Counting Arithmetic Objects

Contents

CHAPTER 1

Analytic number theory and zeta function methods

Frank Thorne

Let $N_3^{\pm}(X)$ count the number of cubic fields K with $0 < \pm\mathrm{Disc}(K) < X$. Using Shintani zeta functions, we have proved the following theorem:

Theorem 0.1. *We have*

$$N_3^{\pm}(X) = C^{\pm}\frac{1}{12\zeta(3)}X + K^{\pm}\frac{4\zeta(1/3)}{5\Gamma(2/3)^3\zeta(5/3)}X^{5/6} + O(X^{2/3+\epsilon}). \tag{1}$$

'We' is Bhargava, Taniguchi, and the speaker; this builds mainly on previous work with Taniguchi, getting an error term of $O(X^{7/9+\epsilon})$, and work of Bhargava, Shankar, and Tsimerman [4] using the geometry of numbers method. The object of these notes is to recall the definition of the Shintani zeta function, to introduce some of the tools that go into the proof of the above theorem, and to compare and contrast the zeta function method with the geometry of numbers method explained by Bhargava.

The zeta function method starts in the same place as the geometry of numbers method: with the parameterization of cubic rings by cubic forms.

1. Definitions and parameterizations

1.1. Cubic rings and cubic forms. A *cubic ring* is a commutative ring which is free of rank 3 as a $\mathbb{Z}$-module. The discriminant of a cubic ring is defined to be the determinant of the trace form $\langle x, y\rangle = \mathrm{Tr}(xy)$, and the discriminant of the maximal order of a cubic field is equal to the discriminant of the field.

The lattice of *integral binary cubic forms* is defined by

$$V_{\mathbb{Z}} := \{au^3 + bu^2v + cuv^2 + dv^3 \ : a, b, c, d \in \mathbb{Z}\}, \tag{2}$$

and the *discriminant* of such a form is given by the usual equation

$$\mathrm{Disc}(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd. \tag{3}$$

There is a natural action of $\mathrm{GL}_2(\mathbb{Z})$ (and also of $\mathrm{SL}_2(\mathbb{Z})$) on $V_{\mathbb{Z}}$, given by

$$(\gamma \cdot f)(u, v) = \frac{1}{\det\gamma} f((u, v)\cdot\gamma). \tag{4}$$

We call a cubic form f *irreducible* if $f(u, v)$ is irreducible as a polynomial over $\mathbb{Q}$, and *nondegenerate* if $\mathrm{Disc}(f) \neq 0$.

Cubic rings are related to cubic forms by the following correspondence of Delone and Faddeev, as further extended by Gan, Gross, and Savin [11]:

Theorem 1.1 **([10, 11]).** *There is a natural, discriminant-preserving bijection between the set of* $\mathrm{GL}_2(\mathbb{Z})$*-equivalence classes of integral binary cubic forms and the set of isomorphism classes of cubic rings. Furthermore, under this correspondence, irreducible cubic forms correspond to orders in cubic fields.*

Finally, if $x \in V_{\mathbb{Z}}$ *is a cubic form corresponding to a cubic ring* R*, we have* $\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(x) \simeq \mathrm{Aut}(R)$.

1.2. Shintani's zeta function. The *Shintani zeta functions* associated to the space of binary cubic forms are defined[1] by the Dirichlet series

$$\xi^{\pