 53.)

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. 28 before Remark 21).

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 45 on p. 54.

Remark 25: 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.⁷⁵ Moreover, the fractality laws determine F up to a finite number of degrees of freedom (compare with the discussion near Footnotes 46 and 47 on p. 22).

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).⁷⁶ Until Langlands, mathematicians wouldn’t suspect that there must be many more points of self-similarity,

⁷³... meaning: outside of a “meagre set of measure 0”.

⁷⁴In 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 63, 65 on p. 28.

⁷⁵In a certain very precise sense a positive fraction of the set of numbers $2\pi R/s$ are horizon-self-similar. Compare with Footnote 136 on p. 55.

⁷⁶In fact, he found another — equivalent — formulation. (In Footnote 90 on p. 40 we have a few more details.)

and that these laws would severely restrict how our red/green coloring (or numbers N_k ; see p. 20) may look like.⁷⁷

All the fractal transformations together: infinities and regularizations

Return back to the situation when “horizon-self-similar points”⁷⁸ 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 $\sum_n N_n e^{int}$ 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”⁷⁹, 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 14 on p. 24.

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 26: 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”.⁸⁰ 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.⁸¹

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

⁷⁷Indeed, 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.

⁷⁸These are defined on p. 28 before Remark 21.

⁷⁹In other words, “a function which may have no value at any particular point, but ‘weighted averages’ of these values still make perfect sense”.

⁸⁰Any 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”.

⁸¹Compare this with the plots in Remark 15 on p. 24. 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 27: 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.⁸² 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.⁸³

For example, compare the graph on p. 32 with the graph on p. 21. 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 [Section “The honest fractality law...”](#) on p. 57.)

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. 17). Recall that (see p. 20) 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”.

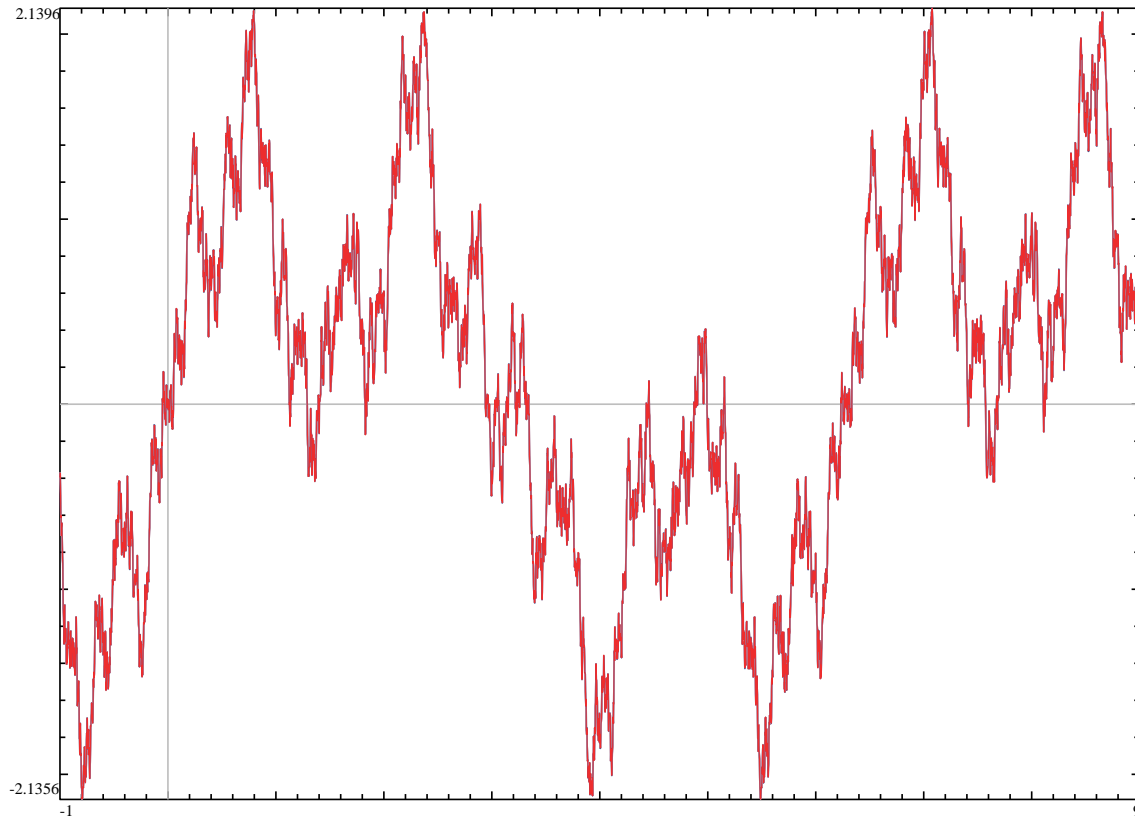
⁸²At least if one understands it as $tF(\text{const} \cdot t) = \text{const} \cdot F(-1/t)$, as in Footnote 65 on p. 29.

⁸³We discuss this extra term [below](#), on p. 57.

- However, in contrast to the “toy fractality law”, the “matching pieces” may be at different heights.

Essentially, these expectations are fully satisfied by the graph on p. 21. For example, near the origin this graph looks very similar to a “toy transform” of itself⁸⁴.

However, the actual graph⁸⁴ of the antiderivative $F^{(-1)}(t)$ (about $1\frac{1}{2}$ periods)

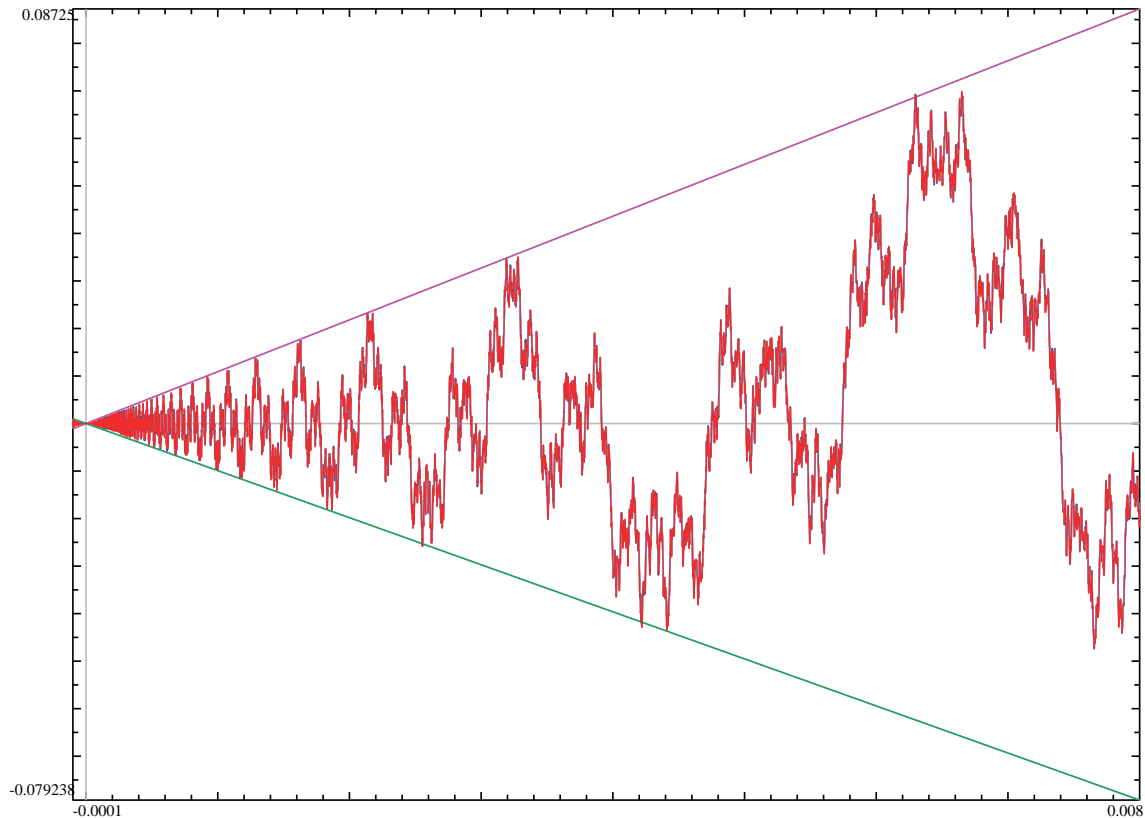


does not look this way — there is no piece similar to the toy transform of this periodic graph! What is the reason for this?

Answer: what is spoiling the fun in the graph above is the conductor! For the graph on p. 21, 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.

⁸⁴The 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, we need to zoom into the graph with a very strong magnification (about 971 times):



Now the pattern is clearly visible.⁸⁵ 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).⁸⁶

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. 28), 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 section on p. 54. For a heuristic estimate of zoom factors needed to expose the extent of fractality see Remark 62 on p. 84.)

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

⁸⁵Note 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 (4,000,000 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 5,000 samples).

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!

⁸⁶Additionally, 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 $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ rescaled. We return to this theme in Footnote 90 and in the section on p. 57.

be dealt with without Landlands program,⁸⁷ we would need to consider different sequences anyway. For example, for a fixed number M , one can consider the sequence⁸⁸ “ $M \times \text{tetrahedral numbers} + 1$ ”.

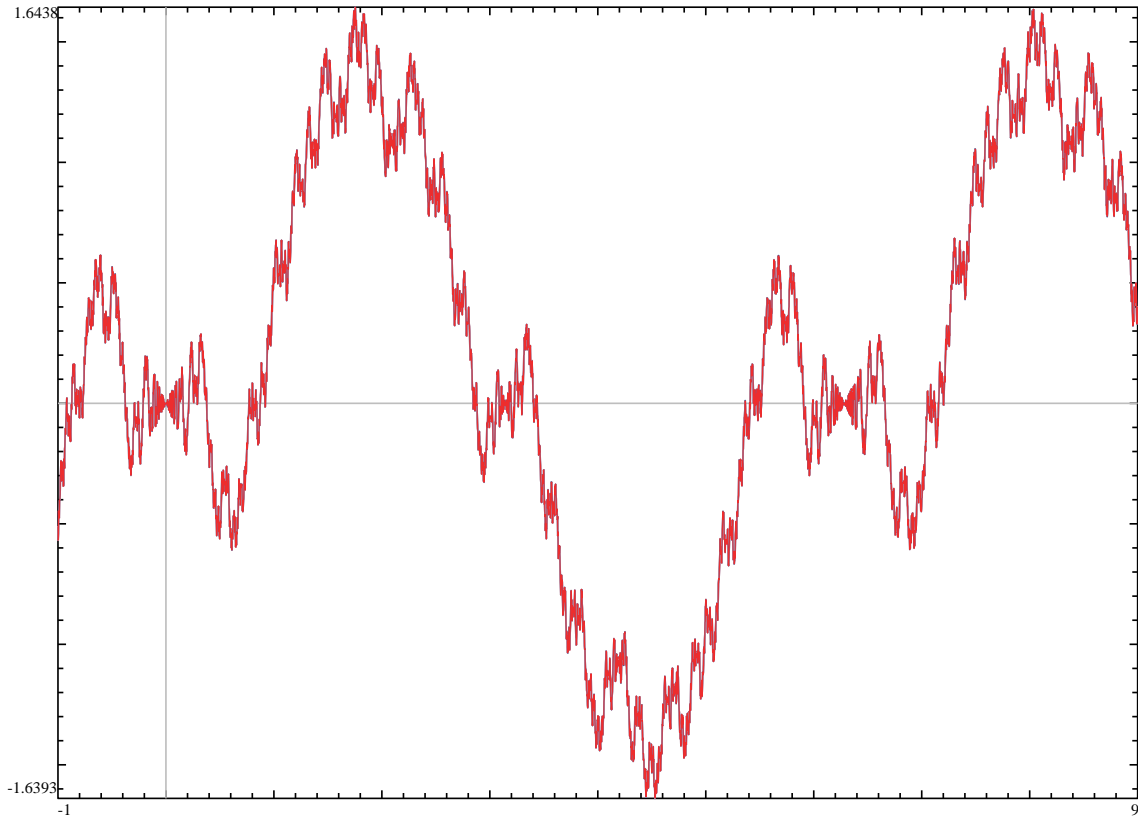
The difficulty encountered in the previous section is related to the fact that discriminants of polynomials of degree 3 tend to be quite large in magnitude (hence the conductors are also expected to be large). For the example of the preceding section, the discriminant is -4×971 . In fact, the smallest 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.⁸⁹

So let’s redo what we did above, starting with the polynomial “ $6 \times \text{tetrahedral numbers} + 1$ ”.

- Assign colors to numbers according to whether they can be divisors of “ $6 \times \text{tetrahedral numbers} + 1$ ”.
- Transliterate colors to numbers N_n (for details, see the section on p. 45).
- Take the Fourier transform $F(t)$ of the sequence N_n .
- Plot the antiderivative $F^{(-1)}(t)$

Here is the result (about $1\frac{1}{2}$ periods)



⁸⁷See the section on p. 66.

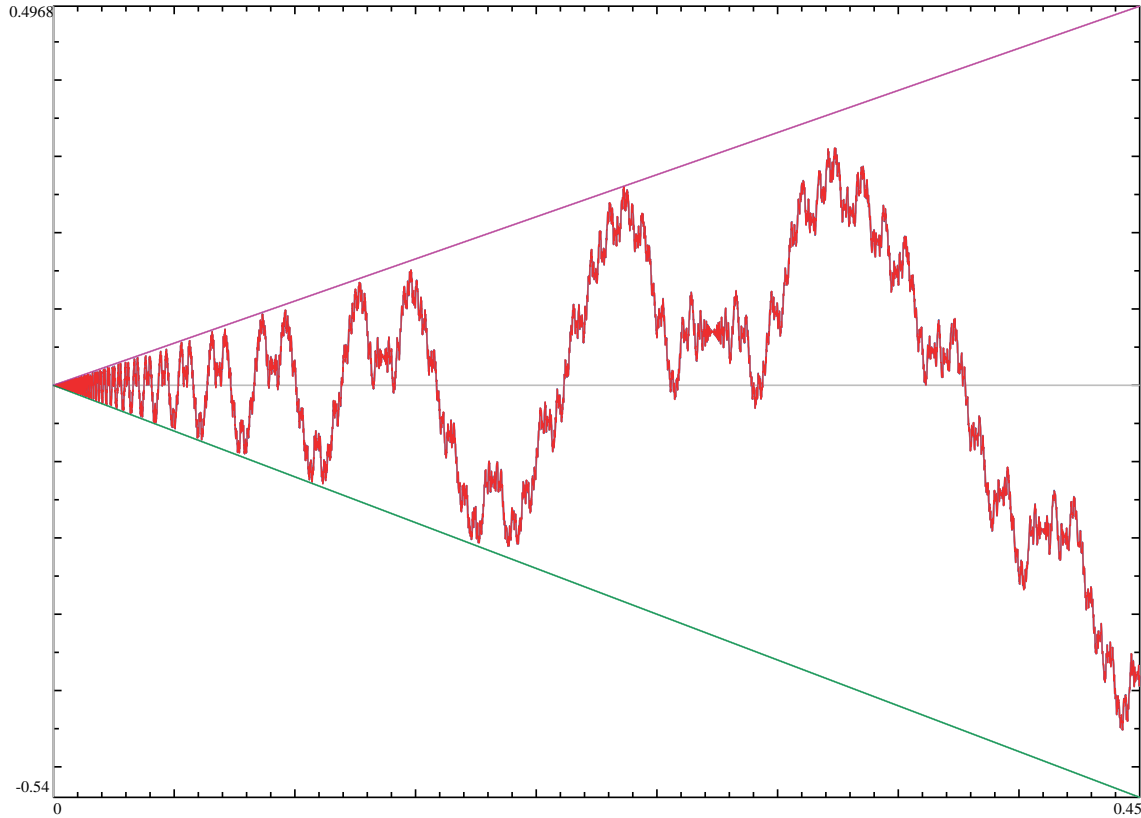
⁸⁸Note that doubling this sequence to become “ $2M \times \text{tetrahedral numbers} + 2$ ” leads to the same prime divisors, with a possible exception of 2. However, such an exclusion is “negligible”, since when matching the patterns of colors, we allow a few primes to be exceptional anyway (compare with 2 above, on p. 11).

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

⁸⁹Again, compare with the section on p. 66.

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:



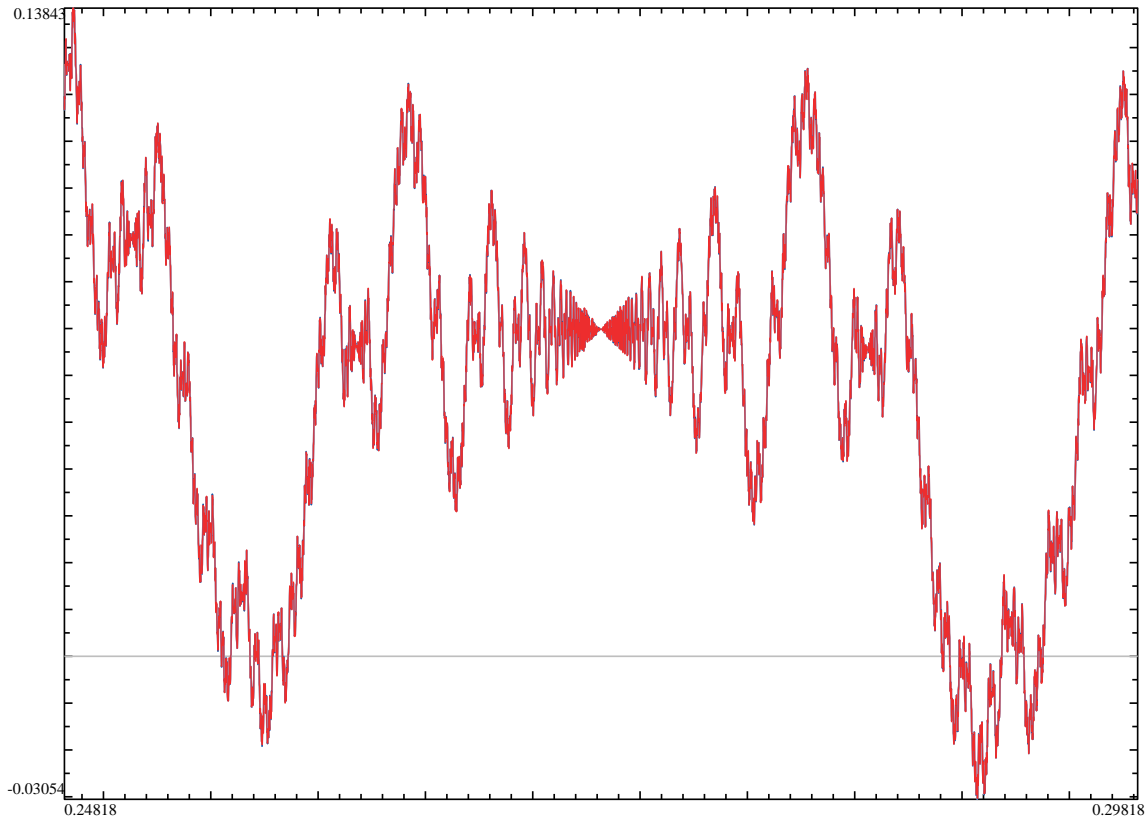
The pattern “a toy transform of a periodic function” is again clearly visible. In notations introduced on p. 28 before Remark 21, this periodic function is $(F^{(-1)})_0(t)$. Moreover, the same as in the previous section, comparison of two preceding graphs shows $(F^{(-1)})_0(t)$ is *different* from $F^{(-1)}(t)$. Again, the parts below the t -axis look very different from the parts above.⁹⁰

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

⁹⁰In 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 $F_0(t) = \text{Im } F_{\mathbb{C}}^{(-1)}(t)$. (Compare with the violet graph of $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ in the top-left plot of Remark 12.)

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. 66 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”⁹¹ point — defined on p. 28.)

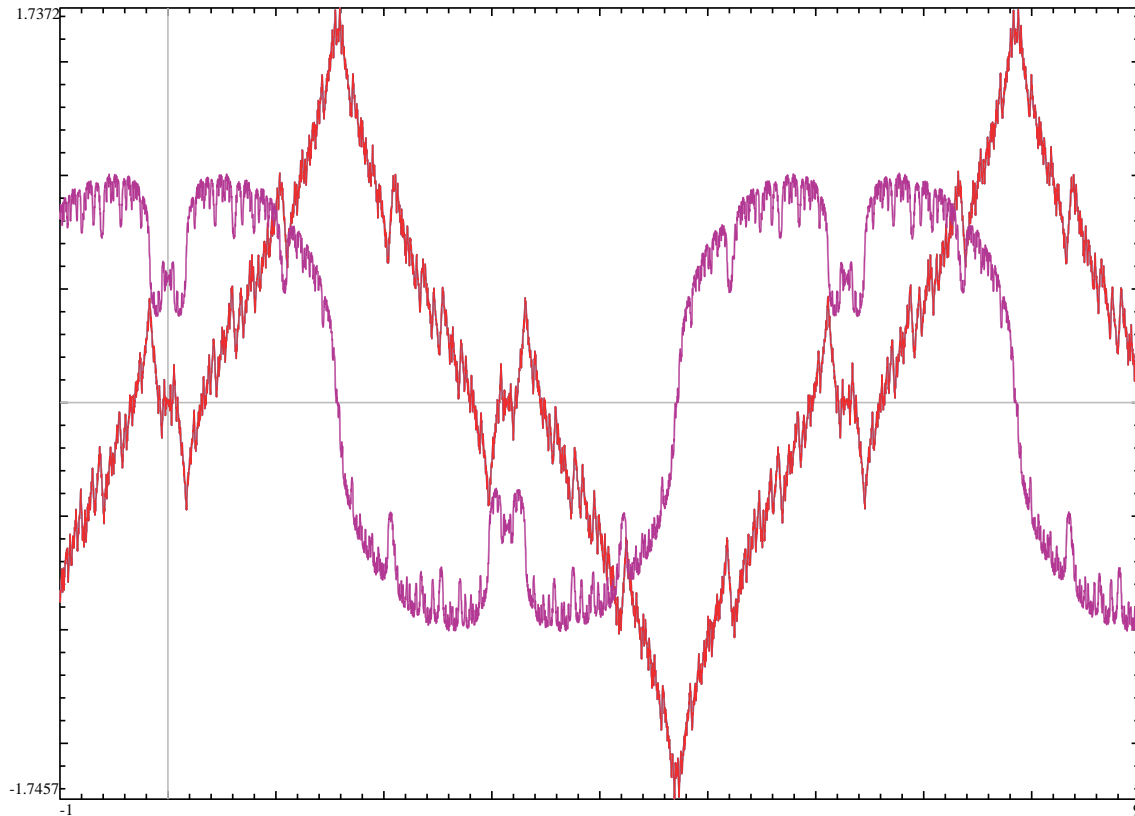
Maass fractality laws

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. 66). This happens for polynomials “ $M \times \text{tetrahedral numbers} + 1$ ” 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).

⁹¹With a correction that here we get not a “toy” transform, but the “honest” fractality law (see p. 57).

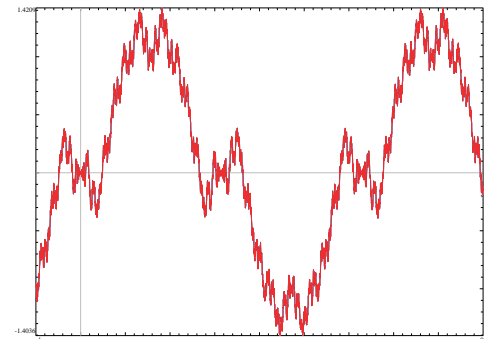
Fortunately, “an extra coincidence” happens, leading to extra symmetries in this case, which make zooming feasible:



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

This extra symmetry (which may be suspected from the factor 2^2 in the conductor $c = 2^2 \times 37 = 148$)⁹² makes our zooming factors behave as if this number was about 4 times smaller. This makes plotting feasible.⁹³

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.⁹⁴ (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. 11.⁹⁵)



⁹²This mirror symmetry is due to the transliteration rules for the prime 2 following the last case of Step (d) on p. 46. Because of this, $N_{2^k} = 0$ for any k , which implies $N_{2k} = 0$ for any k by the rule of Step (e).

⁹³For the graphs below, we used N_n for n up to 1, or 4, or 16 millions.

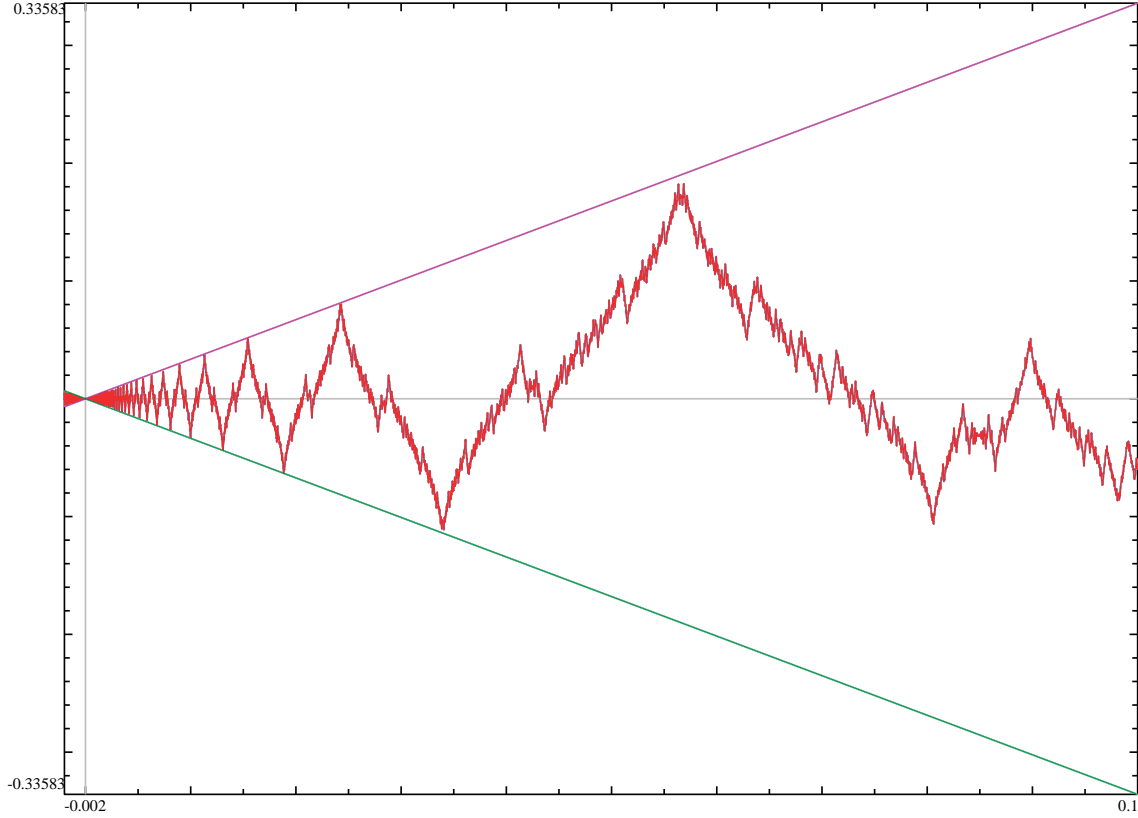
⁹⁴Recall that the case $M = 6$ is the cubic polynomials with negative discriminant, which are covered by the Class Field Theory (see p. 66) — so do not *need* the Langlands program.

⁹⁵On the other hand, Arnold’s *Principle of Fragility of Good Things*⁹⁶ 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 we cannot find such polynomials with conductors which are “large enough” for general polynomials.

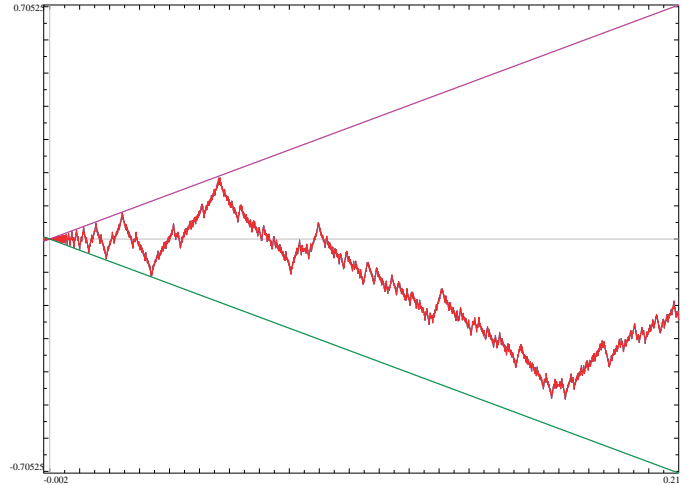
⁹⁶... referenced in Wikipedia article on “Anna Karenina principle”.

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

Recall what we saw in the left column of Remark 12 (discussed in the section on the case $M = 6$ on p. 38): 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 12): near $t = 0$ the graph is visually indistinguishable from the toy transform of *itself*:



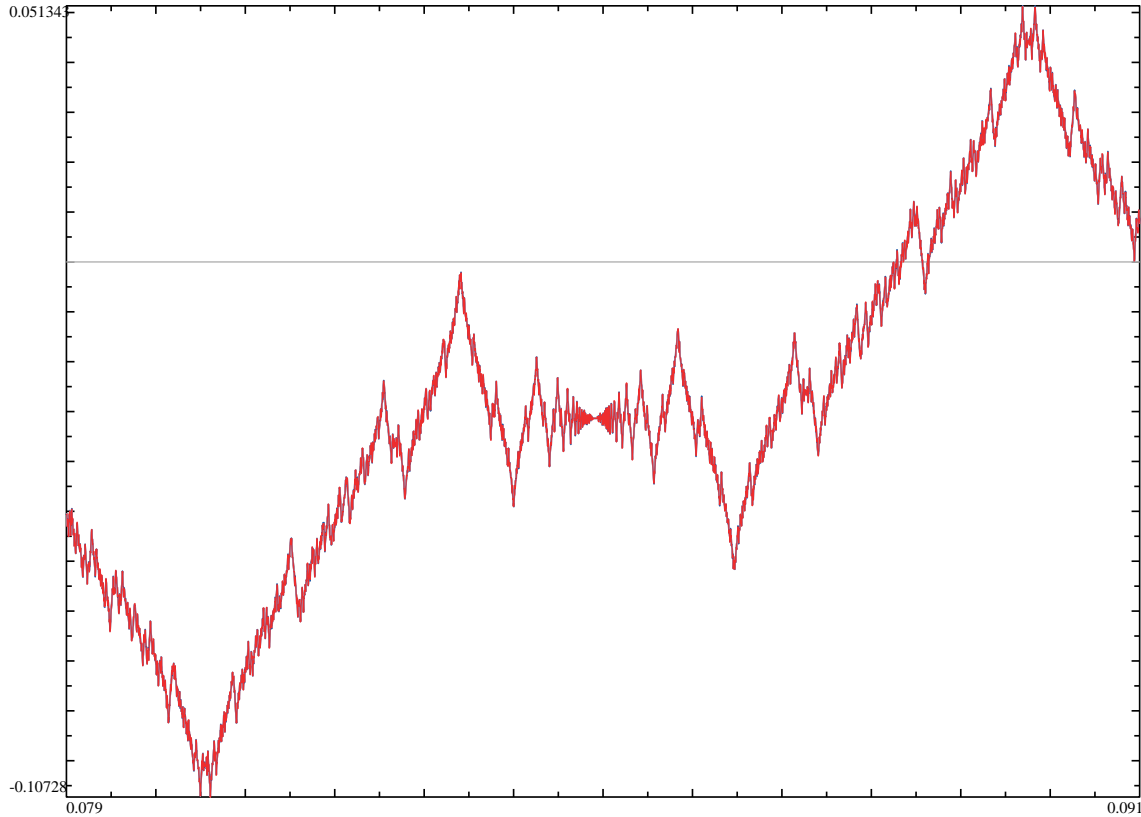
Remark 28: Note that the rightmost maximum, near the violet asymptote, is the transform of the maximum of the first red graph near $-\frac{3\pi}{2}$. In particular, we could have used 3 times smaller magnification so that the extended graph would also include the transform of the minimum at $-\frac{\pi}{2}$ (on the right). Unfortunately, the difference between the honest fractal transform and the toy transform becomes very large in such an extended domain⁹⁷. 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 47).



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

⁹⁷Compare with the section on p. 36, 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 29: Recall what we did in Remark 15: in the “odd” (“modular forms”) case, we would extend the function $F(t)$ to a function $f(z)$ on the upper-half plane (with $z = t + is$); this changes $F_{\mathbb{C}}(t) := \sum_n N_n e^{int}$ to $f(t, s) := \sum_n N_n e^{int - ns}$. In other words, we replaced N_n by $N_n R(ns)$ with $R(s) = e^{-s}$ being the “regularizing factor”. (See also the section on p. 68.)

Such a replacement turned out to be *compatible* with fractal transforms (in the sense of Remark 17). 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 $R(s) := e^{-s}$ used for the “modular forms” regularization, one should write $R(s) := \sqrt{|s|} K_0(|s|)$ with K_0 the Bessel function.⁹⁸ In particular, instead of looking at $f(t + is) := \sum_n N_n e^{int} e^{-ns}$, one should write $f(t + is) := \sum_n N_n \sqrt{|s|} e^{int} K_0(|ns|)$. 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: the main term in the asymptotic of $f(z)$ when $s = \text{Im } z \rightarrow 0$ is $F(t) \sqrt{s} \log s$. Moreover, taking “the trace” is compatible with Lobachevsky-moves (“intertwining”), which means that the fractal transforms of $F(t)$ are the traces of Lobachevsky-moves of $f(z)$.

This means that with this “‘intertwiningly’ 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

⁹⁸The precise form of K_0 is not important for our purposes.

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 18). Functions satisfying this condition are called *algebraic Maass forms*.

Remark 30: 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.⁹⁹ 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.¹⁰⁰

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 31: 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. 66.

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. 20):

- 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. In the next chapter, we try to demystify these rules — 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.¹⁰¹

- (a) In the sequences above (see example for degree 2 on p. 8 and one for degree 3 on p. 17), 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 $0 \bmod n$ appears in this period. If it appears, we mark the number n green, otherwise red.

⁹⁹Recall 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.

¹⁰⁰In 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. 66 and Remark 53 on p. 65.)

This happens when the discriminant is a complete square. For the polynomials of the type we consider here, $M \times \text{Tetrahedral Numbers} + 1$, 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 .)

¹⁰¹The 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.

For degrees higher than 2, and for several variables (as in the beginning of the section on p. 17), 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 $\widetilde{N}_n^{\text{res}}$ (for the residues mod n). For a polynomial of degree 3, for almost all prime numbers p the value $\widetilde{N}_p^{\text{res}}$ is 0, 1 or 3.¹⁰²

- (b) We obtained a sequence $(\widetilde{N}_n^{\text{res}})$ 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 “purify” this sequence a bit.

Recipe: Put $N_p := \widetilde{N}_p^{\text{res}} - 1$ for a prime number p (with exceptions from Footnote 102; we cover them in Step (d)).

- (c) Next, we need to define N_q for $q = p^k$ with a prime number p . **Recipe:**¹⁰³ for every prime p , choose one of the following sequences:

- $-1, 0, 1, -1, 0, 1, \dots$ (3-periodic);
- $0, 1, 0, 1, 0, 1, \dots$ (2-periodic);
- $2, 3, 4, 5, 6, 7, \dots$ (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 102), one cannot find N_{p^k} given $\widetilde{N}_p^{\text{res}}$ 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:

- $1, 1, 1, 1, 1, 1, \dots$ (1-periodic);
- $0, 0, 0, 0, 0, 0, \dots$ (1-periodic).

While there is a recipe explaining which of 5 variants to choose,¹⁰⁴ 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 $F(t)$.¹⁰⁵

- (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. 17. This shows that 11 is green, and 7 is red. So $N_7 = -1$; moreover, checking the table on p. 17 shows that 11 divides only one number of our sequence for sides $1, \dots, 11$ — which is the shortest period of our sequence mod 11. Hence $\widetilde{N}_{11}^{\text{res}} = 1$, and $N_{11} = 0$. Picking up a matching sequence above, $N_{11^2} = 1$ (the second number in the sequence $0, 1, 0, 1, 0, 1, \dots$), and $N_{7^4} = -1$ (the 4th number in the sequence $-1, 0, 1, -1, 0, 1, \dots$). Finally since say, $290,521 = 7^4 \times 11^2$, we conclude that $N_{290,521} = -1$.¹⁰⁶

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-prone¹⁰⁷).

¹⁰²The 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 .

¹⁰³See Remark 37 for more details.

¹⁰⁴See the section on p. 59.

¹⁰⁵Compare 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. 66 for details).

¹⁰⁶This illustrates that in general, whole numbers $|N_n|$ grow very slowly.

¹⁰⁷Compare with Footnote 105.

Remark 32: Replacing the sequence of colors by the sequence of counts $\widetilde{N}_n^{\text{res}}$ (as in Step (a)) 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 a 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!).

Remark 33: As we explained, for degree 3 and residues mod a prime number, the count $\widetilde{N}_n^{\text{res}}$ may be 0, 1 or 3 (with exceptions as above). The count 3 appears less often than the others; in the part of the colored sequence shown above (on p. 17), it appears only for prime 3. The first few other occurrences are for the primes 37, 61, 83,

Essentially, this finishes our first goal (started on p. 20): 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.

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.

Appendix: More patterns, and additional pictorial examples

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 for stating how the Langlands program works in the simplest possible cases. 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 “the fractality of $F(t)$ ” is our hammer, a lot of themes related to the Langlands program happen to work very well as nails!

Finer points of the transliteration rules






What we discuss here is an immediate continuation of what we did in the last section of the preceding chapter.¹⁰⁸


Remark 34: 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 mod 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.

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 $\widetilde{N}_p^{\text{res}}$ 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.12. **Conclusion:** One can find $\widetilde{N}_n^{\text{res}}$ knowing the first color of the number n and the position of n on 971-wheel.¹⁰⁹

Remark 35: The counts $\widetilde{N}_n^{\text{res}}$ 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 $\widetilde{N}_n^{\text{res}}$ 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), (d) on p.46 above by something much easier to explain.

¹⁰⁸The 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.

¹⁰⁹The polynomial we considered above has (cubic) 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.118).

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/subtract/multiply in this arithmetic; we may also divide by any “number” but $0 \bmod 9$, $3 \bmod 9$ and $6 \bmod 9$. What we do above to find $\widetilde{N}_9^{\text{res}}$ is we “replant” our polynomial to this arithmetic, and look how many times it takes the value $0 \bmod 9$.

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’!^{110 111} 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 $\widetilde{N}_9^{\text{res}}$.

So instead of finding the count $\widetilde{N}_9^{\text{res}}$ of residues where the polynomial takes value **0**, we do the same in the Galois arithmetic. Denote these counts $\widetilde{N}_n^{\text{Gal}}$ (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 $\widetilde{N}_n^{\text{Gal}} = \widetilde{N}_n^{\text{res}}$ provided n is prime; this also works if n is not divisible by any square (except 1^2).¹¹² It turns out that this holds for about 61% of numbers! (The exact fraction turns out to be $\frac{6}{\pi^2} \approx 0.6079$.)

Remark 36: We said that to obtain fractal behavior, one must use the counts $\widetilde{N}_n^{\text{Gal}}$ instead of $\widetilde{N}_n^{\text{res}}$. How come that the recipe for N_n given above does not mention $\widetilde{N}_n^{\text{Gal}}$?

In fact, polynomials of degree ≤ 3 are very special: one can find $\widetilde{N}_{p^k}^{\text{Gal}}$ for prime p provided one knows $\widetilde{N}_p^{\text{res}}$. Moreover, this is almost exactly the process we apply on Step (c) of p. 46! **Conclusion:** Step (c) hides recalculation from $\widetilde{N}_{p^k}^{\text{res}}$ to $\widetilde{N}_{p^k}^{\text{Gal}}$. One can omit this step if one uses suitable formulas like $N_{p^2} = (\widetilde{N}_{p^2}^{\text{Gal}} + (\widetilde{N}_p^{\text{Gal}})^2)/2 - \widetilde{N}_p^{\text{Gal}}$ etc.¹¹³

Remark 37: The recipes of Steps (c), (d) on p. 46 look coming out of a clear blue sky. In fact, they come from a very general principle:

The numbers $\widetilde{N}_{p^k}^{\text{Gal}}$ 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 116.) In fact, the same holds for the sequence (N_{p^k}) .¹¹⁴ Additionally, instead of our polynomial of degree 3, one can take any polynomial; the same works for polynomials of any number of variables¹¹⁵—and even when one counts “common zeros”: arguments where several polynomials all take value $0 \bmod p^k$ (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. (As Yu. I. Manin writes, this “forever chang[ed] our

¹¹⁰Gauss’ notebooks show that he also knew about this arithmetic—but he did not publish this.

¹¹¹Essentially, 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 152 on p. 60.)

¹¹²For example, $\widetilde{N}_{30}^{\text{res}}$ is OK, but $\widetilde{N}_{60}^{\text{res}}$ needs to be recalculated, since $2^2 = 4$ divides 60.

¹¹³The 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 $1 + \sum_k \overline{N}_{p^k} \tau^k = \exp \sum_k \widetilde{N}_{p^k}^{\text{Gal}} \tau^k / k$ (equality of Taylor series in τ).

The subtraction of the second term $\widetilde{N}_p^{\text{Gal}}$ is harder to explain, since it is due to purification process of Step (b). Essentially, in the formula above we may replace \overline{N} by N if we replace $\widetilde{N}_{p^k}^{\text{Gal}}$ by $\widetilde{N}_{p^k}^{\text{Gal}} - 1$.

¹¹⁴Moreover, the same also holds for $\widetilde{N}_{p^k}^{\text{res}}$ —but this is trivial: for most p the numbers $\widetilde{N}_{p^k}^{\text{res}}$ do not depend on k .

¹¹⁵Compare with the beginning of the section on p. 17.

understanding of the relationships between continuous 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), (d) on p. 46.¹¹⁶

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:

The “related” powers p^k and p^l of p are “far away”: the connection works¹¹⁷ only if $k|l$.

(For example, the only things in common between the Galois’ replacements for mod p^3 , mod p^4 , mod p^5 are the p residues mod p .¹¹⁸) *This* is the reason why Weil conjectures are so deep (and, for many people, much deeper than they look on the first sight).

Remark 38: 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,¹¹⁹ Weil himself did not believe that Hasse conjectures can keep water.¹²⁰

“Purification” and Motives

The operation we did on Step (b) on p. 46 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” consists of two independent “pure” parts. Each “pure” 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.)

This section turns out to be the most technical in this report.¹²¹ 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.

¹¹⁶The 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), (d). (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 $N_{p^{k+6}} = N_{p^{k+5}} + N_{p^{k+4}} - N_{p^{k+2}} - N_{p^{k+1}} + N_{p^k}$.)

¹¹⁷Another 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} .

¹¹⁸Moreover, 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 mod p . (We discuss such symmetries in Footnote 152 on p. 60.)

¹¹⁹This 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.

¹²⁰Later he changed his mind and confirmed the conjecture in a few cases — and now it is named “Hasse–Weil conjecture”.

¹²¹Except for our calculations with Eisenstein series on p. 85.

Remark 39: One way to explain what happens in Step (b):

There is a hierarchy of “difficulty” of sequences, and:

“Purification” means: “remove” from the given sequence any trace of “simpler” sequences.

Sequences simpler than degree=3 are sequences of degree 0, degree 1 and degree 2. **Conclusion:** in our sequence $\widetilde{N}_n^{\text{res}}$, we need to

- find “the traces” of “ $\widetilde{N}_n^{\text{res}}$ for sequences of degree 2”,
- find “the traces” of “ $\widetilde{N}_n^{\text{res}}$ for sequences of degree 1” (and degree 0), and
- subtract these traces from our sequence $\widetilde{N}_n^{\text{res}}$.

Essentially, we want to write down our counts $\widetilde{N}_n^{\text{res}}$ related to a sequence of degree 3 as

$$\widetilde{N}_n^{\text{res}} \equiv T_n + (\widetilde{N}_n^{\text{res}} - T_n),$$

with T_n being a combination of counts related to sequences of degree 0, 1 or 2, and $\widetilde{N}_n^{\text{res}} - T_n$ “having no similarity to counts $\widetilde{N}_n^{\text{res}}$ related to sequences of degree 0, 1, or 2”.

Remark 40: It turns out that it is easy to characterize sequences “having no similarity to counts $\widetilde{N}_n^{\text{res}}$ 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.¹²²

Moreover, putting $T_n \equiv 1$ in the formula above achieves the goal:

The average value of the sequence $\widetilde{N}_n^{\text{res}}$ 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). As a result, the counts $\widetilde{N}_p^{\text{res}} = 0, 1, 3$ at prime indices p become $-1, 0$, and 2 .

Remark 41: It is not that hard to explain the meaning of the rule in the red frame.

First of all, degree 0 leads to $\widetilde{N}_p^{\text{res}} = 0$ for most of primes p —so we may forget about it.¹²³ Note that degree 1 leads to $\widetilde{N}_p^{\text{res}} = 1$ for most of primes p .

¹²²As 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.66). 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 53 on p.65), this is enough for our purposes.

¹²³On the other hand, tuning an equation of degree 0 (such as an equation $35 = 0$ in an unknown x —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 x 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.

Next, recall the pattern we observed for “ $\widetilde{N}_n^{\text{res}}$ 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 $\widetilde{N}_p^{\text{res}}$ 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 “purify”, 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.

Given a general sequence ν_n , how to “purify” it, making the average of ν_p on primes p in every arithmetic progressions to be 0 after purification? It looks like for every arithmetic progressions we need to subtract the “averaged” value in this progression.¹²⁴

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 42: 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 34 we saw that the counts $\widetilde{N}_p^{\text{res}}$ 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.¹²⁵

Obviously, the average of $\widetilde{N}_p^{\text{res}}$ 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!

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.¹²⁶ In our case, it says that $\widetilde{N}_p^{\text{res}} = 1$ for $\frac{1}{2}$ of primes p (this matches¹²⁷ our claim that $\frac{1}{2}$ of 970 arithmetic progressions have “the second color red” in the table of Remark 34), while $\widetilde{N}_p^{\text{res}} = 0$ for $\frac{1}{3}$ of primes p and $\widetilde{N}_p^{\text{res}} = 3$ for $\frac{1}{6}$ of primes p .

So *under the condition* “the second color is green” in the table of Remark 34, $\frac{2}{3}$ of the primes are going to have $\widetilde{N}_p^{\text{res}} = 0$, and the remaining $\frac{1}{3}$ of them have $\widetilde{N}_p^{\text{res}} = 3$. Obviously, this leads to the average being 1.

Remark 43: 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 $\widetilde{N}_p^{\text{res}} = 1$ appears there (for prime p)—so the all the primes there are green. In the other half, primes are red and green matching $\widetilde{N}_p^{\text{res}} = 0$ and $\widetilde{N}_p^{\text{res}} = 3$.

¹²⁴This is the limit of averages on longer and longer parts of this sequence.

¹²⁵Actually, 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.

¹²⁶In fact, for our question, the earlier version discovered by Frobenius is enough.

¹²⁷Well, this also involves the Dirichlet theorem on primes in arithmetic progressions.



Conclusion: inside this arithmetic progression there is no observable pattern.

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!

In Section “The toy fractality law as a symmetry” (see p.30) we constructed an example of a 2π -periodic function $g(t)$ which has a horizon-self-similar point $t = 0$.¹²⁹ As we saw, the non-smooth points of any such function $G(t)$ are images of $t = \infty$ under chains of transformations¹³⁰ $t' = -1/\gamma t$ 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 44: 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 .

¹²⁸... 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.

¹²⁹Note 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 $\text{Re } G(t)$ and $\text{Im } G(t)$ such that near $t = 0$, $\text{Re } G(t)$ is horizon-similar to $-\text{Im } G(t)$, and $\text{Im } G(t)$ is horizon-similar to $\text{Re } G(t)$.

¹³⁰The --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 174 on p. 69.)

Moreover, if $4\pi^2\gamma_0$ is an integer,¹³¹ then using the new variable $T = t/2\pi$, these fractional-linear transformations are going to have integer coefficients $\alpha, \beta, \gamma, \delta$.¹³²

Remark 45: It turns out that if $\pi^2\gamma_0 > 1$ (as in the example in the section on p. 30, where $\gamma_0 = 2/\pi$, and as in all examples related to divisors of polynomial sequences), then there are other restrictions. 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. 33.) Hence starting with $t = 0$, shifting by multiples of 2π , and applying $t' = 1/\gamma t$ (and combining these transformations in an arbitrary 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!¹³⁵

In fact, we already saw this effect in pictures of the section on p. 30, when such a “prohibited zone” appeared as a “smooth” zone in the graph near $t = \pi$. In the following section (on p. 32) 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

¹³¹This 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.

¹³²To 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 $c = 4\pi^2\tilde{\gamma}_0$, here the base transformation is written as $t' = -1/\tilde{\gamma}t$.)

Actually, it is very important for us that the strongly congruence transformations form a “sufficiently small” collection of fractional-linear transformations: this makes the tessellations of the section on p. 71 possible. The Lobachevsky-rotations sending one “tile” of tessellation to another one coincide with the strongly-congruence transformations.¹³³

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 137).

¹³³Moreover, for $c > 4$ the arguments in Remark 45 on p. 54 show that just a tiny part of the collection of strongly-congruence transformations may be formed by chaining $T' = -1/cT$ and $T' = T \pm 1$.¹³⁴

Indeed, the latter collection was already discussed in Remark 24 on p. 34; 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 135. We discuss such an example in Remark 46 on p. 56.)

With the pictures of the section on p. 71 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 189 on p. 72.)

¹³⁴Note 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.

¹³⁵Compare with strongly congruence transformations: it is not hard to see that for any c , one can make the image β/δ of 0 to be arbitrarily close to any given number.

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. 28 before Remark 21) can be summarized as:

Near every t there is a number $2\pi R/s$ which is a horizon-self-similar point for $F(t)$.

(Recall that the “actual” transform for $F(t)$ implies “the honest transform for antiderivative” for $F^{(-1)}(t)$; see p. 29.)

However, the conductor $c = 997$ 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 12 on p. 23):

If c is prime and $d > 0$, then every number $2\pi R/s$ is horizon-self-similar.

For negative discriminant (the “modular form case”), the situation can be described as

If c is prime and $d < 0$, then every number $2\pi R/s$ is horizon-similar to either $F(t)$, or $\text{Im } F_{\mathbb{C}}(t)$.

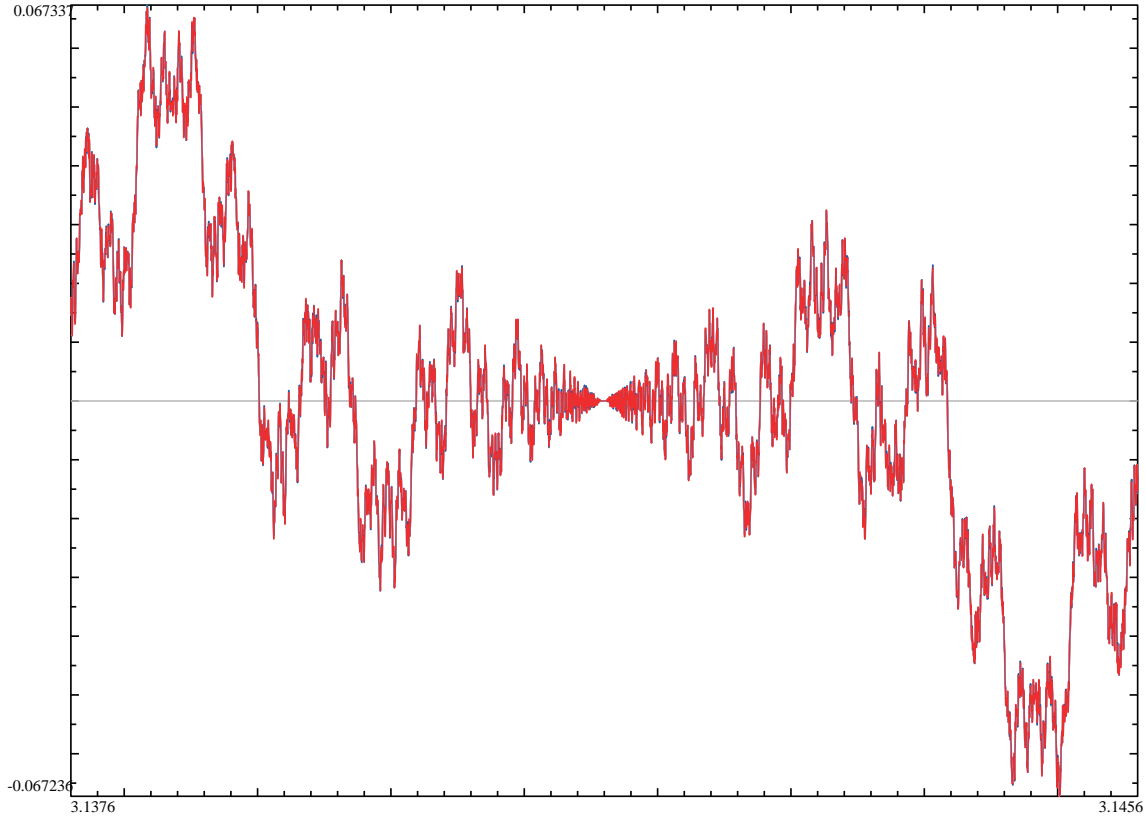
We already saw indications of this in our plots of $F^{(-1)}(t)$ near $t = 0$: these plots were fractal transforms of $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$. Now we know that something similar is going to happen for every rational point: $F^{(-1)}(t + 2\pi R/s)$ is going to be (up to additive constant) similar to¹³⁶ 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 $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$.

Going back to the case of the prime conductor 997 with $d < 0$: while reaching horizon-self-similar points requires zooming about c^2 times, many points which are “horizon-similar to the imaginary

¹³⁶The first case happens when c divides S . Note how the transform $x \mapsto -1/|c|x$ 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!)

part” may require much smaller magnification. For example, here is what happens for $R/S = 1/2$:



Comparing to the graph near $x = 0$ on p. 37, one can observe 3 differences:

- The “oscillating zone” is half as wide for the new graph.
- The sign of oscillations is inverted.¹³⁷ 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 $12/3$). For the maxima on the old graph, the corresponding ratio is about 2 (and the next one is about $1\frac{1}{2}$).¹³⁸

Remark 46: We want to stress that all the preceding examples of graphs of $F^{(-1)}(t)$ but one on p. 44 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 equation¹³⁹ 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.¹⁴⁰

¹³⁷It turns out that this is due to $2 \bmod 971$ being not a square. With Legendre symbol from p. 113, $(\frac{2}{-971}) = (\frac{2}{3}) = -1$.

¹³⁸This non-linear transformation is fractional-linear (see p. 53): $T \mapsto 1/2 + T/2(971T + 2)$; 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 .

¹³⁹See the section on p. 66 for details.

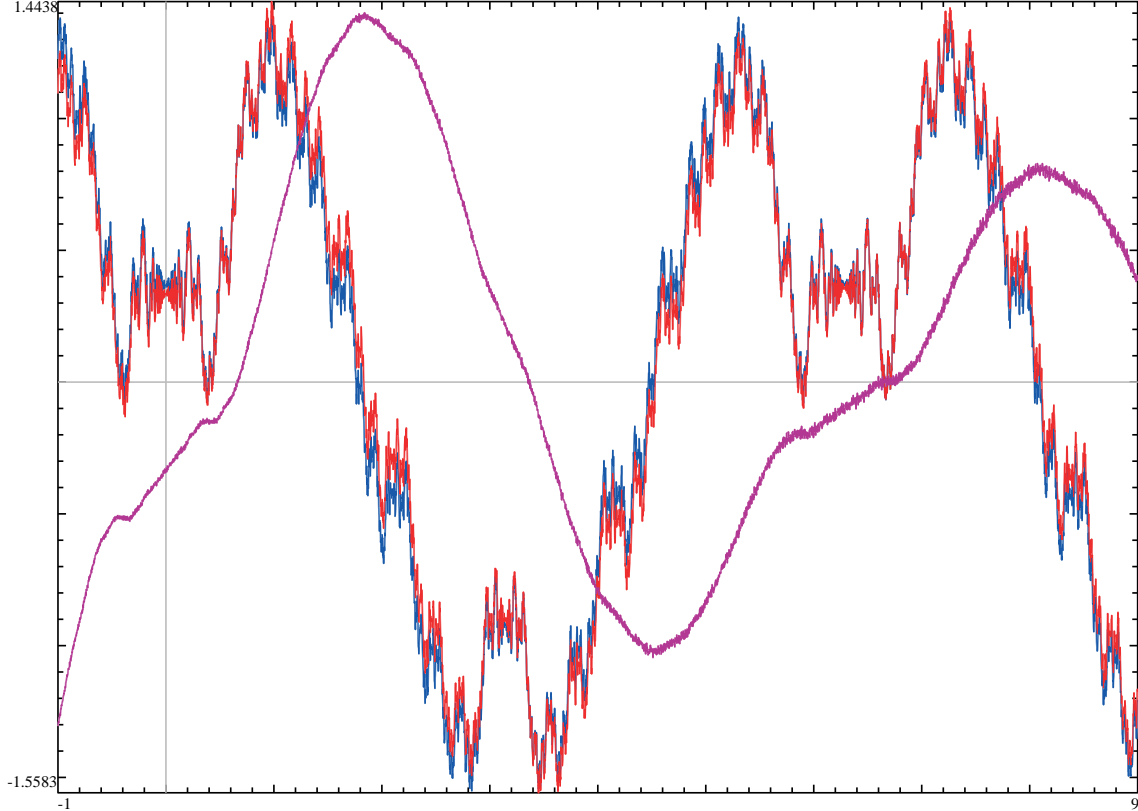
¹⁴⁰I do not know whether such observations were actually made during this period.

However, the last graph illustrates a phenomenon which (as far as I know) cannot be explained by Hecke's result alone.¹⁴¹ (In fact, the majority of points $2\pi R/s$ are of this type; see Footnote 133 on p. 54.)

The honest fractality law for $F^{(-1)}(t)$

Above, on p. 36, 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 from p. 39): the red curve is the plot of the toy transform of $F^{(-1)}(t)$, the blue curve plots $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ (for conductor 23),¹⁴² and violet plots the difference:¹⁴³



The graph for difference is scaled up 10 times; 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.

Calculation: Assume that F and \tilde{F} are related by the “actual” fractality law, so $F(t) = \tilde{F}(1/t)/t$. Integrate by part, denoting the antiderivatives of F and \tilde{F} by g and \tilde{g} (so $F(t) = g'(t)$ and $\tilde{F}(t) = \tilde{g}'(t)$). Then $g(t) =$

¹⁴¹We 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 become known about a decade after Hecke (but it is doubtful people noticed it before 50s). For more details, see the section on p. 66.

¹⁴²Compare with Footnote 90.

¹⁴³The “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).

$-t\tilde{g}(1/t) + \text{Rest}(t)$ (here $\text{Rest}'(t) = \tilde{g}(1/t)$). These three terms are exactly what is plotted above. These relations explain the observations above.¹⁴⁴

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 47: 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. 43.

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 the list of recipes for numbers N_{p^k} (see Item (c) on p. 46) should be augmented: while the first recipe below was relevant for degree 3 too, 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).

Assign these values to N_{p^k} . (Note that the first number matches the value for N_p given by recipe (b) on p. 46.)

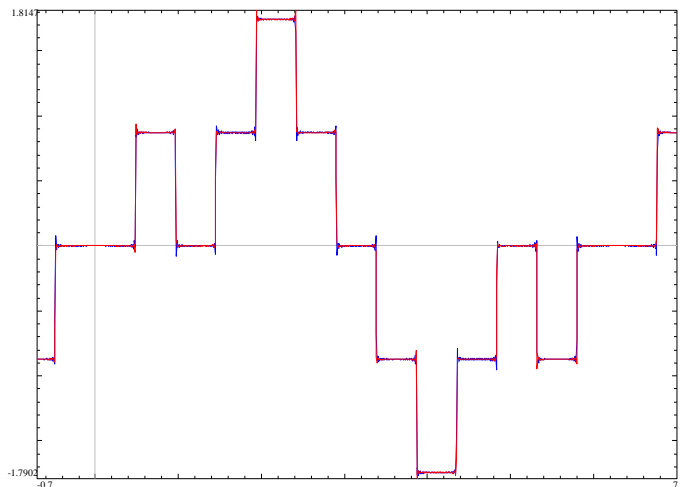
For “exceptional” primes (divisors of the discriminant, of the denominators of coefficients, of the leading coefficient, and possibly for $p = 2$) there is an additional possible choice:

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

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

Above, we wanted to write the recipe in the form similar to our recipe for degree 3 (see (c), (d) on p. 46). 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”: $N_{ab} = N_a N_b$.

Moreover, using the Legendre symbol from p. 113, Quadratic Reciprocity shows that $N_p = \left(\frac{p}{D}\right)$, with D being the discriminant of the polynomial. By top-multiplicativity of Legendre symbol (see p. 113), $N_m = \left(\frac{m}{D}\right)$ for every m . **Conclusion:** the sequence N_m is D -periodic.



¹⁴⁴A very observant reader would note that with the formula above, $\text{Rest}(t)$ would be very singular at 0. To avoid this singularity, we cheated and shifted the argument t in the graph by 2π .

This immediately implies that $F(t)$ 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 is for the polynomial $4n^2 + 2n - 3$ with $D = 13$.¹⁴⁵

Conclusion: in the case of degree 2, the pattern of colors is “already exposed” in our sequence of colors (see p. 7; the corresponding symmetries are the periodicity and mirror symmetry when colors are restricted to prime numbers; see the section on p. 14). 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. 50: every “flavor” of a pure motive needs a *specific* approach to expose its pattern of (hidden) symmetries.¹⁴⁶ 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. (And, as we said before, applying such approaches to “a mix” of pure motives leads to yet messier results. We consider two such examples in the next but one section.)

Denominators in Weil Conjectures

In (c), (d) on p. 46 and in the preceeding section, we saw 6 different cases for the sequences N_{p^k} :

Sequence	How to extend	Series	Or	d
$-1, 0, 1, -1, 0, 1, \dots$	3-periodic	$\frac{1}{1+u+u^2}$	$\frac{1-u}{1-u^3}$	3
$0, 1, 0, 1, 0, 1, \dots$	2-periodic	$\frac{1}{1-u^2}$	$\frac{1-u}{(1-u)(1-u^2)}$	1, 2
$2, 3, 4, 5, 6, 7, \dots$	a linear function	$\frac{1}{(1-u)^2}$	$\frac{1-u}{(1-u)^3}$	1, 1, 1
$-1, 1, -1, 1, -1, \dots$	2-periodic	$\frac{1}{1+u}$	$\frac{1-u}{1-u^2}$	2
$1, 1, 1, 1, 1, 1, 1, \dots$	1-periodic	$\frac{1}{1-u}$	$\frac{1-u}{(1-u)^2}$	1, 1
$0, 0, 0, 0, 0, 0, 0, \dots$	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 “purification”),¹⁴⁷ and the denominator being a product of terms $1 - u^d$, one for each irreducible factor of the reduction mod p of the polynomial

¹⁴⁵Here 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 157 on p. 63.)

¹⁴⁶There is a very nice (and more detailed) summary of relevant issues in the discussion *What is the Langlands Programme?* in **n-Cat Café**.

¹⁴⁷This numerator is going to be always eventually cancelled, since all the possible factors of denominators discussed below are divisible by $1 - u$.

in question (listed on the right).¹⁴⁸ Here d is the degree of the factor. Note that we *ignore* the multiplicity: if a certain factor appears many times, we *still* count it once.

For example, for polynomials of degree 3 having no roots mod p means that the polynomial is irreducible mod p ; hence for the red primes, the denominator is $1 - u^3$, leading to the first sequence. In the same situation, having only one root mod p (with multiplicity 1) leads to $(1 - u)(1 - u^2)$, giving the second case; the third case corresponds to having 3 roots mod p (leading to $(1 - u)^3$). The fourth case matches $1 - u^2$ (irreducible quadratic polynomial). The fourth case corresponds to $(1 - u)^2$; this is either polynomial of degree 2 with 2 roots, or a polynomial of degree 3 with a double root mod p (it automatically has another simple root). The remaining case matches $1 - u$, which means a single multiple root (of multiplicity equal to the degree).

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 37 on p. 49: we represent this rational function as a product of very simple terms.¹⁵⁰

(The general case of Weil conjectures characterizes possible factors of the corresponding rational function in more complicated situations, when the polynomial equations depend on several variables, and there may be more than 1 equation.)

Remark 48: Returning back to the case of one polynomial of one variable: what is crucial for the proof of fractality of $F(t)$ is that for non-exceptional primes, the degree of the denominator is equal to the degree of polynomial. Moreover, the particular factors appearing in the denominator match a particular symmetry type of the (complex) roots of the polynomial.¹⁵²

¹⁴⁸For an “exceptional” prime p , one may need to “improve” the polynomial; one can make a variable change and/or multiply the polynomial by a constant in \mathbb{Q} . The aim is for the “improved” version to have as many distinct roots mod 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 mod p into two $m = 0, 1 \bmod p$. Note that this would not work for $n(n - p) - p$: one cannot split *this* double root. (To be honest, the procedure for “improving” may be more involved than what is suggested 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!¹⁴⁹)

¹⁴⁹Judging 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 absence 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.

¹⁵⁰To further demystify the denominator above, note that it is a characteristic polynomial of a certain permutation matrix.¹⁵¹ The powers d appearing above form the *cycle decomposition* of this permutation. Moreover, one can recognize these cycles as *orbits* of this permutation.

¹⁵¹... which is a matrix with exactly one non-zero entry (equal to 1) in every row and column.

¹⁵²Note 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.

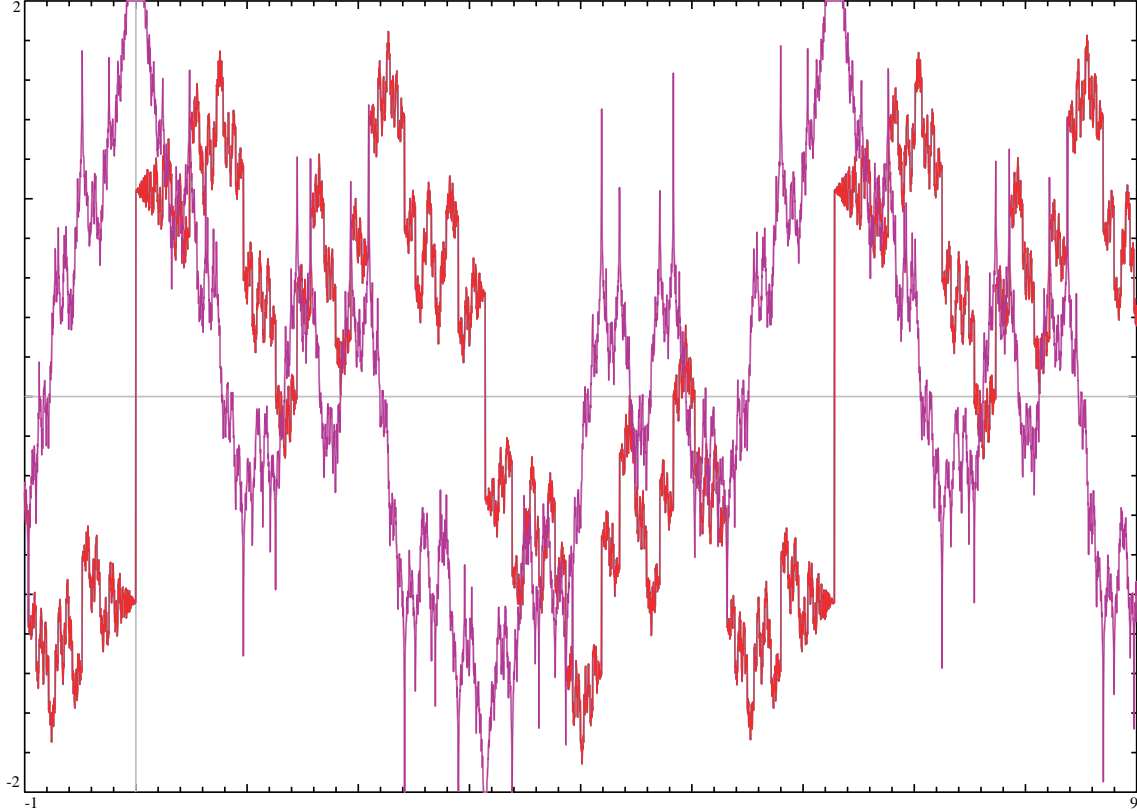
The relation to our denominators is that for every non-exceptional prime p , there is a particular Galois permutation of the roots whose permutation matrix is exactly as described in Footnote 150. This symmetry is named *the Frobenius element* for p (well, we cheated a bit: this symmetry is defined just “up to rotating it by other symmetries”; in other words, it is just a *conjugacy class* of a symmetry).¹⁵³

¹⁵³The 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 $p \gg 0$.

Maass fractality laws: decomposable and abelian cases

As we discussed it on p. 41, the polynomials “ $M \times \text{tetrahedral numbers} + 1$ ” 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. 66).

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 very small, 13. This is much smaller than $c = 148$ considered on p. 41. 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 imaginary part $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ is in violet:



Observe the principal properties of the violet graph:

- The “tips” of the violet graph are cut-off. This is due to the imaginary part having “logarithmic singularities” at all points $2\pi R/s$. The widest of these go outside the y -limits of our plot.
- The violet graph has a lot of “spikes”; this is due to the same singularities. In fact, if we could increase the number of sample points for our graph by many orders of magnitude,¹⁵⁴ one could see that these spikes actually go up or down to infinity!
- **Conclusion:** the low resolution of this plot hides another pathology: the function $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ is unbounded near these points. Since these points are dense, this means that the function is unbounded in every small interval—which means that it is impossible to plot it honestly!¹⁵⁵

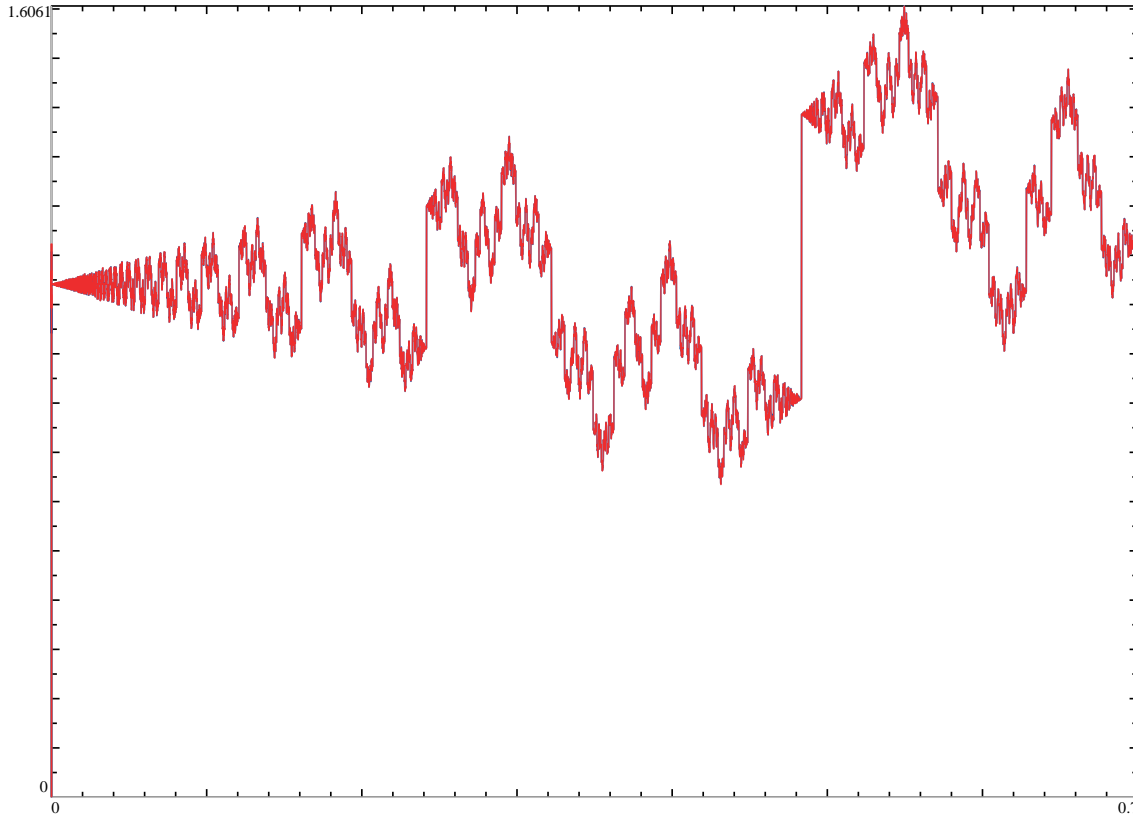
¹⁵⁴A singularity $y = 1/n \log t$ becomes *exponentially* more narrow when $n \rightarrow \infty$. The corresponding jumps on the red graph are visible for n up to hundreds. One would need astronomical number of sample points to see similar number of spikes on the violet graph!

¹⁵⁵While the “spikes” on the graph of $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ happen for t in an everywhere dense subset of \mathbb{R} , their projections to the x -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 “pathologically good” approximations by rational numbers) $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ is close to the violet graph. (To have a plot for *every* t , one needs to take an extra antiderivative: $\text{Im } F_{\mathbb{C}}^{(-2)}(t)$ has a honest plot.)

For the red graph:

- The red graph has a jump at every point $2\pi^R/s$.
- Moreover, on 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 $t > t_0$).¹⁵⁶
- On the right of every jump (say, for $t > t_0$), the red graph behaves as a Lipschitz function: $|\Phi(t) - \Phi(t_0+0)| \leq C \cdot (t - t_0)$. (Likewise on the left.)

Recall what we saw for $M = 24$ (on p. 41): 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 $J \approx 2.08$, 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. 113). All this works for $13 \nmid S$.

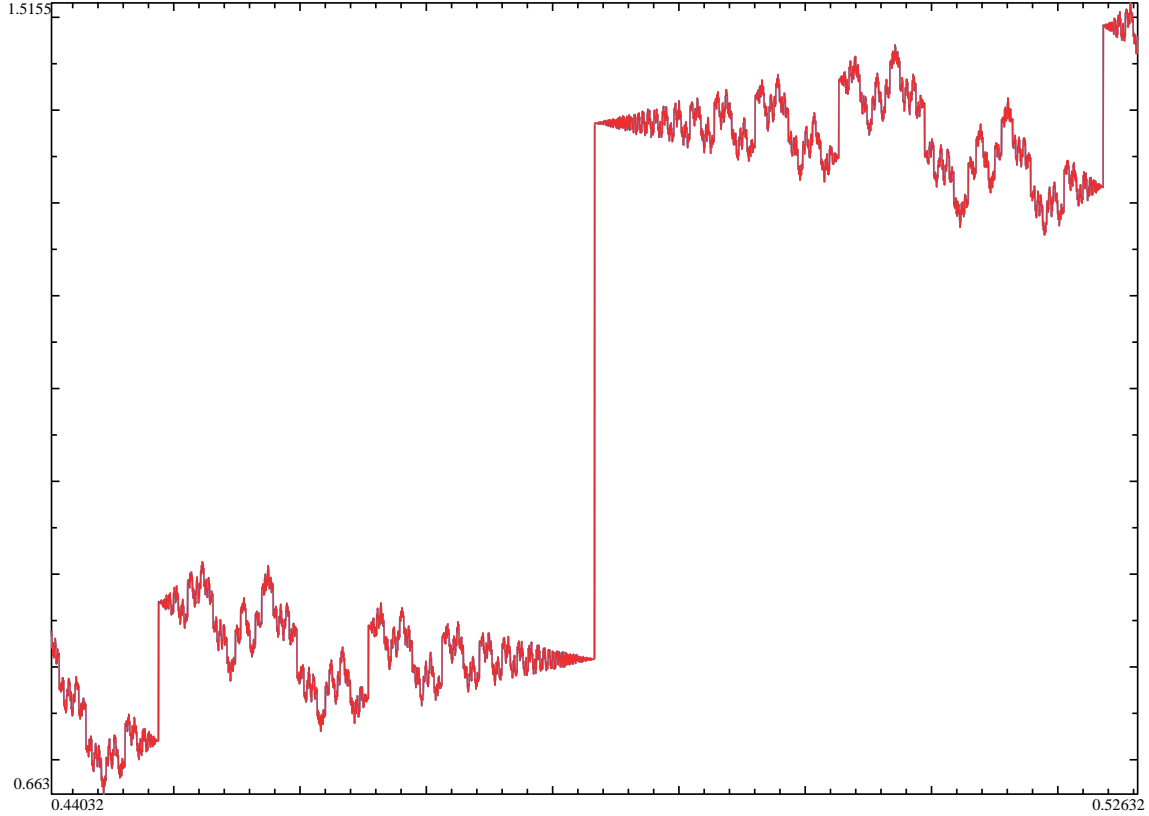
¹⁵⁶One should be extremely careful with statements like this, since this sum is only conditionally converging. There is a way to overcome this (see Footnote 241 on p. 88). However, the result is strikingly unexpected: the sum of jumps is twice the variation $\Phi(t) - \Phi(t_0)$ of the function!

In short: a few paragraphs below, we 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. 85—does not lead to more understanding of what happens.)

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 $\left(\frac{R}{13}\right)$.

For example, zoom in a lot¹⁵⁷ near $t = 2\pi/13 \approx 0.48332$:



The jump is by about $0.577 \approx 2.08/\sqrt{13}$, as predicted above.

As in the case $c = 148$, the shape of “one oscillation” in the pattern of oscillations near the points of jump matches the shape of the graph of one period (compare with the first graph of this section). In this way, the shape behaves very similar to the shapes in horizon-self-similar points. However, the jumps themselves break horizon-similarity completely (we discuss how to fix it in Remark 52 on p. 65).

How to explain the difference between what we see here (for $M = 16$) and what we saw for $M = 24$ (on p. 41)? The reason is very simple: the polynomial “ $16 \times$ tetrahedral numbers $+ 1$ ” vanishes at the point $\frac{1}{2}$. In other words, $8x(x^2 - 1) + 3 = (2x - 1)(4x^2 + 2x - 3)$. 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 $2x - 1$ and zeros of $4x^2 + 2x - 3$. Note that $2x - 1$ 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. 17) the color of primes $p \geq 3$ is going to be always green. Moreover, the number of solutions mod p (used in the section on p. 45) is one more¹⁵⁸ than the number of solutions for $4x^2 + 2x - 3 = 0$.

In the language of the section on p. 50 “the corresponding motive is not pure” — and the patterns corresponding to the factors are “overlayed on top of each other”, contaminating these patterns.

¹⁵⁷The “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.

¹⁵⁸There is no collision between the solutions, since the value of $4x^2 + 2x - 3$ at $x = 1/2$ is -1 — which has no prime factors!

In short: for a decomposable polynomial the sequence of red/green colors *is a “mix”* of colors for the factors of polynomials. Likewise for numbers N_k from the section on p. 45: they are determined by the corresponding numbers N_k^{quadr} for $4x^2 + 2x - 3 = 0$.

Remark 49: 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”).¹⁵⁹ 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 156 on p. 62), 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 formula¹⁶⁰ for $F^{(-1)}(t)$.

Remark 50: While decomposable polynomials are not covered by Langlands’ approach,¹⁶² 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 points of the first and the second type; moreover, after multiplication by $\sqrt{13}$ (and taking into account the law for how δ -functions change 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.¹⁶³

Together with Remark 49, 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 $(\frac{a}{13})$). 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 51: Note that Eisenstein series are *direct analogues* of Euler’s formulation of Quadratic Reciprocity. Indeed, essentially Euler’s formulation claims that the corresponding numbers N_k^{quadr} form a periodic sequence.¹⁶⁴ In terms of $F(t)$, this means that 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 $(\frac{R}{c})$). Compare with the graph on p. 58).

For the Eisenstein series for $(2x - 1)(4x^2 + 2x - 3)$ above, the only thing which changes is that we allow jumps with *any* denominator S with $13|S$ (instead of $S = 13$ for $4x^2 + 2x - 3$).

¹⁵⁹As 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 appearance” of the graph: the contribution of this term would be lost in all the “fractal noise” of the jumps in the graph.

¹⁶⁰To do this, one needs to rearrange this sum *smartly*, since it is obviously not absolutely convergent; we discuss this in the section on p. 85. 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 .¹⁶¹

Mathematically, such objects are described using modular symbols.

¹⁶²The functions $F(t)$ predicted by the Langlands program are “cusp form”—which are, in a certain very precise sense, “functions ‘opposite’ to Eisenstein series”.

¹⁶³However, 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 21 on p. 29.

¹⁶⁴Compare with the section on p. 58.

Remark 52: 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 17 on p. 25. 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:

the t -axis with the added point $t = \infty$ 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.

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 .^{165 166}

Remark 53: The problem with the plots above is that the Langlands program focuses on the behaviour of “pure” motives. *This* is why we needed to “purify” our sequence N_m for its Fourier transform to have the “expected” fractal properties. For an indecomposable polynomial P , the motive for $P(x) = 0$ is a mix of a motive of a point (recall that a point is a solution to $x = \text{const}$; the corresponding N_m are all 1) with “what remains”; however, if $P = P_1 P_2$, 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 “purification” would remove one copy of the point-motive; what remains is “a point” mixed with “what remains” for P_2 — which is exactly the “unpurified motive for P_2 ”! So in addition to showing what happens for decomposable P , the pictures above also show the result of our procedure applied to a quadratic polynomial $4x^2 + 2x - 3$, but without the step of “purification”.

Another case in which our naive procedure of purification $N_p = \widetilde{N}_p^{\text{res}} - 1$ does not result in a pure motive is the case of an “abelian=cyclic” polynomial P of degree 3. While in this case the periodicity $N_p = N_p^{\text{per}}$ still holds (here N_m^{per} is a certain periodic sequence), the identities $N_{p^k} = N_p^k$

¹⁶⁵Due 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.

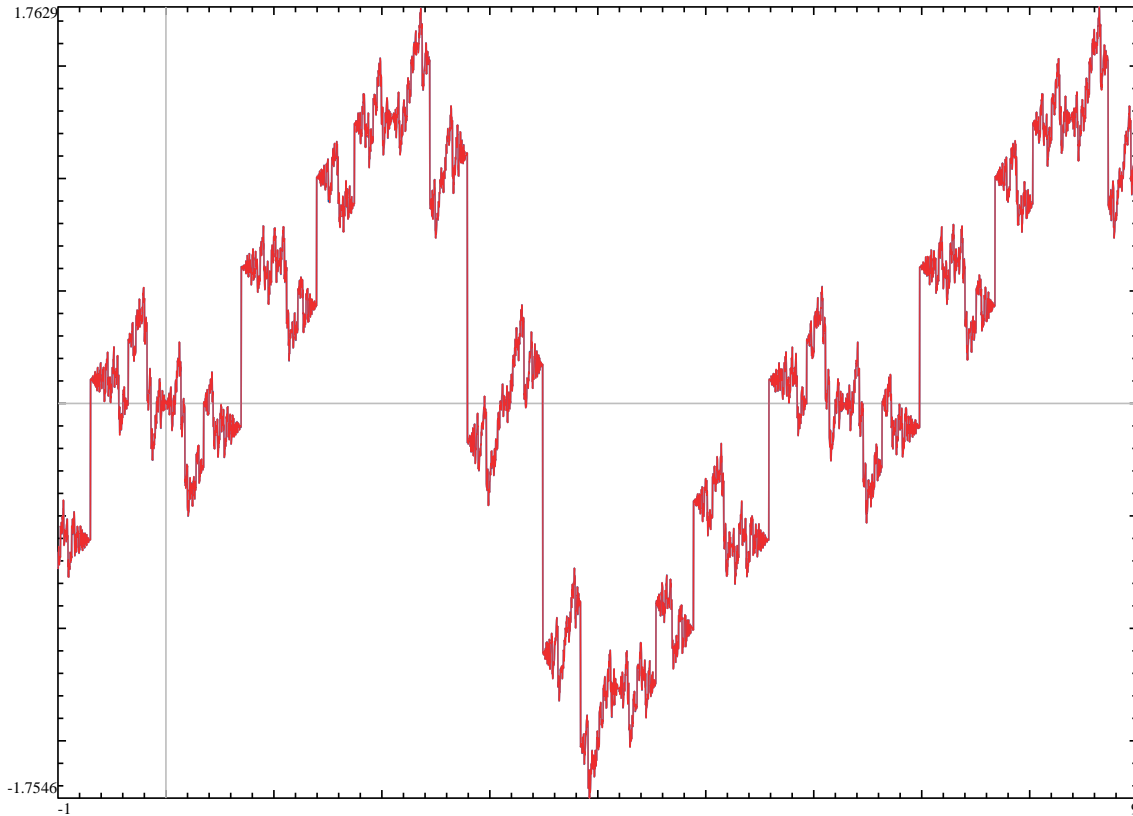
Moreover, strictly speaking, the Langlands program does not cover decomposable cases.

¹⁶⁶We 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 $\frac{1}{2}$ -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.)

and $N_{p^k}^{\text{per}} = (N_p^{\text{per}})^k$ do not hold. The corresponding graph looks (again!) like an antiderivative of an Eisenstein series:



(Here $M = 18$ and the discriminant is $D = 9^2$; compare with Footnote 100 and the discussion in the next section.)

The reason for similarity to the preceding case is that motives for an abelian polynomial split into “motives of rank 1”; in other words, the number of pure parts is equal to the degree of the polynomial. In the case above the degree is 3, so we start with 3 pure parts, and “what remains after purification” is still a mixture of two pure parts.

In a certain sense, this example is much more complicated than the preceding one! Before we were mixing a motive of a point (zero equations with zero unknowns!) with a motive for a quadratic equation. They corresponded to two functions $\widehat{N}_p^{\text{res}} \equiv 1$ and $F_p^{(-1)} \equiv N_p^{\text{periodic}}$ (for prime p). Now we are mixing two motives which are both periodic.¹⁶⁷

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

¹⁶⁷Compare with our discussion of a quartic polynomial with field discriminant 117 on p. 97. The corresponding motive is also a mix of two pure motives. That time it is a “periodic” motive mixed with a “modular form” motive.

$P = 0$ may be “relatively uncomplicated”¹⁶⁸ 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), 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 $N_p = N_p^{\text{per}}$ for prime p . (Compare with Remark 53 on p. 65.) The latter case is what we called the “odd” (or “modular form”) case in Remark 12 on p. 23. 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.¹⁶⁹

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: “Maass form”) case.

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 24 and 25, as well as Remark 45). 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. 37 for $D = -23 < 0$, and one on p. 55 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 and applying $-1/cT$. The only graph which satisfies both restrictions is one on p. 55 with $c = 37$. (Compare with Remark 46 on p. 56.)

¹⁶⁸How to do this was discovered about two centuries ago.

¹⁶⁹Apparently, 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 $q = \exp 2\pi iz$ 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.

(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 51 on p. 64. So this gives a very different example of a need to “purify” 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 15 on p. 24 — 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. 74. 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 15 on p. 24. 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 12 on p. 23) — 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 29 on p. 44.)

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 “repetition 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 $t + is =: z$ converts the Fourier transform of (N_n) into a function $f(z)$ defined on the upper half-plane $\mathfrak{H} := \{\text{Im } z > 0\}$.

It turns out that if one considers the complex Fourier transform (as in $F_{\mathbb{C}}$) then

- The function $f(z)$ is complex-analytic.¹⁷⁰ The “boundary trace” of this function is $F_{\mathbb{C}}(t)$.¹⁷¹
- 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 16 on p. 25 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 12 on p. 23 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 same¹⁷²). The even case would

¹⁷⁰This should be replaced with real-analyticity in the “even” case.

¹⁷¹Recall 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.

¹⁷²The smallest conductor for which both even and odd cases are possible is 756 with the corresponding “even” and “odd” polynomials $x^3 - 6x - 2$, $x^3 - 6x - 12$.

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. 77.

Lobachevsky-symmetries: the case $c = 1$

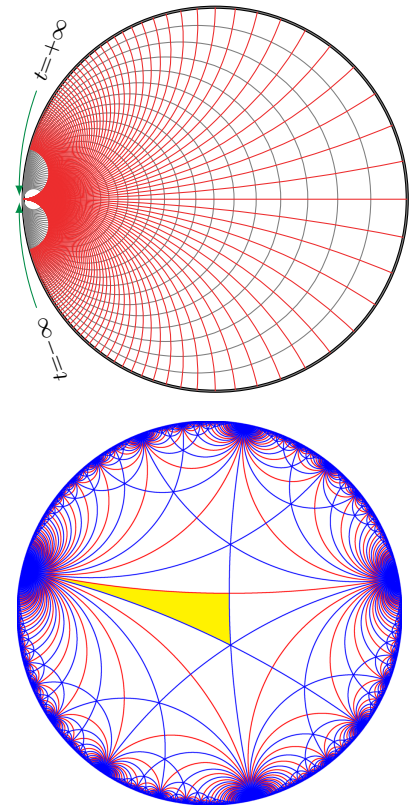
The easiest way to deal with a collection of symmetries is to find a picture such that these symmetries coincide with the symmetries of the picture. A particular case is when the picture consists of cut line which tessellate (“tile”) the plane into pieces of the same shape and the same size. Moreover, 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:

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 $F(t)$.¹⁷³

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.)¹⁷⁴ 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 tessellation of Lobachevsky geometry (on the right; on Wikipedia, it is in the article [Truncated triapeirogonal tiling](#)¹⁷⁵ together with a few other examples, some of which are for small conductors). Every piece of tessellations we consider is made of several copies of “an elementary tile”. This tile (in yellow on the right)¹⁷⁶ is marked as “index 1” in the Wikipedia article above. How the piece is made of these elementary tiles

¹⁷³Essentially, 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 $f(z)$.

¹⁷⁴In 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 17 on p. 25.

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 130 on p. 53.)

¹⁷⁵I do not know anybody using such bizarre names in real life, or in real math.

¹⁷⁶Note 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.¹⁷⁷ The yellow tile “combined” with any one of its neighbor tiles forms the piece good for¹⁷⁸ $c = 1$.

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.¹⁷⁹

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.

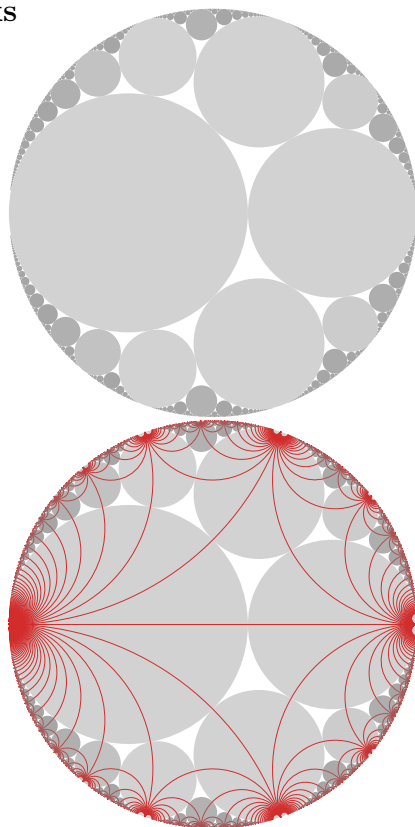
Conclusion: What is drawn above is *just a hint* of what is going on in Lobachevsky geometry. In fact, it takes a lot of training to be able to interpret these hints fully! Below, the reader may need a long leap of faith with our recurring claims like “that picture demonstrates this symmetry”.¹⁸⁰

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 tessellation). 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 gray¹⁸¹).

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.¹⁸² **Conclusion:** two pictures above have the same symmetries; moreover, if one combines these two pictures, the result still has the same symmetries.¹⁸³ (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



¹⁷⁷Keep 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.

¹⁷⁸Without doubling the yellow tile is a “piece” if we allow mirror symmetries. Compare with Footnote 223 on p. 78.

¹⁷⁹In particular, there is a (unique!) Lobachevsky-symmetry of the picture above sending any “elementary tile” to any other tile.

¹⁸⁰There are videos visualizing geometry and movements of the Lobachevsky plane. [Google for movement OR visualizing hyperbolic demo OR projections video](#).

¹⁸¹We use darker gray for smaller disks to make them easier to see. This tint has no math significance.

¹⁸²Indeed, 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.

¹⁸³In 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 179.)

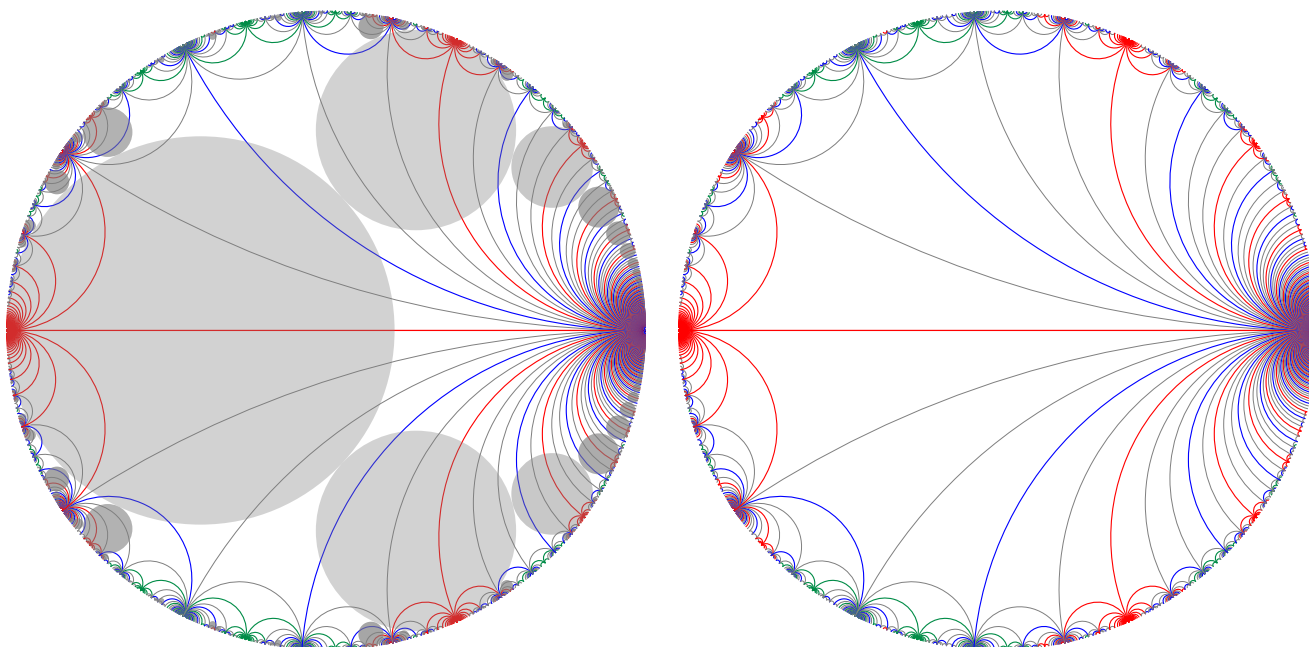
colored lines. (Additionally, we would need to rescale the coordinate t on the disk—and its boundary.)

The case $c = 5$

For larger conductors c one can draw pictures which are very similar in spirit. Here we consider the case of $c = 5$.¹⁸⁴ 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 colored¹⁸⁵ lines on the right picture match the red lines on the picture on p. 69.¹⁸⁶
- These lines cut the picture into “triangles”. Every such triangle matches a red-sided triangle which on the picture on p. 69 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.¹⁸⁷
- 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).

¹⁸⁴As we discuss it in the section on p. 77, there are several different analogues. Until then, it is enough to say that here we consider the “smallest useful” collection of symmetries.

¹⁸⁵Here and below “colored lines” means “non-gray” lines.

¹⁸⁶For this and other matches below, it is better to Lobachevsky-move the picture on p. 69 (“squeeze it to the left”).

¹⁸⁷Well, 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,¹⁸⁸ 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.¹⁸⁹

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.¹⁹⁰

Remark 57: For any gray disk, the unit “necklace rotation” of Footnote 183 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 $c = 5$.)¹⁹¹

The gray disks and the “special zones”

Above, we constructed a coloring of the Lobachevsky plane such that its symmetries coincide with the symmetries¹⁹² of the (generalized) function $F(t)$ on the absolute and of the function $f(t, s)$ on the Lobachevsky plane.^{193 194} Here we reap the fruits, using the symmetries of the picture of gray disks to inspect the fractal transforms¹⁹⁵ 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 17 on p. 25 we identify the leftmost point of the absolute with $t = \infty$ (as on the picture on p. 69). 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 in the graph on p. 41.)

¹⁸⁸Indeed, 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 223 on p. 78.) Anyway, in this chapter we ignore reflections!

¹⁸⁹Moreover, 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.

¹⁹⁰The same holds if we also take into account the coloring of the edges of the pieces.

¹⁹¹**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 $[-3\pi/c, 3\pi/c]$).

¹⁹²Here it helps to interpret the functions F , $F^{(-1)}$, and f as tensor-valued, as in Footnote 166 on p. 65. Then the toy/actual transforms become just coordinate-changes applied to tensor fields, and horizon-self-similarity may be interpreted as “being symmetrical”.

¹⁹³Recall 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 Lobachevsky-move the other in exactly the same way. See Remark 17 on p. 25.

¹⁹⁴The 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*.)

¹⁹⁵See p. 21.

As we explained above,¹⁹⁶ 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. 53). 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*¹⁹⁷), use as “the unit of measure” “the projection of the gray disk near $t = \infty$ to the absolute”.¹⁹⁸ 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*.¹⁹⁹

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 17 on p. 25, 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 190 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.

¹⁹⁶See Footnote 174 on p. 69.

¹⁹⁷Indeed, “mathematically” the fractal transform is defined *everywhere*. However, it is not everywhere “visually recognizable”: the non-linearities hide the similarities.

¹⁹⁸Making 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.

¹⁹⁹In 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.

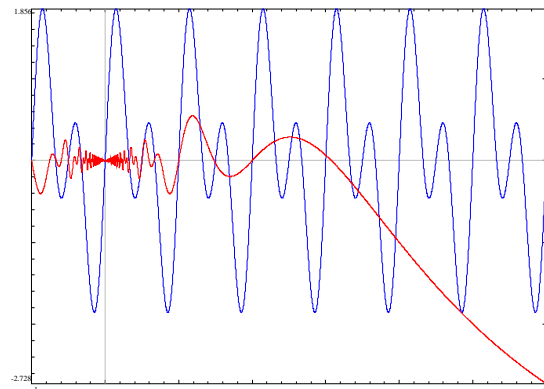
Later (in the section on p. 80), when we work with harder-to-understand pictures, we are going to have *both* the “original” and the $4/c$ -disk marked on the picture.

Recall that the whole idea of the coloring above, on p. 71, 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 .²⁰⁰

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 183 from p. 70 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 function²⁰¹ $\Phi(T)$ with period 1, and graph $\Phi(T)$ and $1/4T\Phi(1/T)$ 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).²⁰² 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 c s, 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

The pattern on the graph of $T\Phi(1/cT)$ is visible when $T = 1/cT'$ 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 one²⁰³ period around 0.

²⁰⁰Doing 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 $0, \pi, 6\pi/5, 4\pi/3, 2\pi, \infty$. For the piece to the right of it, the values are $0, 2\pi/5, \pi/2, 2\pi/3, 4\pi/5, \pi$.)

²⁰¹We use the “same” letter T for the variable as before, when we had $t = 2\pi T$ and 2π -periodic functions of t .

²⁰²The situation does not improve for $|T| > 6$, where $T\Phi(1/T)$ quickly converges to a certain limit.

²⁰³This 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 $\Phi(t)$.²⁰⁴

However, in practice, we need not the toy transform, but the “honest law for antiderivative”, see p. 36. 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 28 on p. 43.) 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. 71). 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. 40 (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!)

²⁰⁴... as we have seen in the section on p. 41, with $\Phi(T)$ having “extra” symmetries.

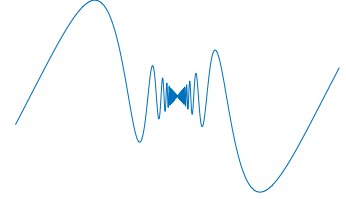
Covering properties of the zones of horizon-self-similarity

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 periodicity²⁰⁵ 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.²⁰⁶

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 needs²⁰⁷ to allow just a tiny amount of exceptions.²⁰⁸

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. 71, 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!

²⁰⁵... together with $F^{(-1)}(t)$ being actually a tensor field! (See Footnote 192 on p. 72.)

(The difference between $t = \infty$ and $t \neq \infty$ is due to extreme non-linearity of the coordinate t near $t = \infty$.)

²⁰⁶Furthermore, 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. 77). This means that we do not yet list *all* the possible zones. We complete this list later, in the section on p. 80. In the same section we also discuss horizon-similar but non-self-similar zones.

²⁰⁷To 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. 79) 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. 79, 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.)

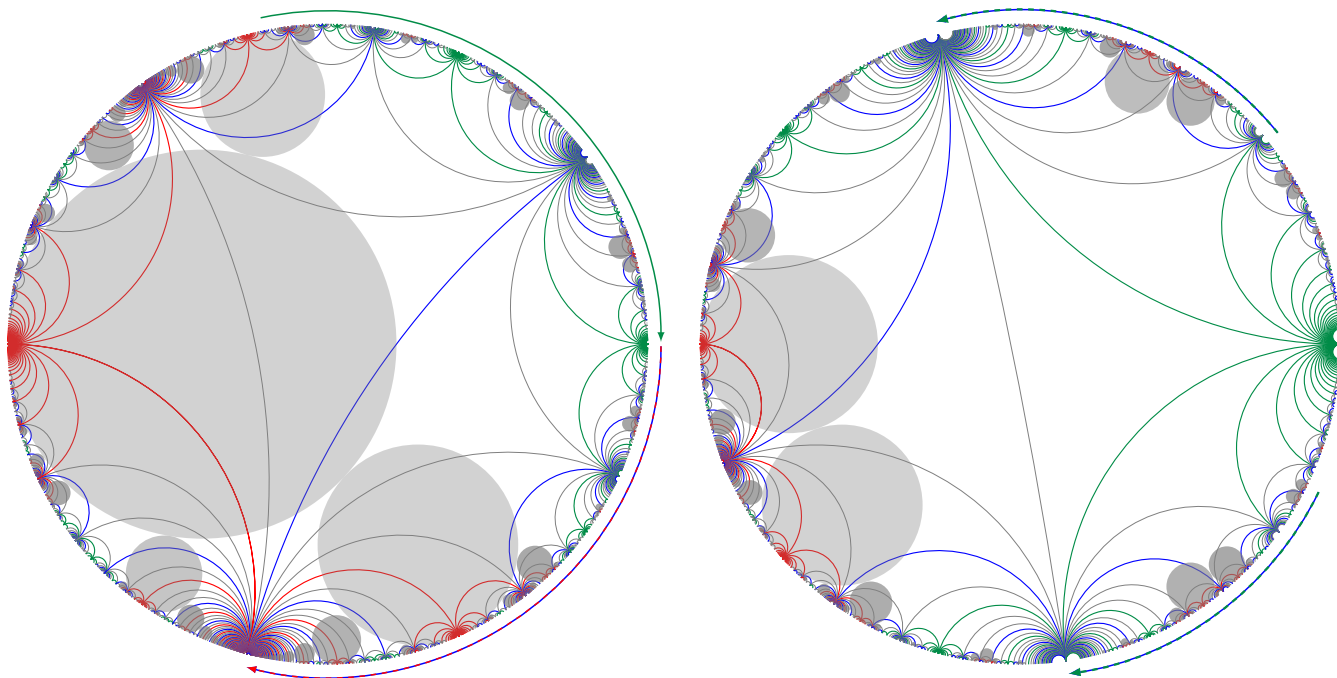
²⁰⁸The uncovered set is a “meagre” subset of measure 0.

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*, 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!²⁰⁹

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”.²¹⁰

Before we can zoom this way near our “worst offender” point on the picture on p. 71 (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 rightmost point intact and sends a slice into the next slice clockwise. (This is similar to what we did in

²⁰⁹As in Footnote 207.

²¹⁰This 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.

Remark 58 on p. 73.) **Conclusion:** all slices have the same shape and the same size (in Lobachevsky sense) and are “colored in the same way”.

In particular,

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

In any chunk of the absolute, the projections of two largest gray disks to the boundary cover about $1/3$ 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 132 on p. 54 these symmetries are both the “strongly-congruence” type, and the “keeping sign” type (these types of “congruence” transforms coincide²¹¹ for $c = 5$). Because these symmetries keep sign of $F(t)$,²¹² 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*.²¹³

One can also show²¹⁴ 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. 78 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. 79, 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.²¹⁵

²¹¹Compare with Footnote 218 on p. 78.

²¹²In pedantic mode: ... *would “keep”* the sign — if F with such a tiny conductor existed.

²¹³In terms of formula of Footnote 65 on p. 29, this means $\varepsilon > 0$.

²¹⁴**N.B. (???) Check!!!**

²¹⁵Compare with Footnote 137 on p. 56; 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 $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ ”).

Moreover, the “sign-flipping” symmetries can be described geometrically, as symmetries of the right picture on p. 71 which exchange red and green lines. So the newly added disks are tangent to the absolute at the points where the green lines meet²¹⁶, and the newly added $4/c$ -central zones are the central regions inside projections of these disks.²¹⁷

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.²¹⁸ 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.²¹⁹

Remark 61: To add insult to injury, on our graphs we saw still *other* zones of fractality, for example the zone near $t = 0$.²²⁰ As we already mentioned, the corresponding transformation $t' = 1/\gamma t$ is directly related to Hecke’s functional equation (see the section on p. 66 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 $T' = -1/cT$ ”).²²¹ So this provides 3 more classes (“as above, but possibly combined with $T' = -1/cT$ ”) to consider.²²²

We investigate the largest of these augmented types in the section on p. 80.²²³

²¹⁶... (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)

²¹⁷For 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 199 of p. 73. Compare with the picture on p. 79 — where we color the added disks red.

²¹⁸This 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 137 on p. 56. Such c s are divisors of $2^3 \times 3 \times 5 = 120$.)

²¹⁹The symmetries of the first and third arrangements have names: $\Gamma_1(c)$ and $\Gamma_0(c)$.

²²⁰While we saw that the behaviour of $F^{(-1)}(t)$ in such zones is different in “even” and “odd” cases (see Remark 12 on p. 23), the geometry of the *zones themselves* are the same. So here we treat these cases uniformly.

²²¹The 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.

²²²This is related to the fact that the suitable symmetries live in $\mathrm{PGL}_2\mathbb{Q}$ and not in $\mathrm{PSL}_2\mathbb{Z}$.

²²³Since $F(t)$ 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 132 on p. 54 — 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.

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 $T' = -1/cT$.

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 199 on p. 73) 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.

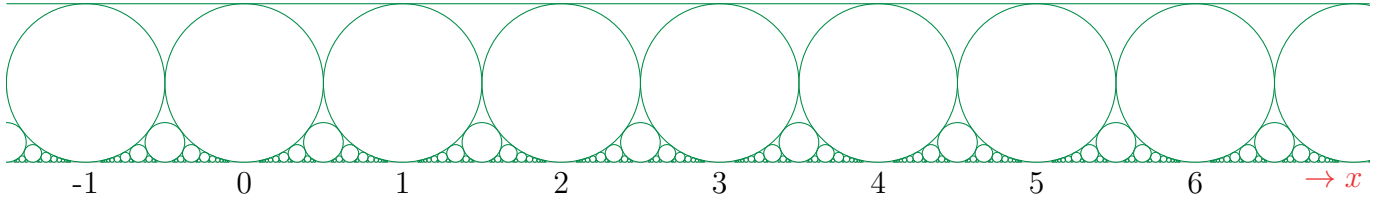
With this modification, the recipes we used before become:

- 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.

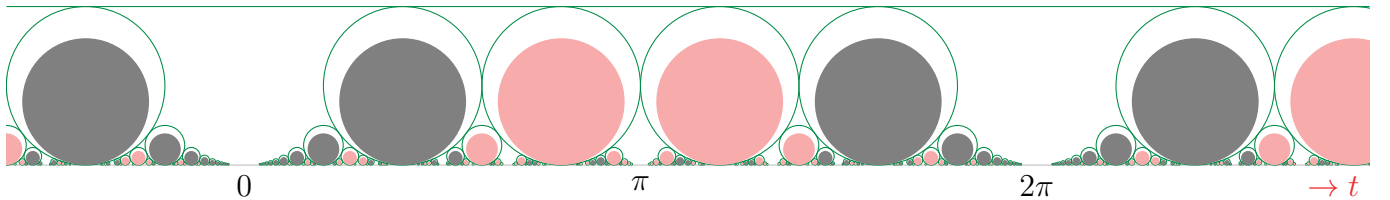
Second, we want to use a different model of the Lobachevsky plane. While the model in the disk used above simplifies visualization of Lobachevsky moves, the required book-keeping is too complicated.²²⁴ The half-plane model of the Lobachevsky plane allows us to state a more explicit description of the 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:

The red disks “match” the “sign-flipping” Lobachevsky-symmetries of $f(t, s)$.

²²⁴For example, we could not state explicitly which of the Apollonian circles are omitted on our pictures for $c = 5$.

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 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 $(\frac{R}{c})$.)²²⁵

The new process may be summarized as:

Remove some Ford circles, and “inscribe” smaller disks in the remaining circles.

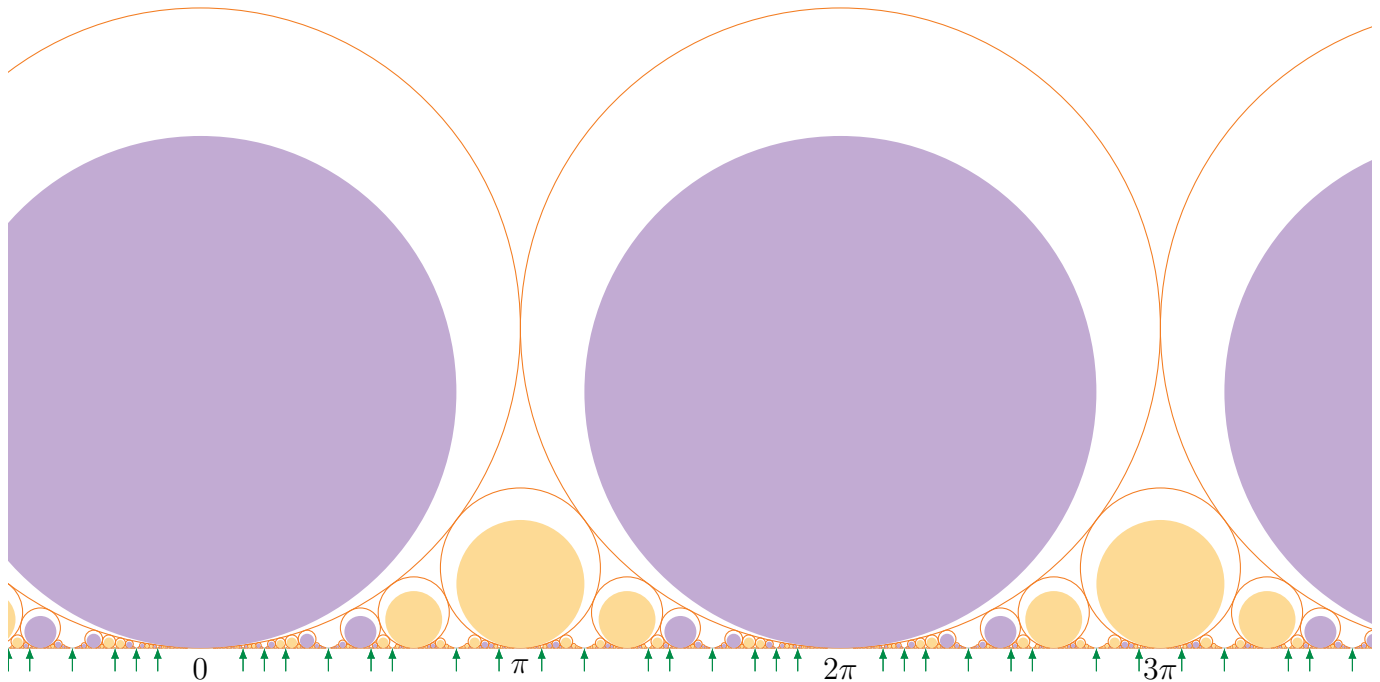
(The scaling factor for the smaller disks is $4/c$.)

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. 78 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 multiply²²⁶ $f_{\mathbb{C}}$ 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 $z := x + iy$ on the upper half-plane with $\text{Im } z \geq 0$.) 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):



²²⁵Note 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 $\text{Im } x > \text{const}$. We do not shade it, since it does not contribute to the zones in question anyway!

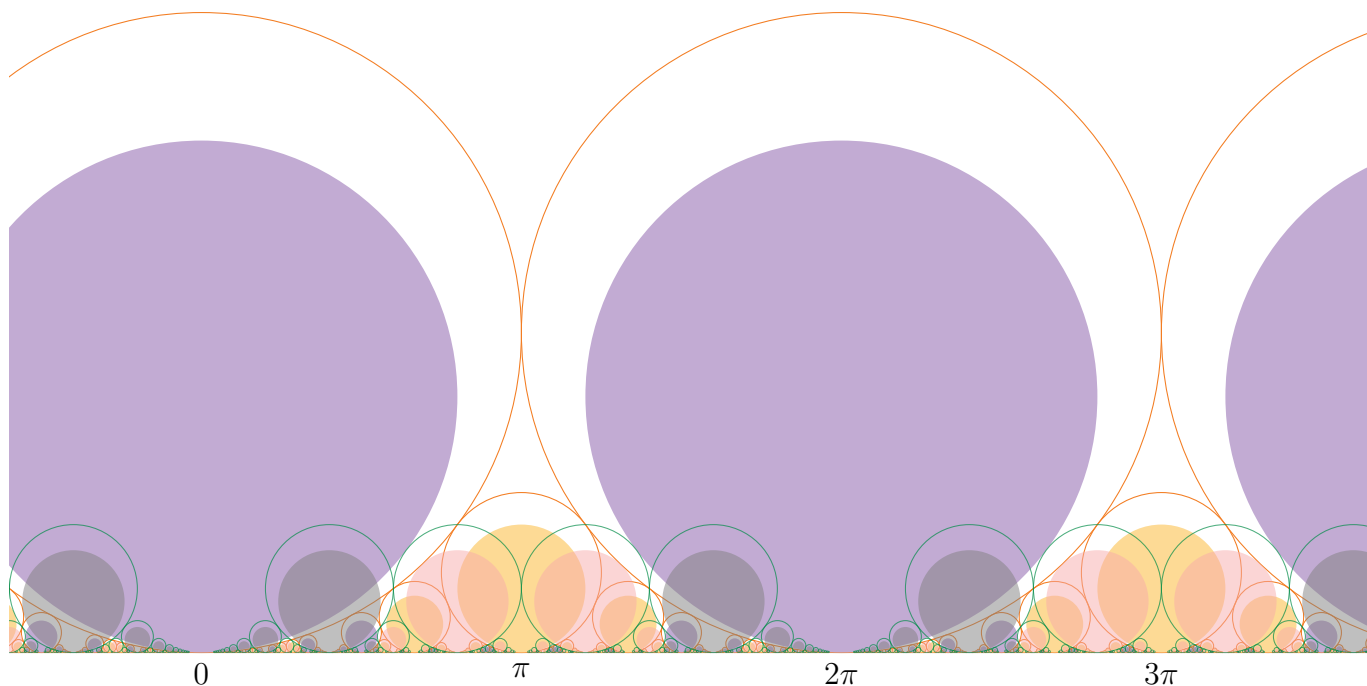
²²⁶Here again we need to consider F and f as *tensor* fields. (See Footnote 192 on p. 72.)

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. 79 coincide with positions of “omitted” Ford circles in the pattern of orange *circles*.

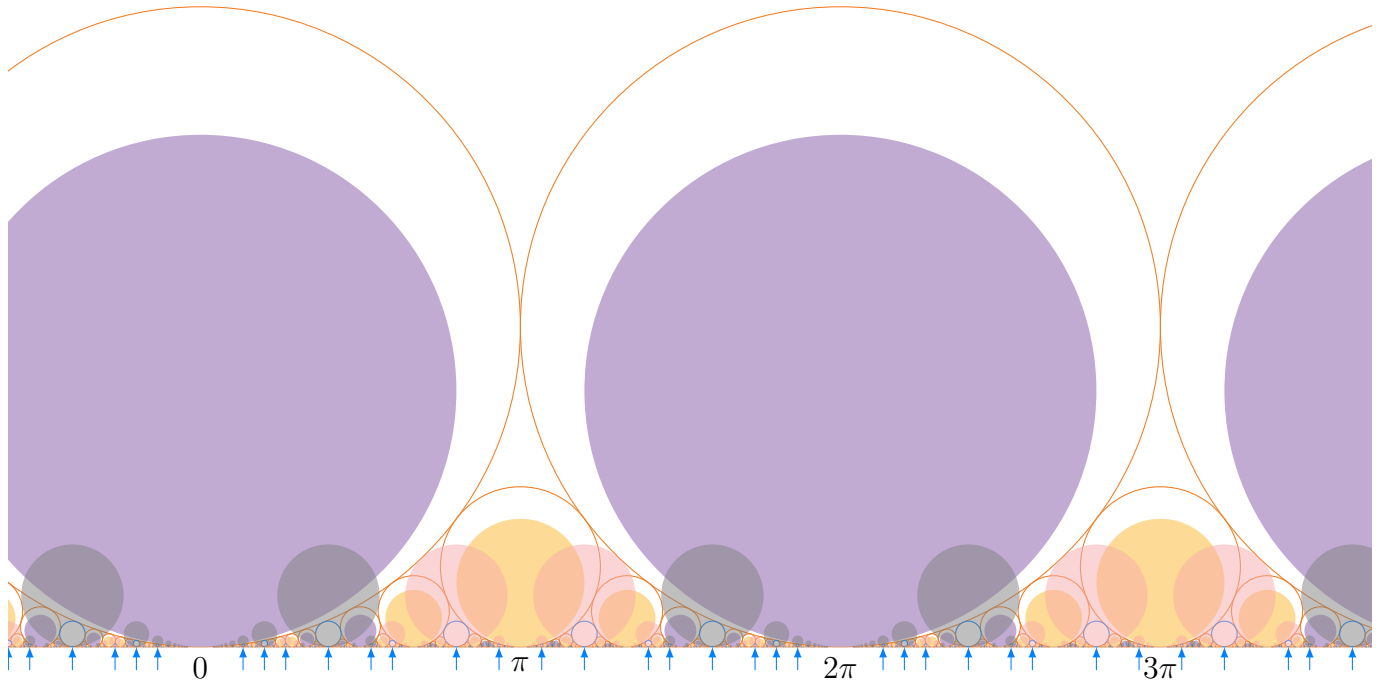
Overlaying the last two pictures on top of each other gives:



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. 79), 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:

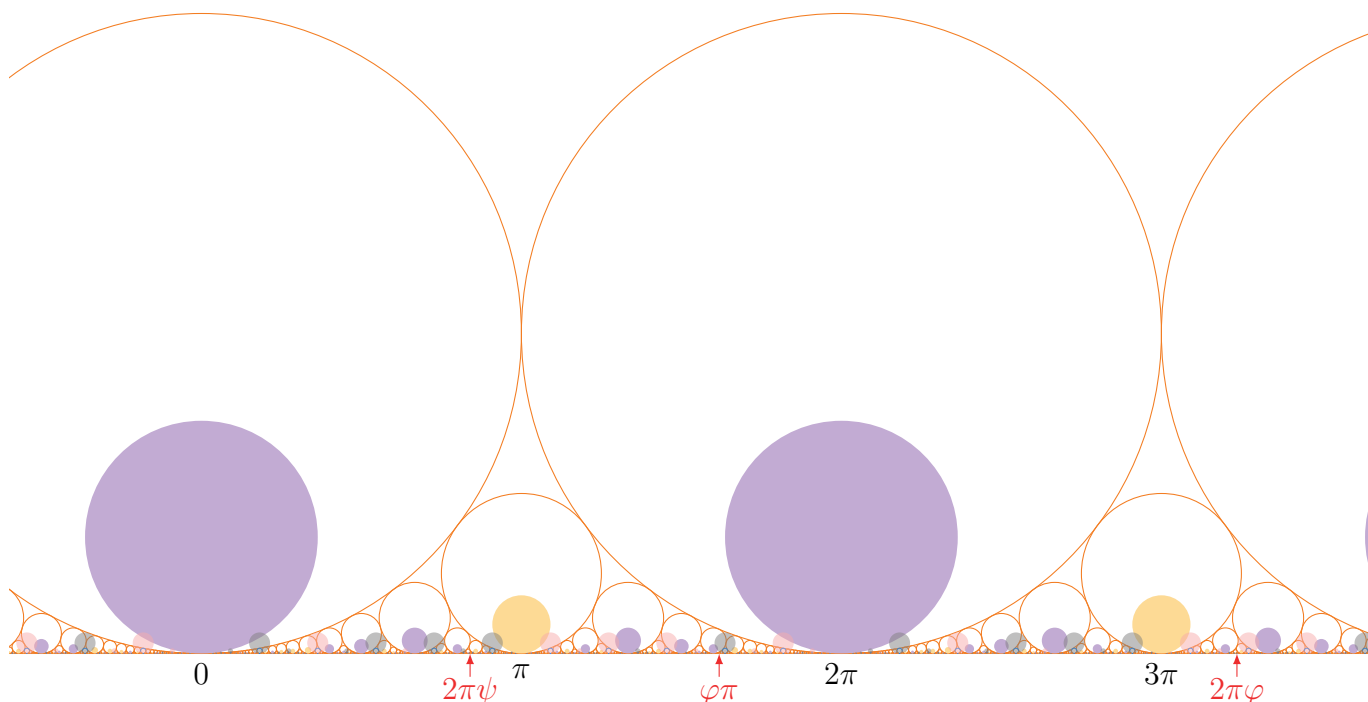
“Inscribe” smaller disks into the orange Ford circles. For every blue circle, “outscribe” a larger disk.

(The scaling factors are $4/c$ and 4 for orange and blue circles correspondingly. The color is chosen 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- $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$ 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 scarce (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 incircled in orange circles, gray and red are outcircled “in” (tiny!) blue circles. Four different colors of disks correspond to the types of symmetry described in the bullet list above.

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 form²²⁷ 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,²²⁸ 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 $c = 11$.^{229 230}

²²⁷It is still the same formulation as we had in Footnote 208 on p.75. 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 $|\alpha - R/D| \geq 2/c \cdot 1/D^2 \forall R, D$ works as an estimate for our exceptional set as well.

²²⁸It cannot be approximated by rationals with the required precision for $c > 2\sqrt{5} \approx 4.472136$.

²²⁹This is *almost* visible on the picture above—but one may need to zoom in *a lot*.

²³⁰Moreover, 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.²³¹

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 $\prod_{D=1}^d (1 - \varepsilon_D \varphi(D)/D^2)$ for the relative size of what is not covered by projections of disks with $D \leq d$.

This product decreases as $\text{const} \cdot d^{-6/\pi^2 \varepsilon}$. 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).²³²

²³¹Similar 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 mod 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.)

²³²In 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 $1/2$ 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 $1/2$.

Appendix²: Eisenstein series

In construction!

This part of the notes has very little to do with our principal aims. However, the situation uncovered in Remark 49 on p. 64 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 49 on p. 64, 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 very 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 $M \times \text{Tetrahedral number} + 1$ is proportional to the polynomial $(2n - 1)(4n^2 + 2n - 3)$, which can be rewritten with $k := 2n - 1$ as $k(k^2 + 3k - 1)$. Applying our recipes above for the numbers N_m (see the section on p. 59) *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. 113). (The latter case may require the rule from the section on 59.) One can immediately see that $N_m = \sum_{d|m} \left(\frac{d}{13}\right)$ when $m = p^k$; since both sides are “multiplicative” 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 $\lim_{s \rightarrow -0}$.

Our aim is²³³ to calculate the Fourier transform of the sequence σ_m . One can rewrite 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

$$\frac{1}{2} \sum_{d>0} \left(\frac{d}{13}\right) d^{s-1} \sum_{r \bmod d} \sum_{m>0} (\mathbf{e}^{im(t+2\pi r/d)} + \mathbf{e}^{im(t-2\pi r/d)}).$$

In this form, the complex conjugation replaces summation over $m > 0$ by $m < 0$; hence taking the real part gives (on $[0, 2\pi]$)

$$\frac{1}{4} \sum_{d>0} \left(\frac{d}{13}\right) d^{s-1} \sum_{r \bmod d} \left(\sum_m (\mathbf{e}^{im(t+2\pi r/d)} + \mathbf{e}^{im(t-2\pi r/d)}) - 2 \right) = \sum_{d>0} \left(\frac{d}{13}\right) d^{s-1} \sum_{0 \leq r < d} (\pi \delta(t - 2\pi r/d) - \frac{1}{2}).$$

²³³In fact, we came to this calculation “going backwards”: we took the conjectured formula for jumps in the function $F^{(-1)}(t)$ from p. 62, 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 $s < -1$)

$$\pi \sum_{d>0} \left(\frac{d}{13}\right) d^{s-1} \sum_{0 \leq r < d} \delta(t - 2\pi r/d) - \frac{1}{2} \ell_s.$$

Writing $d = Du$, $r = Ru$ with $u = (d, r)$, this becomes

$$\pi \ell_{s-1} \sum_{D>0} \left(\frac{D}{13}\right) D^{s-1} \sum_{0 \leq R < D, (R,D)=1} \delta(t - 2\pi R/D) - \frac{1}{2} \ell_s.$$

Periodic extension from $[0, 2\pi]$ gives

$$\pi \ell_{s-1} \sum_{D>0, R, (R,D)=1} \left(\frac{D}{13}\right) D^{s-1} \delta(t - 2\pi R/D) - \frac{1}{2} \ell_s.$$

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 .²³⁴ 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 numbers 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 the 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. 63 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 print²³⁵ in theorems on analytic dependence on parameters 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. 63 has a non-trivial jump there.

[There is a lot to say](#) about $s \geq -1$.

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 $p \neq 13$ 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 $\tilde{\sigma}_m := \tilde{\sigma}_m(s) := \left(\frac{m'}{13}\right) \sum_{d|m} \left(\frac{d}{13}\right) d^s$ with the same limit²³⁶ N_m when $s \rightarrow -0$.

To calculate Fourier transform, rewrite the factor $\left(\frac{m'}{13}\right)$. Let $\rho_l(m) := \sum_{v \bmod 13^l} \left(\frac{v}{13}\right) e(m \cdot v/13^l)$ for $l \geq 1$. 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.

²³⁴(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 $\lim_n n(\delta(t - 1/n) - \delta(t))$.)

²³⁵Which one???!?

²³⁶Another way to see this is to note that $\sum_{d|m} \left(\frac{d}{13}\right) = \sum_{d|m'} \left(\frac{d}{13}\right) = \sum_{d|m'} \left(\frac{m'}{13}\right) \left(\frac{d}{m'}\right) = \left(\frac{m'}{13}\right) \sum_{d|m} \left(\frac{d}{13}\right)$.

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 $\sigma_m(s) = m'^s \tilde{\sigma}_m(-s)$.

Indeed, write residues mod 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 mod 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

$$\delta_{l,k+1} \tilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{v \bmod 13^l} \left(\frac{v}{13}\right) \mathbf{e}(m \cdot v/13^l) \sum_{d|m} \left(\frac{d}{13}\right) d^s;$$

moreover, one may assume that $d|m'$. As above, we may rewrite the condition $d|m'$, getting

$$\delta_{l,k+1} \tilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{v \bmod 13^l} \left(\frac{v}{13}\right) \mathbf{e}(m \cdot v/13^l) \sum_d \left(\frac{d}{13}\right) d^{s-1} \sum_{r \bmod d} \mathbf{e}(m' \cdot r/d);$$

additionally, $\mathbf{e}(m' \cdot r/d) = \mathbf{e}(m \cdot r/13^k d)$. Hence one can rewrite this as

$$\delta_{l,k+1} \tilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{13 \nmid d} d^{s-1} \sum_{v \bmod 13^l} \left(\frac{vd}{13}\right) \mathbf{e}(m \cdot vd/13^l d) \sum_{r \bmod d} \mathbf{e}(m \cdot 13r/13^{k+1} d).$$

Additionally, the residues of $13^k r \bmod d$ are all distinct, so one can replace $13r$ by $13^{k+1}r$. Hence,

$$\delta_{l,k+1} \tilde{\sigma}_m = \frac{\sqrt{13}}{13^l} \sum_{13 \nmid d} d^{s-1} \sum_{v \bmod 13^l} \left(\frac{vd}{13}\right) \sum_{r \bmod d} \mathbf{e}(m \cdot (vd + 13^l r)/13^l d).$$

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 $l \geq 1$:

$$\tilde{\sigma}_m = \sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{R \bmod D} \left(\frac{R}{13}\right) \mathbf{e}(m \cdot R/D).$$

Calculating the real part of Fourier transform as above, we get (on $[0, 2\pi]$)

$$\sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{0 \leq R < D} \left(\frac{R}{13}\right) (\pi \delta(t - 2\pi R/D) - \frac{1}{2}) = \pi \sqrt{13} \sum_{13|D} D_{13}^{-s} D^{s-1} \sum_{0 \leq R < D} \left(\frac{R}{13}\right) \delta(t - 2\pi R/D).$$

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 $\tilde{\sigma}_m$ 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. 62 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 derivatives²³⁷ 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 parallelepiped 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 \mathcal{L} in \mathbb{R}^n and a bounded function α on \mathcal{L} , the sum $\sum_{\tau \in \mathcal{L}} \alpha(\tau)(\Psi(\tau) - \int_{\square} \Psi(\tau + \tau') d\tau' / |\square|)$ converges absolutely for $d - 2 < -n$; here \sum° means skipping²³⁸ τ with $|\tau| \leq |\tau_0|$. 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 \mathcal{L}' -periodic (here \mathcal{L}' is a sublattice of \mathcal{L}) with average 0. Take a centrally-symmetric collection U of representatives of all translations $\tau + \mathcal{L}'$ of \mathcal{L}' inside \mathcal{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 $\sum'_{\tau \in \mathcal{L}} \sum_{\tau_0 \in U} \alpha(\tau + \tau_0)\Psi(\tau + \tau_0)$ 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 $\sum_{\tau \in \mathcal{L} \setminus U} \alpha(\tau)\Psi(\tau)$ (which converges absolutely).

Conclusion: the former summation method (with added $\sum'_{\tau \in U} \alpha(\tau)\Psi(\tau)$) gives a generalization of summing $\sum'_{\tau \in \mathcal{L}} \alpha(\tau)\Psi(\tau)$: it gives correct answers for $d < -n$, and makes sense on a larger set $d - 2 < -n$ of degrees d .^{239 240}

Remark 63: An important related question is the possibility of analytic continuation when the sum is restricted to points $\tau \in \mathcal{L} \cap C$; here C is a cone with the vertex at 0.²⁴¹ When one restricts summation to shifts $\tau + U$ of U with $\tau \in \mathcal{L}'$ 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

²³⁷In fact, the arguments below work also when the first derivative is Lipschitz with the constant M , and the function is bounded by M .

²³⁸We need to skip such values since Ψ is not defined at the origin (at least when d is negative). In what follows, skipping $0 \in \mathcal{L}$ 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.

²³⁹One can also consider complex d . Then the conditions are $\operatorname{Re} d < -n$ and $\operatorname{Re} d - 2 < -n$.

²⁴⁰Likewise, 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 .

²⁴¹The 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 dependence on this “method” disappears 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 interested 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.²⁴²

²⁴²Indeed, 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 $\varphi(b', b)$ 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 .²⁴³

However, consideration of the “remaining” terms is more delicate.

Remark 64: 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 $\tau \in \mathcal{L}'$ 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.²⁴⁴ Assume that the kernel of Π is spanned by a vector l_0 in the lattice \mathcal{L}' . Then the image \mathcal{L}° of \mathcal{L}' under this projection may be lifted back to \mathcal{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 \mathcal{L}° .

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 $\sum_{\tau \in \mathcal{L}^\circ} \xi(\beta(\tau))\Psi(\tau)$ (or similar sums over polyhedral cones in $\mathcal{L}^\circ \subset \mathbb{R}^{n-1}$); here β is a linear function on \mathcal{L}° , 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 \mathcal{L}° satisfy $d < 1 - n'$.

Note that when $\xi \circ \beta$ is periodic (and automatically odd) on \mathcal{L}° , the argument in Footnote 240 implies the required analytic dependence. This happens when the hyperplane has a normal in the dual lattice to \mathcal{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 \mathcal{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 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 $\sum'_{\tau \in \mathcal{L}} \alpha(\tau)\Psi(\tau)$ 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 sufficiently quickly when $R \rightarrow \infty$. This immediately implies that $\sum'_{(R,D)} \alpha(R, D)D^{s-1}\delta(t -$

²⁴³If 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.)

²⁴⁴In 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

$$\sum'_{(R,D) \in U} \alpha(R, D) D^{s-1} \delta(t - R/D) + \sum'_{(R',D') \in \mathcal{L}'} \sum_{(R,D) \in U} \alpha(R' + R, D' + D) (D' + D)^{s-1} \delta\left(t - \frac{R' + R}{D' + D}\right).$$

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 \mathcal{L}_1 . In particular, the third antiderivative is a well-defined absolutely continuous function.

Remark 65: 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 value²⁴⁵ 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 $|\sum_{m \leq M} \xi(m\gamma)|$ grows not quicker than $M/\log^2 M$, then $\sum_m \xi(m\gamma)/m^s$ converges for $s \geq 1$. In fact, for the aim of analytic continuation, it is enough if the “ M -summation” $\sum_M 1/M^2 \sum_{m \leq M} \xi(m\gamma)$ converges absolutely, and that $1/M \sum_{m \leq M} \xi(m\gamma) \rightarrow 0$. 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, assume²⁴⁷ $|\gamma - A/B| < 1/B^2$, 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 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 estimate²⁴⁸ $|\sum_{m \leq M} \xi(m\gamma)| \leq 3K_M$ 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 $K_M \leq \sum_{k \leq l+1} a_k$; here a_k are coefficients of the continued fraction of γ . This may be improved to $K_M \leq \sum_{k \leq l} a_k + M/Q_l$.

Typically, the sequence Q_l grows much quicker than a_l . Let $A_l = \sum_{k \leq l} a_k$; 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 $\sum_{Q_l \leq M < Q_{l+1}} (A_l + M/Q_l)/M^2$. The term A_l contributes at most $A_l/Q_l = \sum_{k \leq l} a_k/Q_l$; summing such terms over l , the number a_k comes with a coefficient $\sum_{l \geq k} 1/Q_l$; since $Q_{k+2} > 2Q_k$, this coefficient is bounded

²⁴⁵For 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.²⁴⁶

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.

²⁴⁶**N.B. Is it???**

²⁴⁷As in Dirichlet's approximation theorem.

²⁴⁸For 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 $\int_1^{a_{l+1}+1} 1/Q_l^2 \pi Q_l d\kappa = \log(a_{l+1}+1)/Q_l$. **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 $\tau \in \mathcal{L} \cap C$ “makes sense” for $s = 0$. Moreover, one can find it by

- For every $\tau' \in \mathcal{L}'$ add together the terms corresponding to points of $\mathcal{L} \cap C$ 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 compatible²⁴⁹ 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 $|C|$). First²⁵⁰ restrict attention to the rational slopes of the boundary rays. To avoid poles of analytic continuation (see Footnote [245](#)), assume that the average value of α on any hyperplane in \mathcal{L} is 0.²⁵¹

One can break the sum into 4 parts (below U^τ is the translation $\tau + U$ of U with $\tau \in \mathcal{L}'$) running over:

- The U^τ s fully contained inside C .
- The U^τ s which are bisected by the second ray, but not the first one.²⁵²
- The $U^\tau \setminus C$ 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.

²⁴⁹Of course, the ray separating the angles should be included in one of the angles only.

²⁵⁰**N.B. Check???**

²⁵¹Note that this is not very restrictive. For example, if \mathcal{L}' is of a prime index in \mathcal{L} , this adds the condition $\alpha(0) = 0$.

²⁵²Here we sum over the “whole” U^τ — as opposed to $U^\tau \cap C$.

However, the latter estimate does not ensure continuity, 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 $\max_M \left| \sum_{m \leq M} \Xi(m\gamma) \right| / Q$, here Q is the length of the period (the denominator of γ)²⁵³. 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 appropriate 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 $(1/N + 1/Q) \cdot (v + 2 \max_M \left| \sum_{m \leq M} \Xi(m \cdot r/q) \right|)$ for $\max_M \left| \sum_{m \leq M} \Xi(m\gamma) \right| / Q$. (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 $\tau \in \mathcal{L}$ 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 \mathcal{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 \mathcal{L} going in the direction γ_1 .

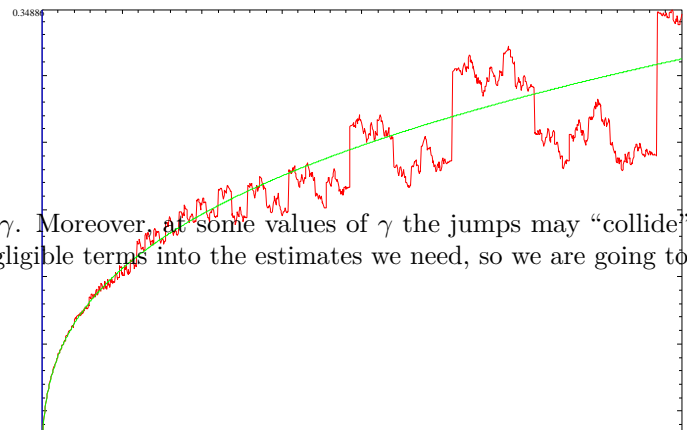
Together with additivity, this gives a partial description of the behaviour of the sum in an open angle C when one varies one of the rays \mathcal{R} of the angle. Call an \mathcal{L} -rational direction *admissible* if the average of α on the line of this direction through 0 vanishes; call it *strongly admissible* if the same holds for all translations of this line. Restrict attention to angles with admissible direction of \mathcal{R} ; then near a strongly admissible direction, there is a jump of $\sum_{\tau \in \mathcal{R}} \alpha(\tau)\Psi(\tau)$, 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 $\mathcal{L} \subset \mathbb{Z}^2$ 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. 61. (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 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 on the right 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



²⁵³Note 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.

the “expected” jumps at points p/q with $13 \nmid p$, of magnitude $\sqrt{13}/q_{13}^s q^{1+s} \binom{p}{13}$. In general, the curve to fit when $t \rightarrow +0$ is²⁵⁴ $C_s \cdot t^{-s}$. When $s \rightarrow 0$, the coefficient C_s goes to $1/2$, and y goes to $1/2$ for positive t . Since the function is odd, with $s = 0$ we get a jump of 1 at $t = 0$.

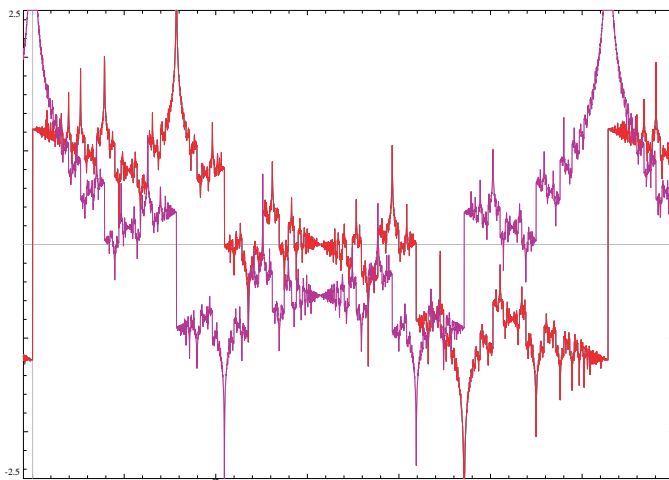
Conclusion: the spurious jumps appear only when $s = 0$. For other values of s , one gets “less confusing” power-law singularities at the positions of spurious jumps.

Remark 66: 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. 61 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 324 on p. 122). 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” graphs²⁵⁵ 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. 14).

In other words: in this context, the Maass case is much more interesting than the “case of modular forms”.



²⁵⁴Heuristically, this may be explained by the last identity of Footnote 236. 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 $\tilde{\sigma}(-s)$. 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 $\tilde{\sigma}(-s)$, 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 $\tilde{\sigma}$ 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 $12/(13 - 13^s) \approx 0.9543$.)

²⁵⁵As opposed to simulations made by interpolating from a small collection of values of t .

Appendix: On verification, — and the future

In construction!

We start with a section we could not find place elsewhere.

More details on the M -family

On p. 16, we introduced the M -family of polynomials “ $M \cdot \text{tetrahedral numbers} + N$ ” (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.²⁵⁶

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*”²⁵⁷ 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 231 on p. 84 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 $2^{1/3} \cdot 2^{-K}$. 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.)

Remark 67: 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 $-1,228$.

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.²⁵⁸

²⁵⁶Here 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.

²⁵⁷Given 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.)

²⁵⁸Indeed, by Class Field Theory one should start with subgroups of index 3 in $(\mathbb{Z}/m)^\times$; exclude subgroups induced by surjections $(\mathbb{Z}/m')^\times \rightarrow \mathbb{Z}/3$ 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.

The simplest coincidences of non-cyclic extensions are related,²⁵⁹ by Class Field Theory, to having more than one subgroup of index 3 in the Class Group $\text{Cl}(\mathbb{Q}[\sqrt{d}])$. (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 68: 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.²⁶⁰

On gamma-factors

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 of 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 $\widetilde{N}_n^{\text{res}}$ (or better, $\widetilde{N}_n^{\text{Gal}}$ which needs very little massaging; see p. 46). 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. 66). This symmetry sends $G(T)$ to $\bar{G}(-1/cT/T)$. 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 $\widetilde{N}_n^{\text{Gal}}$ of a polynomial equation in 1 variable is

- There is a function $\varkappa_{d,r_1}(s)$ defined for $s > 0$ and depending only on the degree d and the number of real roots r_1 of the equation;
- There are numbers $c \in \mathbb{N}$ and C such that the sum $K(s) := \sum_n \widetilde{N}_n^{\text{Gal}} k(ns)$ is symmetric: $K(-1/cs) = C \cdot K(s)$.

What Langlands program predicts is that a similar thing works for any system of polynomial equations — at least after a suitable purification (which may be much less trivial than what we did above).

²⁵⁹Unfortunately, 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.)

²⁶⁰Indeed, 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.)

Examples: degree 4

The algorithms used for plots in this section are not fully tested yet!

As we explain in the following section, while in degree 4 it is not a problem to produce the sequence N_n , there is no reason to expect that *the same* patterns of fractality will still work for its Fourier transform $F(t)$. Here we provide the plots which show that


- Indeed, the graphs *do not* look like a toy transform of a *periodic* function.
- However, the graphs show *some* of characteristic features of the toy transform (the “hourglass” shapes ).

I have no idea how to explain the appearance of these features. Below, the plots are provided without any explanation! (Our magnified views are *not* at points in the Cantor hyper-family — since the behaviour near such points is known to be governed by the “actual” fractality law.)

Since discriminants of indecomposable polynomials of degree 4 cannot be very low, it is not computationally feasible to show what *would be* a region of horizon-self-similarity in the case of degree 3 (with the same field discriminant). Instead, we proceed as on p. 55 and show what *would be* a non-trivial²⁶¹ region of *just* horizon-similarity (in fact, it would be a horizon-similarity to $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$; see Footnote 136 on p. 55).

Since we do not expose fractality here, we do not mention “conductors”. Instead, we use *field discriminants*: for non-abelian case of degree 3, the conductor was equal to the field discriminant.²⁶²

When inspecting the graphs below, it makes sense to pay attention to:

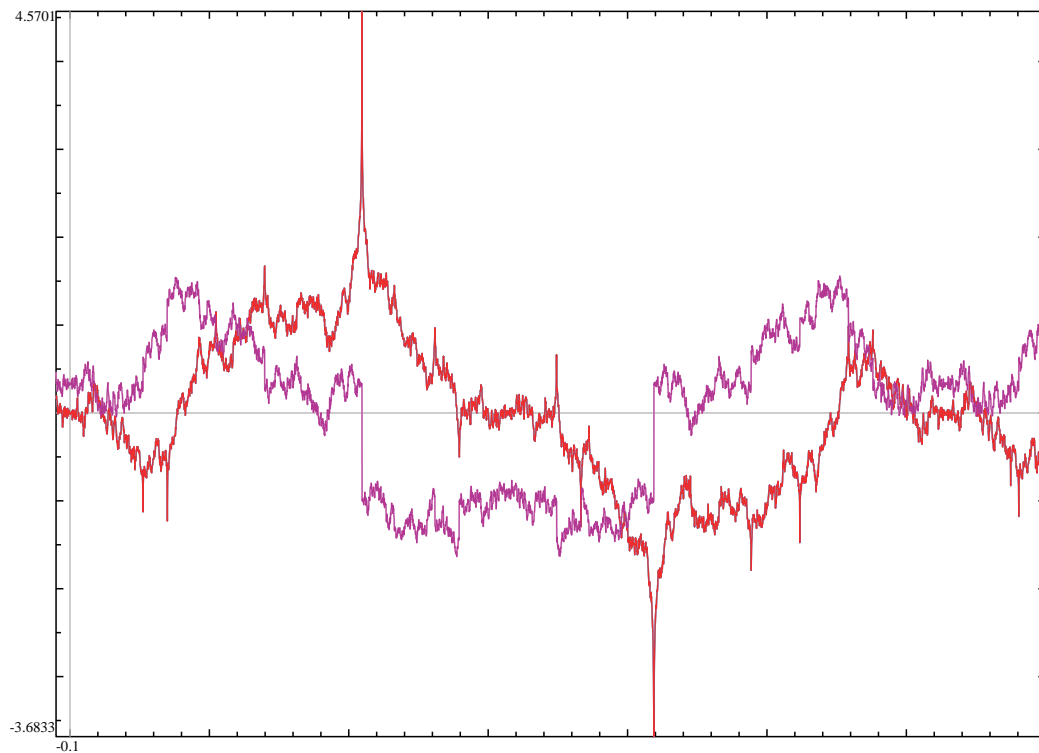
- The appearance of the “hourglass” shapes .
- The hourglasses near rational multiples of π *do not* look like toy transforms of *periodic* functions: the “shape” of oscillations changes when we get closer to the waist.
- Hence the function is not horizon-similar (and definitely not self-similar). However, the shape being “a hourglass” shows that it is a toy transform of a function oscillating *between two values*.
- Absence of jumps and/or log-spikes (we saw them on our Eisenstein plots on p. 60) indicates that what we have may be related to something *cuspidal* (see Footnote 162 on p. 64).

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 pure (it breaks into two, of ranks $1 + 2$; compare with Footnote 276 on p. 109), so it is not surprising that the corresponding graphs of $F^{(-1)}(t)$ “change via jumps”! About one period of the real and imaginary

²⁶¹“Trivial” points of self-similarity are those in Cantor hyper-family. If function has extra symmetries (as in the section on p. 41), the images of trivial points under these symmetries are also “trivial”.

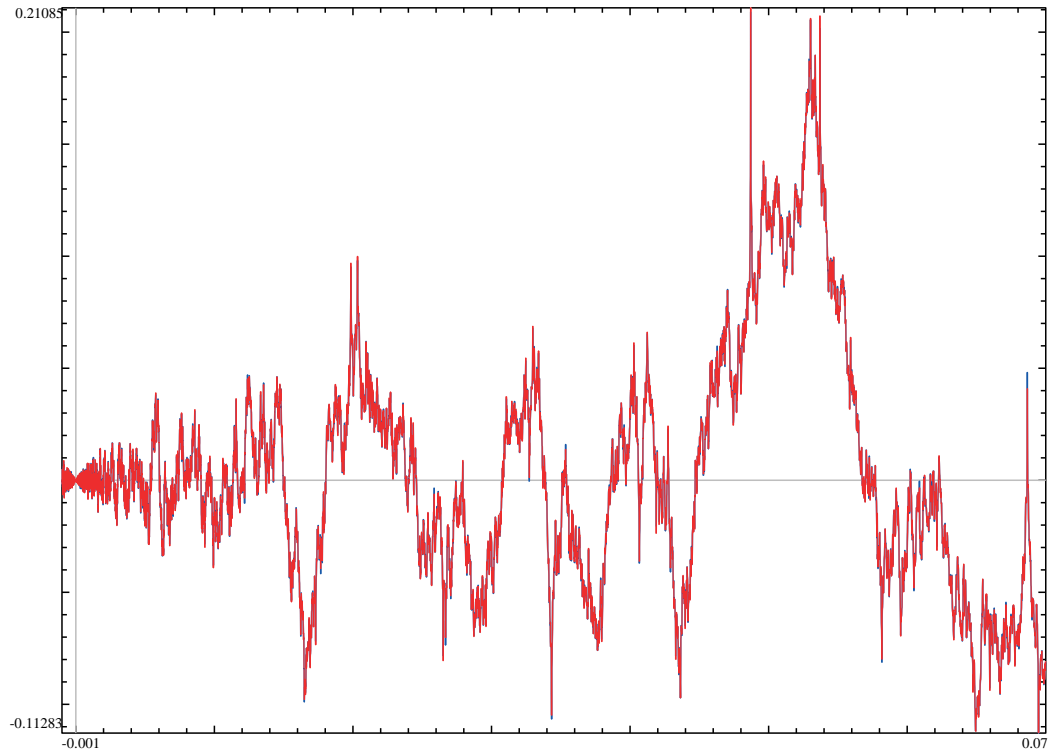
²⁶²Moreover, this number governs fractal transforms in the “trivial” points in the same way as for degree 3. Indeed, the Hecke’s functional equation works for any degree; it implies horizon-similarity (“self” or to $\text{Im } F_{\mathbb{C}}^{(-1)}(t)$) in “trivial points” (see p. 67).

parts:

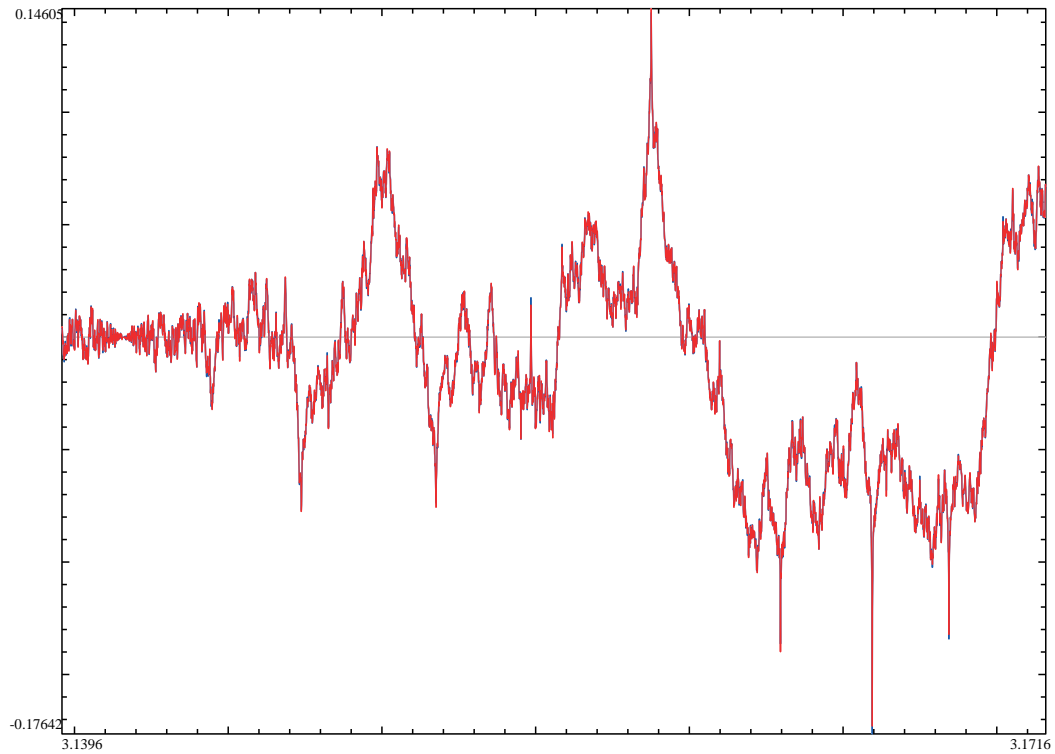


The real part has log-spikes, while the imaginary parts has jumps (this is similar to what we already saw in other examples of non-pure motives). Note that this is a much more complicated example than the non-pure motives we saw before: it mixes a motive with "periodic" N_p (for prime p) with a "modular form" motive.

Here is how it looks²⁶³ near 0:



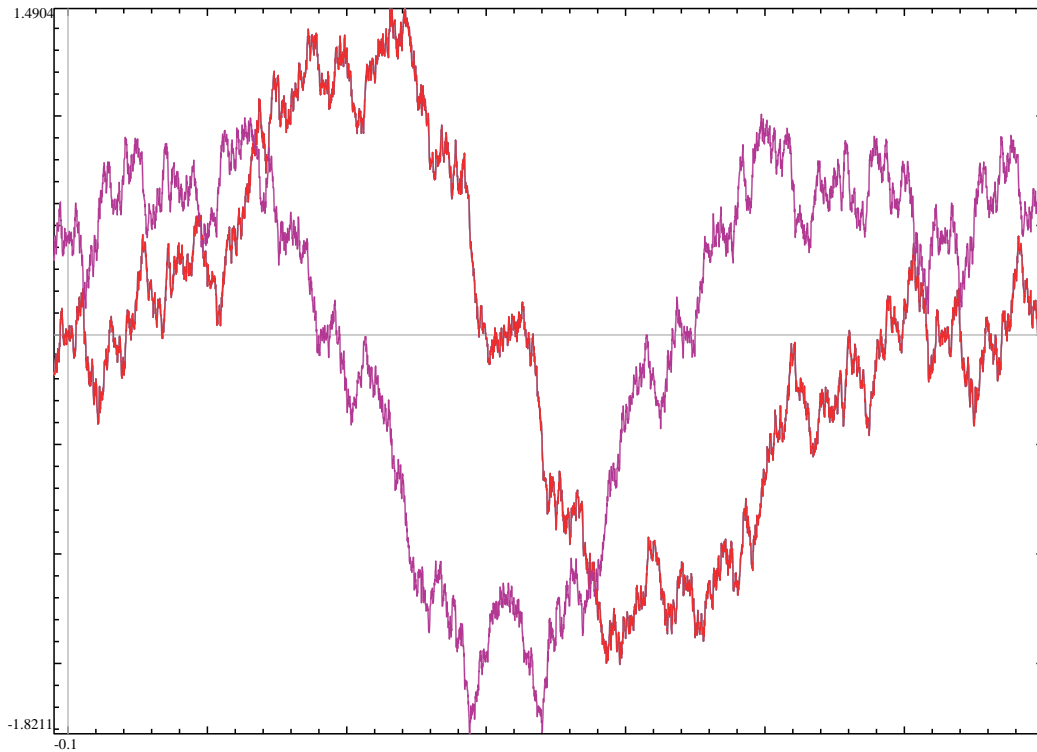
Here is the magnified view near $t = \pi$:



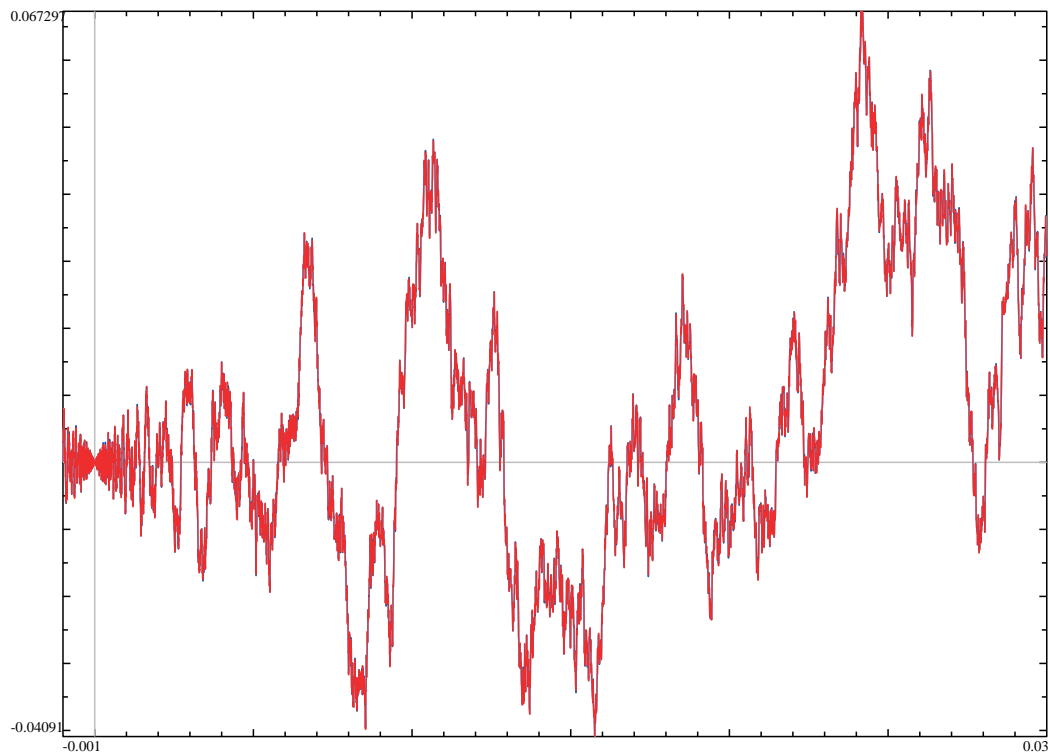
²⁶³A more careful examination shows that the top/bottom asymptotic near 0 are not linear, and follow the $C|t \log(t/K)|$ -law we discuss below for field discriminant $D = 229$.

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 a *bounded oscillating* function! (Another interesting feature is the presence of noticable “spikes” even that close to $t = \pi$.)

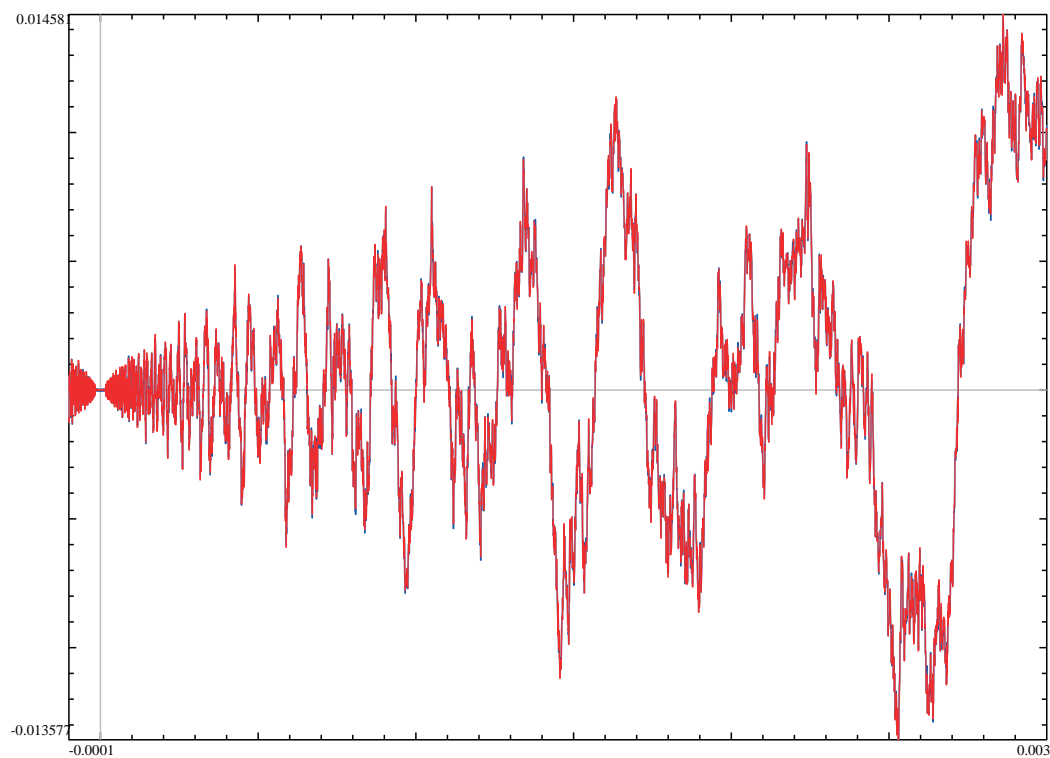
Our next example is the polynomial $x^4 - x + 1$ with the smallest magnitude of the field discriminant for the case of a pure motive: $D = 229$ (which is a prime number). It again has no real roots, and the Galois symmetries form the symmetric group S_4 . Observing about one period of the real and imaginary parts:



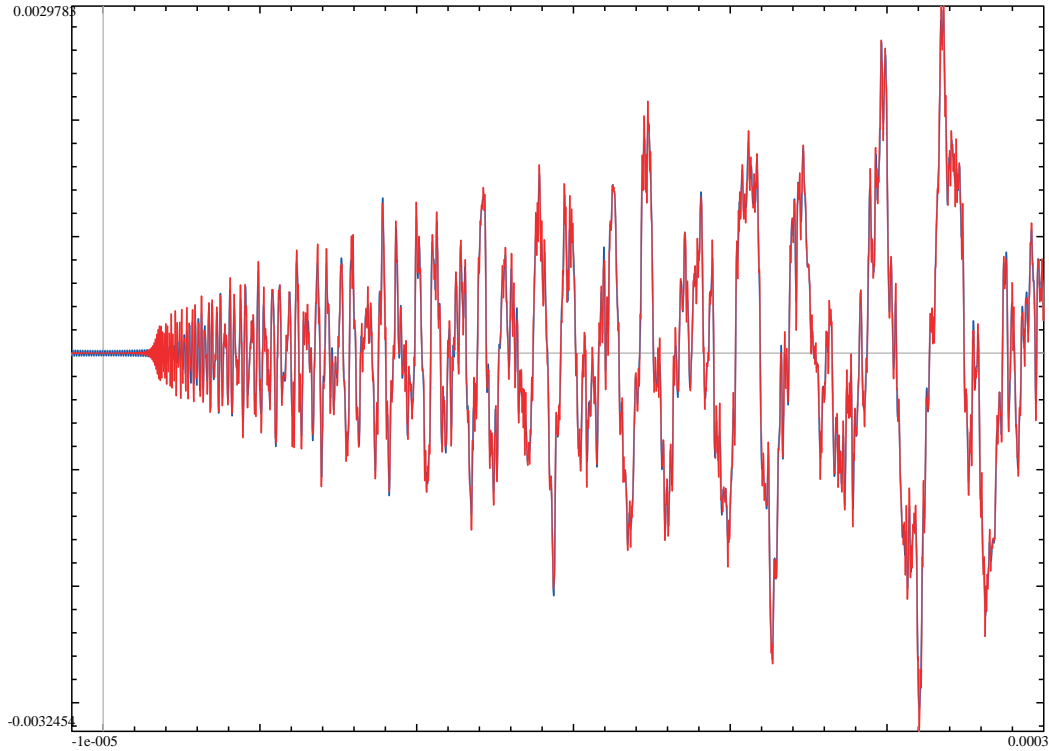
shows neither jumps/spikes, nor other types of discontinuity. When we look near 0:



we again can see a “hourglass”: the behaviour resembling two (straight?) top- and bottom-asymptotes as $t \rightarrow \pi + 0$. Moreover, zooming in 10 times shows that these “asymptotes” become steeper when we get closer to the “waist”:

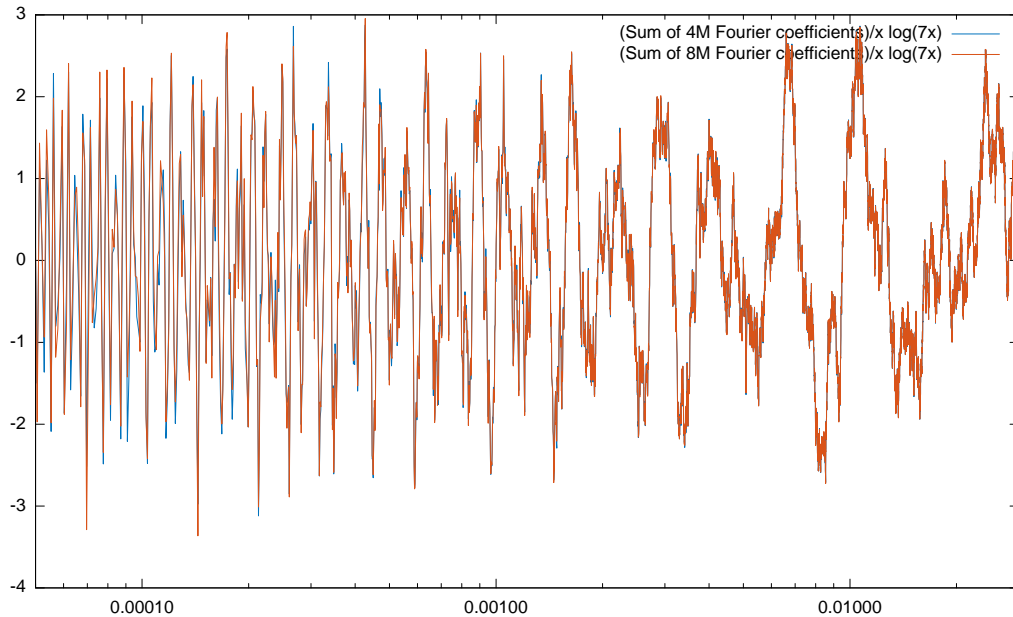


(So we can start to suspect that the “hourglass” is actually non-linear!) Zoom in 10 times more:



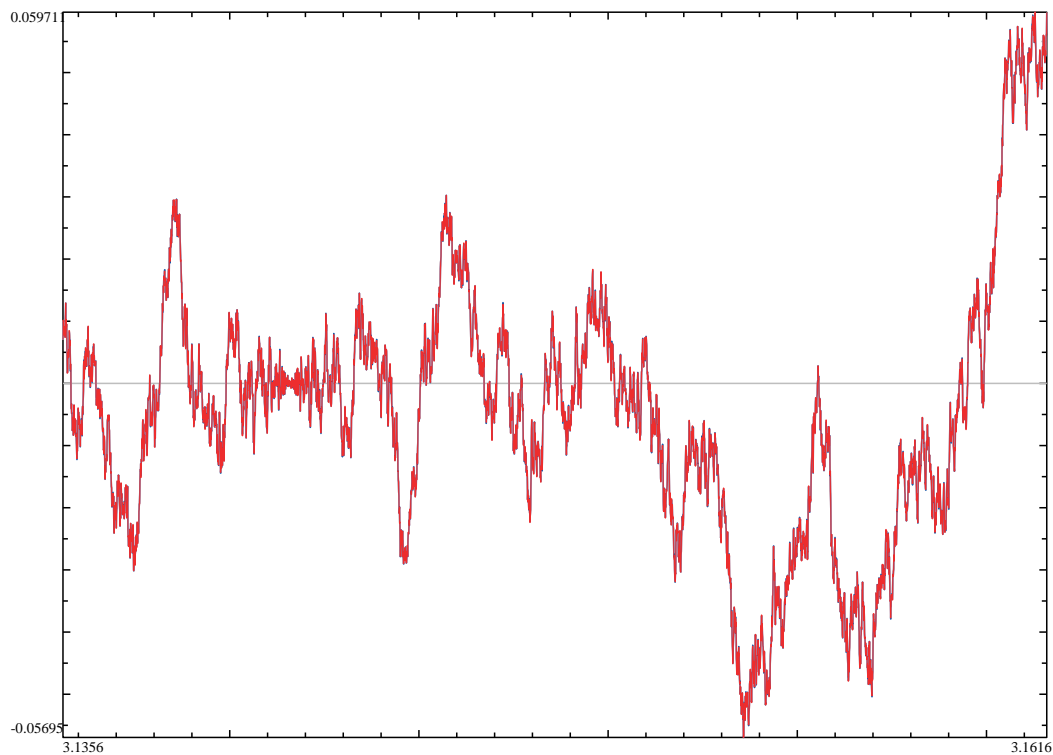
Here the “hourglass asymptotes” look almost linear now.²⁶⁴

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 in the part on the left is abysmal):

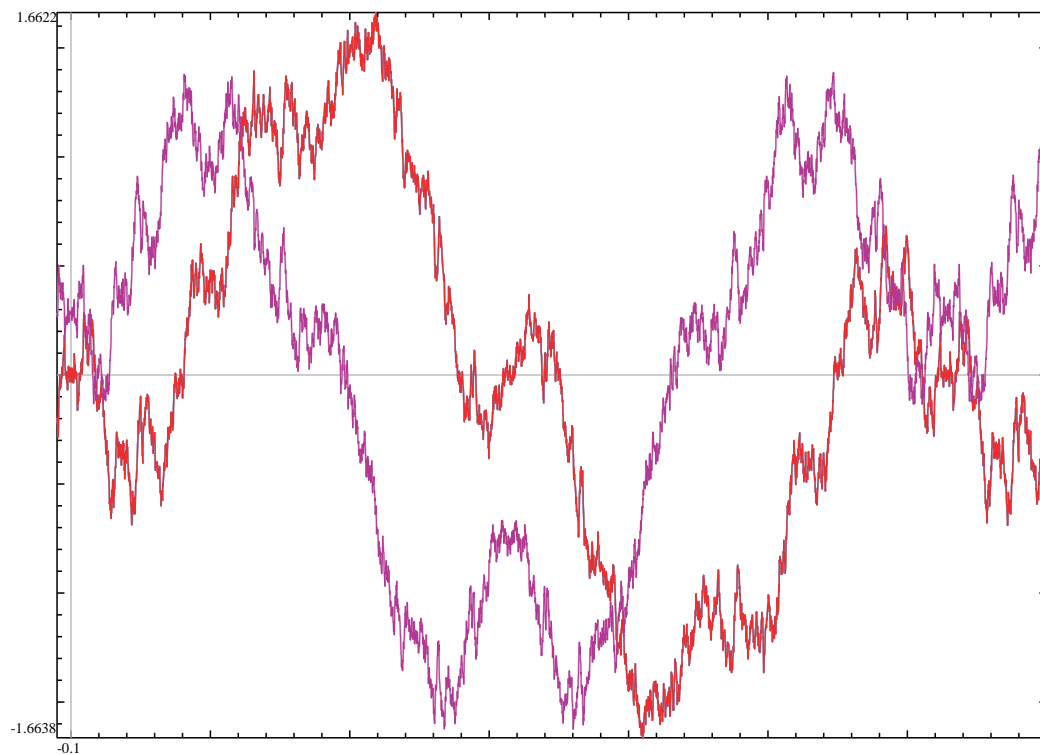


²⁶⁴However, a flattened zone becomes very visible. See Footnote 85 on p. 38.

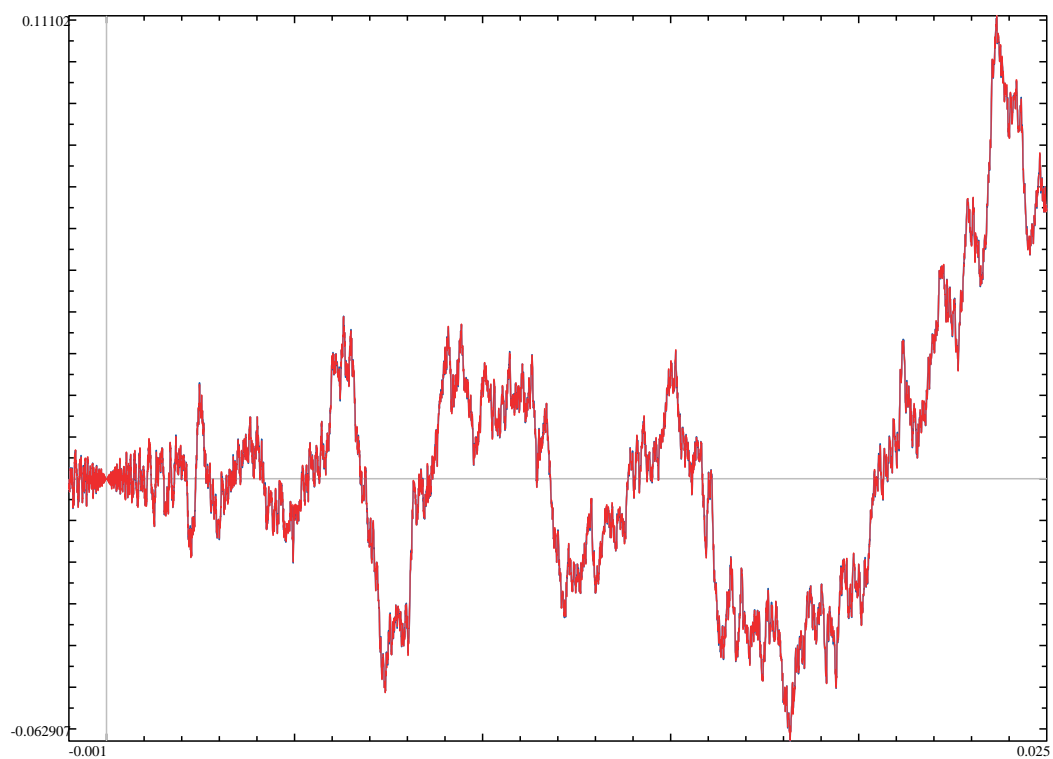
One can check that near $t = \pi$ very similar effects appear (we do not show the pictures with higher zooms — but they behave as above):



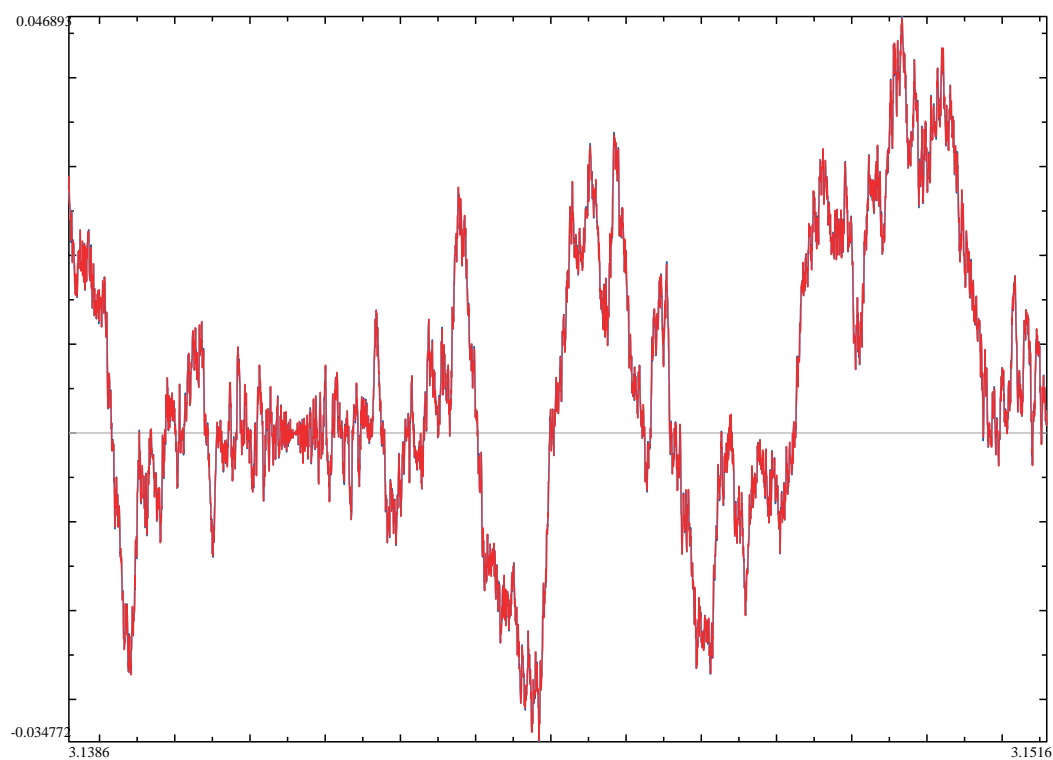
The minimal magnitude of a negative field discriminant with a pure motive is $D = -283$ (which is a prime number). The polynomial is $x^4 - x - 1$ with two real roots and the Galois group S_4 . The plots as above still show no visible discontinuities:



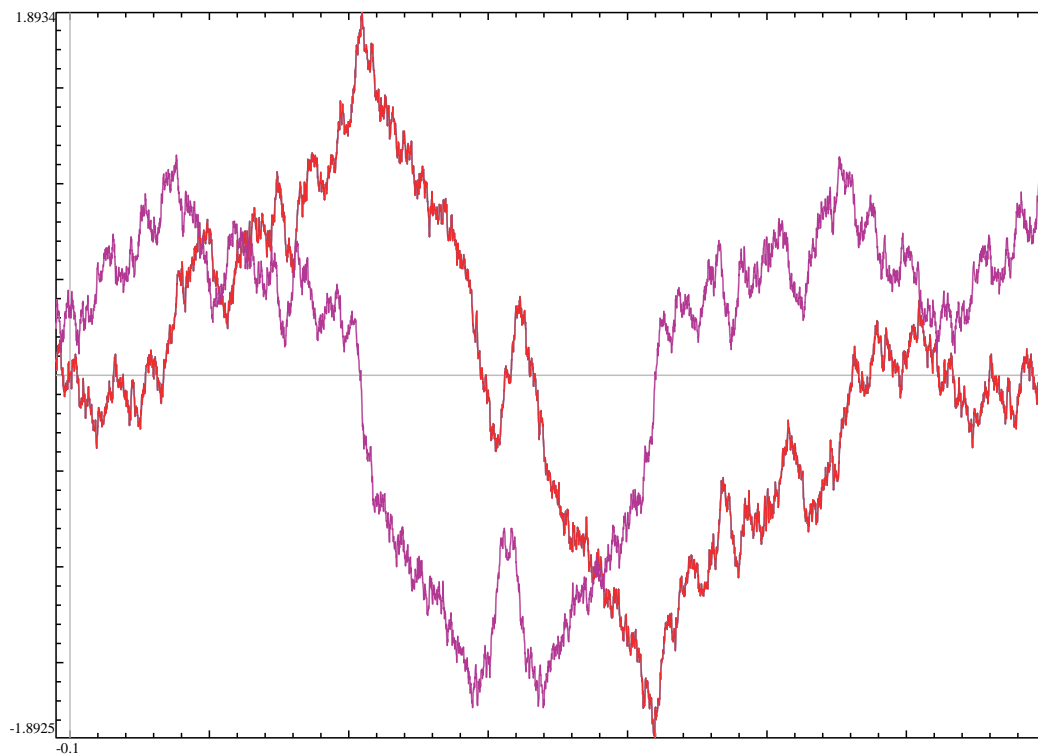
Here is how it looks near 0:



Near $t = \pi$ one can see yet another “irregular hourglass”:

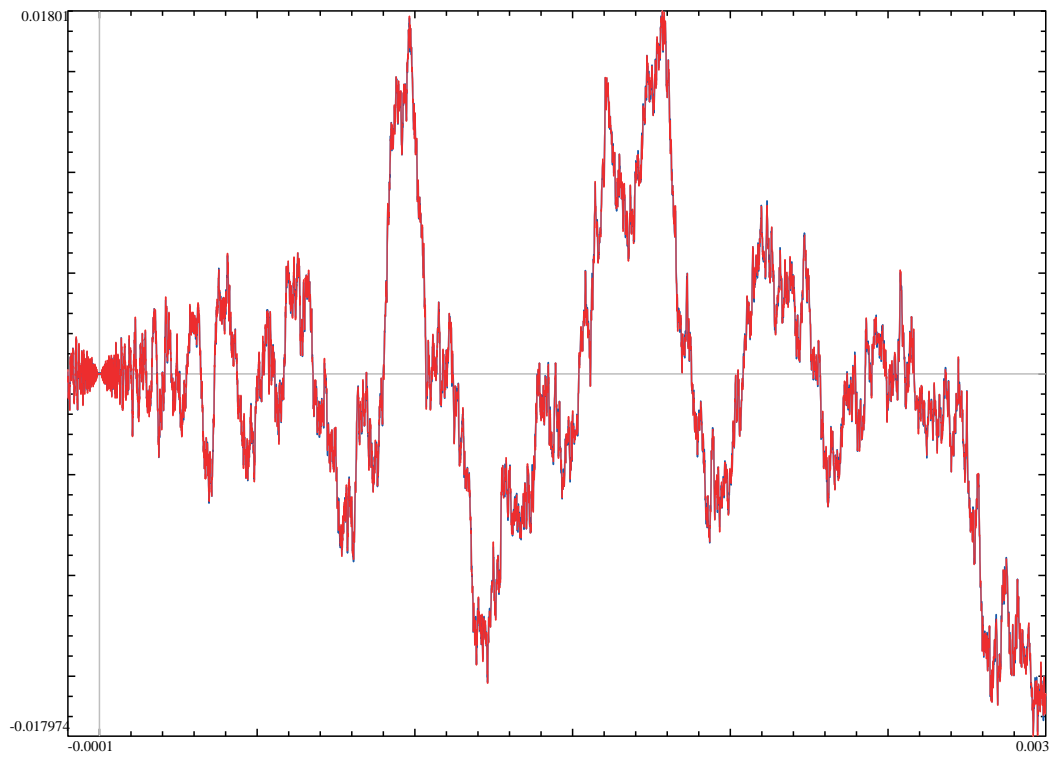


Next, consider $x^4 - 4x^2 - x + 1$; it has 4 real roots, and the smallest field discriminant for such a case of a pure motive²⁶⁵: $D = 19 \times 103 = 1,957$. Its Galois group is S_4 . Here is one period of the real and imaginary parts:

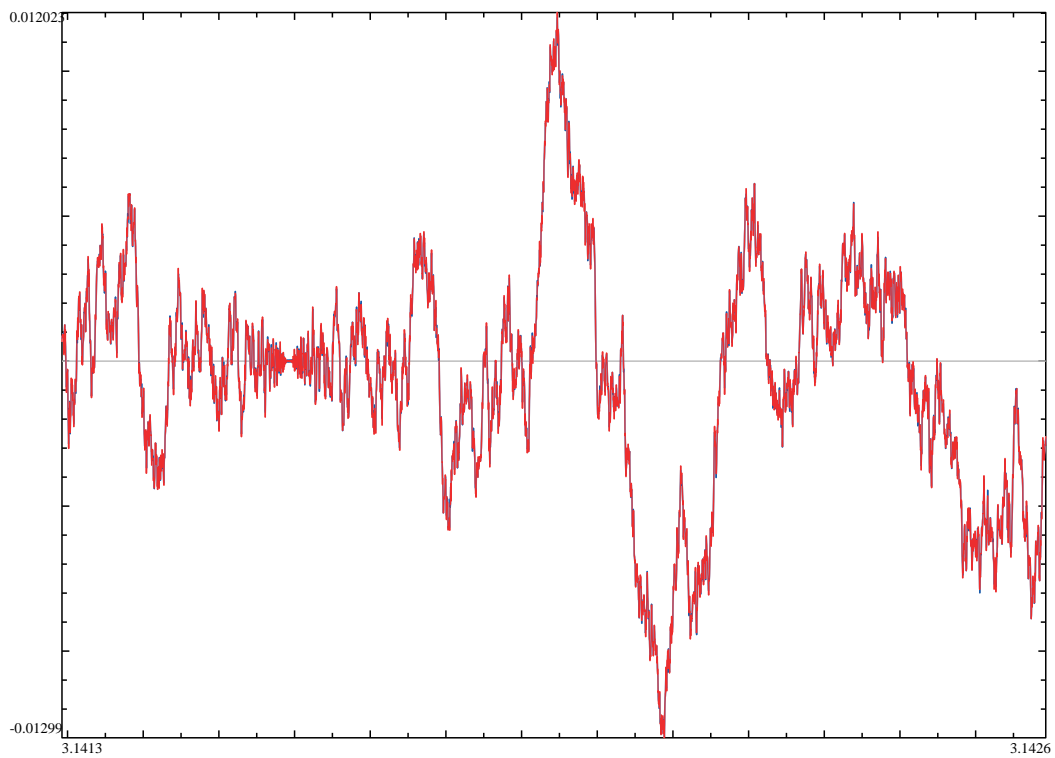


²⁶⁵Moreover, another measure of complexity, the narrow class number (hence class number) turns out to be trivial:
1. So this example is “the simplest one” in all the possible senses.

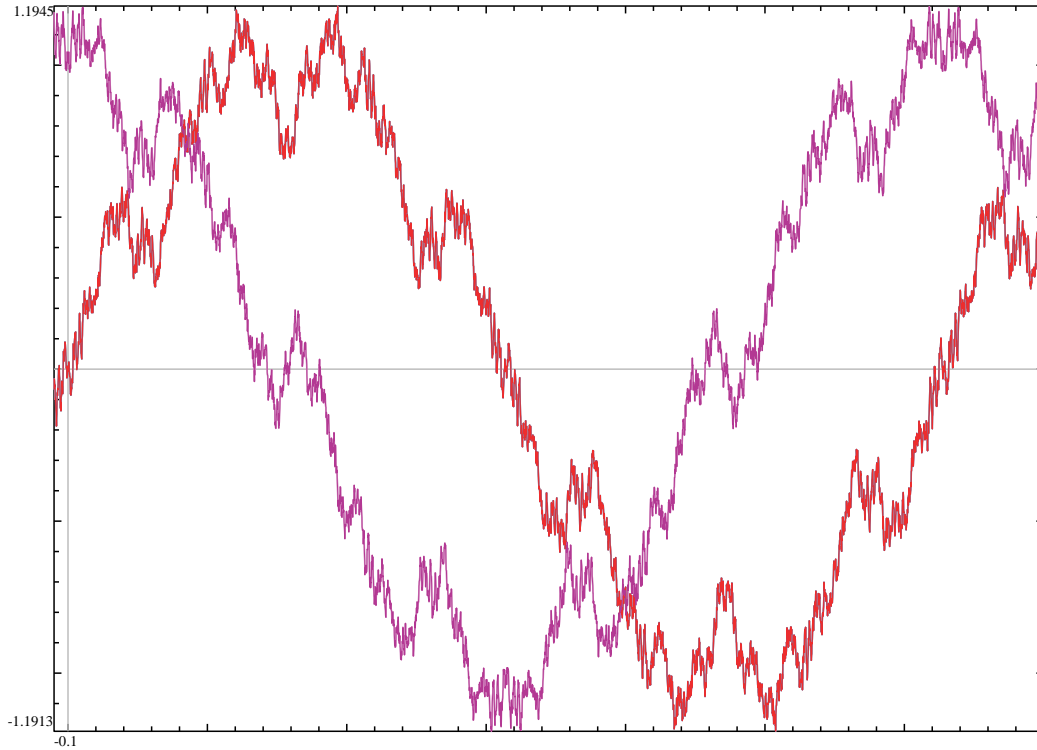
Near $t = 0$ it still shows no horizon-similarity:



The behaviour near $t = \pi$ should not be surprising now:

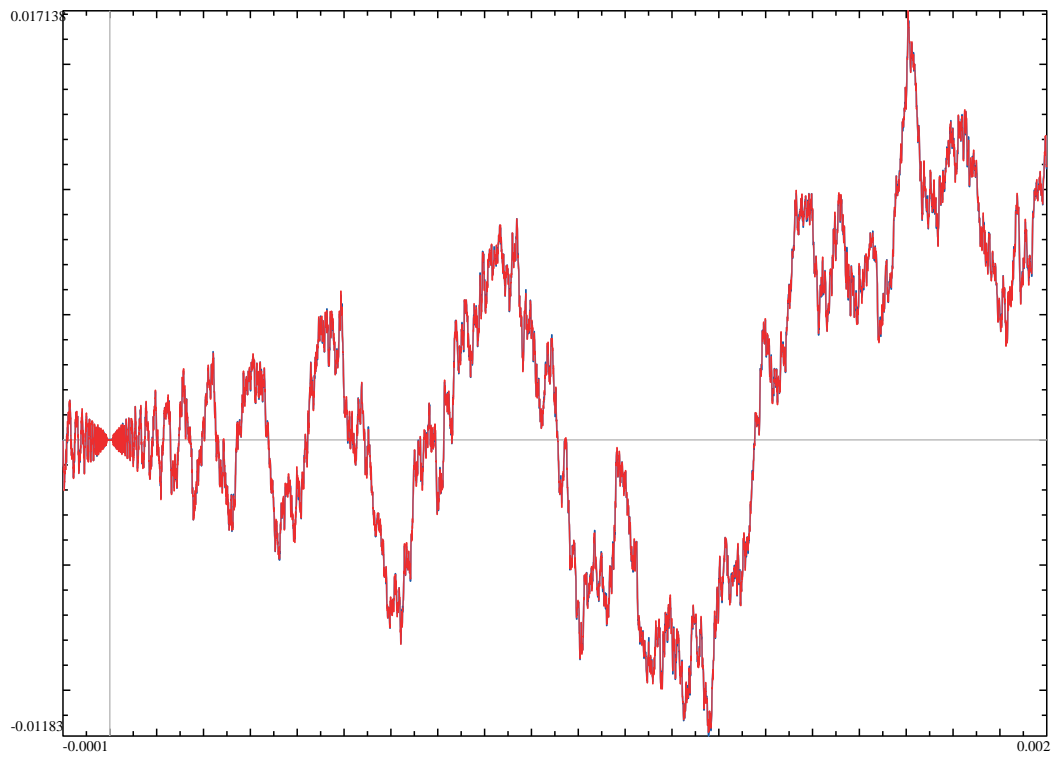


Finally, consider $x^4 - 2x^3 + 2x^2 + 2$, with the smallest field discriminant which is a square, and leads to a pure motive: $D = 56^2 = 3,136$. It again has no real roots,²⁶⁶ and (since the discriminant is a square) the Galois symmetries form the alternating group A_4 . The graph of about one period of the real and imaginary parts shows that the plot has “extra symmetries” (as in the section on p. 41): it is π -antiperiodic:

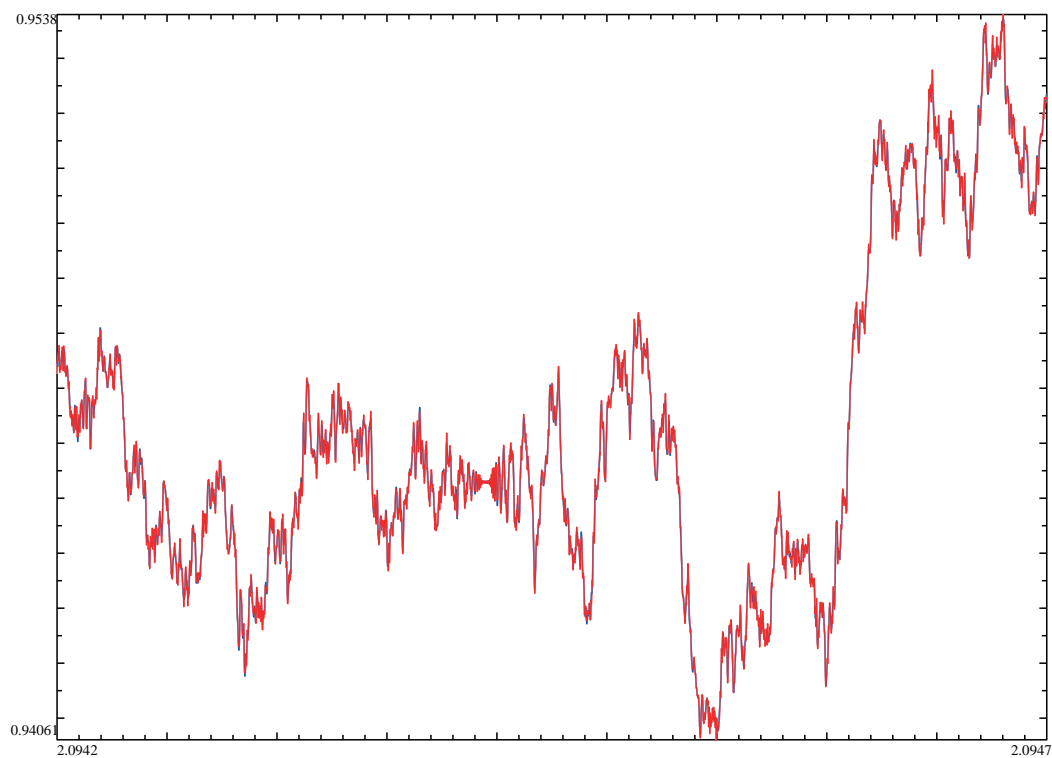


²⁶⁶The smallest square field discriminant $D = 163^2 = 26,569$ for the case with real roots ($x^4 - x^3 - 7x^2 + 2x + 9$) is too large to hope to see patterns in the graphs.

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”:



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 observed fractality *should* take place. The key ingredient is our description of Weil denominators as characteristic polynomials of matrices for Frobenius permutations (see Footnote 152 on p. 60). From another point of view, Frobenius elements permute 3 roots, and their 3×3 permutation matrices can be thought of as permuting *real weights* assigned to the roots. 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 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 Frobenius permutation.²⁶⁷
- 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.²⁶⁸
- Our numbers²⁶⁹ $N_{p^k} =: a_k$ are defined by this recurrence relation. (We start with $a_0 = 1$, and $a_k = 0$ for $k < 0$.)

People who have heard of Artin L -function can immediately recognize²⁷⁰ that our numbers N_m are exactly the coefficients of this function (for our assignment of 2×2 matrices).²⁷¹

Finally, recall that *in the simplest cases* this part of Langlands program is *already known*:

$F(t)$ has required fractal properties when N_m are coefficients of an “uncomplicated” Artin L -function.

According to Langlands–Tunnell results (of ≈ 1980)²⁷² 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.²⁷³

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.²⁷⁴

²⁶⁷Well, only a conjugacy class — but all permutations in a class have the same characteristic polynomial.

²⁶⁸For example, a polynomial $1 - 3u + 2u^2$ gives a recursion relation $a_k - 3a_{k-1} + 2a_{k-2} = 0$.

²⁶⁹... from the section on p. 45.

²⁷⁰In addition to what we did in the section on p. 59, 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.

²⁷¹Since our language is not good enough for a *general* description of what happens in *exceptional* primes, this does verify the match if m is divisible by an exceptional prime. Still, in our particular case one can check this match as well.

²⁷²In 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).

²⁷³The 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*.

²⁷⁴These 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

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 53 on p. 65). 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.^{275 276}

Moreover, in the 2×2 case Gelbart explicitly states²⁷⁷ 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.²⁷⁸ “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 formulations²⁷⁹ 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 formulas²⁸⁰ than $N_p := \widetilde{N}_p - 1$ even for non-special p . Moreover, only a few “flavors” of polynomials allow *such* inclusions into 2×2 -matrices.²⁸¹

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 into²⁸² $SO_3\mathbb{R} \simeq PSU_2 \subset PSL_2\mathbb{C} \subset PGL_2\mathbb{C}$. The same²⁸³ 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 $F(1/\gamma t)/t$, 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. 67). 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 $(d-1) \times (d-1)$

the same multiplicities, 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.)

²⁷⁵For 2×2 -matrices, block-diagonalization and diagonalization are equivalent.

²⁷⁶For example, this happens for abelian polynomials in degree ≥ 3 (as we already saw in Remark 53 on p. 65), and, in degree ≥ 4 , for polynomials with the Galois group being the *dihedral group* D_k .

²⁷⁷In 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!

²⁷⁸In [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 $1/4$ ” (in [Automorphic forms and representations](#), CUP 1998).

²⁷⁹Compare with Footnote 273.

²⁸⁰Moreover, 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 34 on p. 48.) Still, the resulting numbers remain remote cousins of our original coloring of numbers into red/green on p. 17: 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.

²⁸¹One needs to consider *inclusions* since the kernel would lead to us, essentially, studying a polynomial of a smaller degree.

²⁸²Note that this mapping does not lift to a mapping to $GL_2\mathbb{C}$. While there is *another* mapping into $GL_2\mathbb{C}$, it passes through $S_4 \rightarrow S_3$, hence has a kernel. — Therefore the corresponding sequence N_n corresponds to a related polynomial (the “*cubic resolvent*”) of degree 3 (from .

²⁸³... only in this case there is no non-trivial homomorphism into $GL_2\mathbb{C}$ whatsoever.

matrices defined above:

Would the Fourier transform of N_n 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 $\mathrm{GL}_{d-1}\mathbb{C}$. 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.²⁸⁴

Fortunately, the Galois group we started with was finite, so this is the so called “Artin case”,²⁸⁵ 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 \hat{T} with the action of the group $\mathrm{GL}_{d-1}\mathbb{R}$ and a tensor field $\hat{F}(\hat{t})$, $\hat{t} \in \hat{T}$, which is preserved by a certain congruence-subgroup of $\mathrm{GL}_{d-1}\mathbb{Z}$. (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 $\mathrm{GL}_{d-1}\mathbb{R}$ does not act on the projective line! This is why the fractality of $F(t)$ requires a separate consideration. Anyway, a question remains:

Is there a recipe for $\hat{F}(\hat{t})$ in terms of d and the sequence N_n ?

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:**

Is it possible to extract any additional info about functions $F(t)$ from Langlands functoriality?

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. 46) 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”²⁸⁶ 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:²⁸⁷

- This vector space has an increasing filtration indexed by a certain “complexity degree”.²⁸⁸
- 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

²⁸⁴The 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} .)

²⁸⁵Essentially, this means that we consider “motives of dimension 0” — indeed, our polynomial is 1 equation with 1 unknown.

²⁸⁶Unfortunate, my almost complete illiteracy in these topics does not let me say anything more precise.

²⁸⁷I suspect that this approach should be well-known to specialists in the Langlands program — but I never saw it mentioned explicitly.

²⁸⁸This measure of complexity is “orthogonal” to dimension, degree, discriminant and height.

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!).²⁸⁹

- The existence of our “purification 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.²⁹⁰

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.²⁹¹ This suggests²⁹²:

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:²⁹³ given a submanifold $Z \subset X$, we can introduce a relation $[X] = [Z] + [X \setminus Z]$ 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.²⁹⁴

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?

²⁸⁹In 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 $1 + p$).

²⁹⁰Moreover, 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 $\mathrm{GL}_1(\mathbb{C})$) one considers subspaces of c -periodic sequences.

²⁹¹Although 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.

²⁹²I cannot follow it close enough, but I suspect that Cogdell’s paper (see Footnote 278 on p. 109) investigates what happens in this directions.

²⁹³Warning: this should not be confused with the (completely unrelated) K -group of an *individual* manifold!

²⁹⁴Very 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!

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. 14). 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. 14) is way more natural (compare with Footnote 22).

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”.²⁹⁵ 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.²⁹⁶

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.^{298 299}

In fact, the (anti-)palindromicity is a particular case of top-multiplicativity (considered in the following section).

²⁹⁵Any indecomposable quadratic P satisfies this condition. An indecomposable cubic P satisfies it if and only if its discriminant is a perfect square.

²⁹⁶This is a 2-colors variant of Euler’s formulation. Above, on p. 14, we discussed a coloring into 3 colors, when the residues not mutually prime with $|4N|$ were colored gray.²⁹⁷ 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.

²⁹⁷In fact, we described “gray” differently: as “this residue has only a finite number of prime number representatives”. However, a residue mod 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.

²⁹⁸It 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.

²⁹⁹For 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 .³⁰⁰

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 \bullet , \circ and \circ (see p.14; this is compatible with our rules $-\bullet = \circ$ and $-\circ = \bullet$). 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

The third color is “the product” of two other colors.

(Here the “product” is calculated according to the assignments of numbers 0, ± 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 mod p being invertible).

For example, from the list on p.8 we can see that 23 cannot be a divisor of “squares $+ 3$ ”, and the list on 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.³⁰¹

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

³⁰⁰In 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”.

³⁰¹This 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 check³⁰² $p = 3, 5, 7, 17$: these prime numbers cover all the odd residues mod $|4n|$ —and the even residues are going to be gray anyway.

Remark 69: 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 $p \pmod{4}$ or $\pmod{8}$ 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!)³⁰³

Legendre's $p \leftrightarrow q$ -reciprocity

Top-multiplicativity and two cases of Remark 69 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”):

The sign in $\left(\frac{q}{p}\right) = \pm \left(\frac{p}{q}\right)$ is “−” if $p \equiv_4 q \equiv_4 -1$, otherwise it is “+”.

This assumes that p and q are distinct odd positive primes.

Remark 70: 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 $p \pmod{q}$ (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 $\left(\frac{p_1}{q}\right), \dots, \left(\frac{p_r}{q}\right)$ (and now p_1, \dots, 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's 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.³⁰⁴)

If $p \equiv_4 q$, this argument does not even need palindromicity, just periodicity! Indeed, write $q = p + 4n$; then

$$\left(\frac{q}{p}\right) \stackrel{\circ}{=} \left(\frac{q-p}{p}\right) = \left(\frac{4n}{p}\right) \stackrel{*}{=} \left(\frac{n}{p}\right) \stackrel{\circ}{=} \left(\frac{n}{p+4n}\right) = \left(\frac{n}{q}\right) \stackrel{*}{=} \left(\frac{4n}{q}\right) \stackrel{\circ}{=} \left(\frac{4n-q}{q}\right) = \left(\frac{-p}{q}\right) \stackrel{*}{=} \left(\frac{-1}{q}\right) \left(\frac{p}{q}\right).$$

³⁰²In fact, this was already checked in the section on p. 10. 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 $\pm 1 \pmod{3}$ (enough to check $m = 0$ and $m = \pm 1$) which shows that $m^2 - 2$ cannot be divisible by 3.

³⁰³Well, having p^2 in the formula for $\left(\frac{2}{p}\right)$ has an advantage: it makes palindromicity explicit.

³⁰⁴In 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 $\left(\frac{-1}{q}\right) \stackrel{\circ}{=} \left(\frac{-1}{p}\right)$). Here we mark $\stackrel{\circ}{=}$ -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 $\stackrel{\circ}{=}$ is non-obvious!)

Likewise, if $p \equiv_4 -q$, write $p + q = 4n$ (with $n > 0$). Then

$$\left(\frac{q}{p}\right) = \left(\frac{4n-p}{p}\right) \stackrel{\circ}{=} \left(\frac{4n}{p}\right) \stackrel{*}{=} \left(\frac{n}{p}\right) \stackrel{!}{=} \left(\frac{n}{q}\right),$$

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 69, 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. 14!), we may assume that q is odd.

So what we need to show is $\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$ 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 71: 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:

$$\left(\frac{n}{a}\right)\left(\frac{n}{b}\right) = \left(\frac{n}{ab}\right) \quad \text{with } a, b \text{ residues mod } |4n|.$$

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. 113). In fact, Jacobi *defined*³⁰⁵ 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 .³⁰⁶ 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 any³⁰⁷ 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 $m \bmod |4n|$ 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 72: 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 $|n| \leq \frac{1}{2}m$, or change m to make $m \leq 2|n|$, 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 $m \neq 0$).

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.³⁰⁸

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.

³⁰⁵About half a century after Legendre.

³⁰⁶We 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 74.

³⁰⁷There are several convenient ways to extend this to $m \leq 0$, but different contexts benefit from different extensions. So it is reasonable to restrict attention to $m > 0$.

³⁰⁸Indeed, 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 mod 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.³⁰⁹

³⁰⁹I 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. 113, 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 $2 \bmod c$ 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. 14); second, this residue may be colored gray (as we saw for the same $n = -3$ on the 6-wheel on p. 14)!

In particular, the answer depends on our choice of c , the size of the conductor wheel.³¹⁰ 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 out³¹¹ that the answer is³¹³ $C = 2^m n_0$, depending only on the square-free part³¹⁴ n_0 of n , here m depends only on $n_0 \bmod 4$: if $n_0 \equiv_4 1$, then $m = 0$; otherwise $m = 2$. Moreover.³¹⁵

On the C -wheel, there is no exception unless $n_0 \equiv_8 5$, when 2 is the only exception.

(We already saw such an exception happening for $n = -3$ on p. 11.) Hence $2 \bmod C$ 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,³¹⁶ 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.

³¹⁰For example, if $2|c$, then the color of $2 \bmod c$ must be gray, which would lead to $\left(\frac{n}{2}\right) = 0$.

³¹¹Indeed, when we deduced the bottom-periodicity from Legendre formulation (see p. 115), 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_i 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 $\left(\frac{p}{q_1}\right), \dots, \left(\frac{p}{q_k}\right)$ 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).³¹² 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, odd p mutually prime with n cannot be exceptions.) 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 $n_0 \bmod 4$: if n_0 is even, then $m = 3$; if $n_0 \equiv_4 3$, then $m = 2$; otherwise $m = 0$. Hence $2 \bmod C$ 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 $\left(\frac{2}{C}\right)$.

³¹²All these statements follow from the Chinese remainder theorem, and from Dirichlet theorem on arithmetic progressions.

³¹³One can recognize this as the discriminant of the field $\mathbb{Q}[\sqrt{n}]$.

³¹⁴Here we write $n = n_0 K^2$ with the maximal possible K .

³¹⁵Indeed, all prime divisors of C appear as 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 $m = 0$.

³¹⁶See Footnote 297.

Remark 73: 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.³¹⁷

Remark 74: For example, observe two colored rows we matched on p. 7. Now one can recognize the bottom row as colored according to $\left(\frac{-7}{m}\right)$ in the sense of Kronecker.³¹⁸ (The 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 5). Indeed, if P is decomposable, then as we saw on p. 4, already the divisors of one linear factor “would cover” all prime numbers (with just a finite number of exceptions).

If P is indecomposable, then

- Prime divisors p of the numbers in the sequence coincide with p such that $P(x) = 0$ has solutions mod 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 mod 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 $x^2 - D$.

³¹⁷In 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 ”. However, not every number is a possible value of C , since C is a *fundamental discriminant*. Hence if we consider $\left(\frac{n}{m}\right) = \left(\frac{C}{m}\right)$ as a function of C , then it is defined not on every number C , but just on the fundamental discriminants.

³¹⁸This is not exactly true since we were using slightly different notations on p. 7. 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 taught me the relation of the tower of congruence-groups with the adelic approach), the first half³¹⁹ 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,³²⁰ 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’s 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. 21 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, $r_1=3*$, $|D|=1..1000$, $\text{sort}_1=\text{Gal}$, $\text{sort}_2=|D|$, $\text{sort}_3=h$

would result in a list of 27 real cubic fields of small discriminant (first cyclic, then non-cyclic ones).³²¹

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 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(P*Den))*Den*polcoeff(P, poldegree(P)))[, 1].
```

One can check that P is irreducible by `1==factor(P)[, 2]`.

³¹⁹The second half of this review is dedicated to Scholtzefication, which looks unrelated to what we discuss here.

³²⁰See Footnote 277 on p. 109.

³²¹One 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 $M \cdot \text{“Tetrahedral numbers”} + N$ 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. 38.

For a non-exceptional prime p and an irreducible P one can find $\widetilde{N}_p^{\text{res}}$ 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 $\widetilde{N}_{p^k}^{\text{Gal}}$: for the first one, one should replace p by `[ffinit(p,k,varlower("PP")),p]` in the first expression, or replace X^p-X by $X^{p^k}-X$ in the second.

For the following discussion, assume we initialized a few pieces of data with

```
NN = lfunan(lf = lfuninit(lfuncreate(nf = nfinit(P)),[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. 46 it is enough to know the pair $[N_p, N_{p^2}]$.

To see the prime decomposition of p in the field `nf`, inspect³²²

```
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 “cleaning” (which proceeds “as if it removes” a point) would *divide* these local factors by $1 - p$.

Conclusion: in these 5 cases, after cleaning 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 \mathbf{p} and inverting, one gets 5 series in \mathbf{p} , and the coefficients at \mathbf{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.³²³

```
NtypeNonSpec(p,P)=my(F);[# select(x -> x == 1, (F=factormod(P,p,1)[,1])) - 1,0,poldegree(P),#F];
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));my(d=poldegree(nf.pol));my([f,F]=idealfactorBase2(nf,p));\
return([#f-1,sum(k=1,#F,F[k][1][3]-1),d,#F]); \ \ 2: "extra" ramification; 4: total number of factors
coeffNpow(k,t)=if(t[3]==3,coeff3Npow(k,t),t[3]==4,coeff4Npow(k,t),coeff2Npow(k,t));
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));
coeff2Npow(k,t)=t[1]^k;
ppFactor(x)=["base-prime",x[1][1],"ramification",x[1][3],"ff-degree",x[1][4],"multiplicity",x[2]];
```

³²²Take into account “multiplicity” as well???

³²³One can cut-and-paste the code below (*including* the intervening text) into `gp`.

```

specPrimes(P)=my(Den = denominator(content(P))); factor(abs(polcoeff(P,P*Den))*Den*polcoeff(P,poldegree(P)))[,1];
reportSpecFactors(P,nf=nfinit(P))=my(ps=specPrimes(P));\
    for(n=1,#ps,printp(Mat(apply(x -> ppFactor(x), Col(idealfactor(nf,ps[n])))))));
\\ 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,t)));
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));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)= 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 $PN_nINIT(LIM,P)$,³²⁴ one can plot with $plot(X=-0.1,7,[N_n_Ft(X,2),N_n_Ft(X)])$. 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 $my(c);plot(X=-0.1,7,[N_n_Ft(X,2),imag(c=N_n_cFt(X)),real(c)])$ would show the real and the imaginary part.)

If one wants to cover the case of degree 4, one should add this code (only cursorily tested):

```

coeff4_2Npow(k,t)=if(t[2],1,1+k\2);
coeff4_1Npow(k,t)=if(t[2]>1,0,t[2],!(k%2),!(k%3));
coeff4_0Npow(k,t)=if(t[2],1-2*(k%2),t[4]>1,(1+k\2)*(1-2*(k%2)),[1,-1,0,0][1+k%4]);
coeff4Npow(k,t)=if(t[1]==3,(k+1)*(k+2)/2,t[1]==2,k+1,t[1]==1,coeff4_2Npow(k,t),t[1]==-1,coeff4_0Npow(k,t),coeff4_1Npow(k,t));

```

To compare our manually-computed array N_n with $|PARI|$'s $lfunan()$, 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=#LL)=forprime(p=2,L,ckPr(p,L));
ckP(P,LIM=1000000)=N_n_INIT(LIM,P);NN = lfunan(lf = lfunit(lfuncreate(nf = nfinit(P)),[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. 86) as $lfun(13,-s)$. In the case of modular forms, if one knows N_p for a few values of p , one can use $mfeigensearch([1..LIMc],1,[p_1,N_{p_1}],\dots,[p_k,N_{p_k}])$ to list all cases for sequences N_m with these particular values, up to $c = LIMc$.

³²⁴This usage assumes that P is indecomposable. Otherwise one needs to specify the parameter LST explicitly. For example, for the case $M = 16$ considered on 60,

$$LST = [[2, [0, 0, 3]], [3, [2, 0, 3]], [13, [1, 0, 3]]]$$

One can find the primes to include by $specPrimes(P)$. To find the suitable arrays, follow the explanations above and the section on p. 59.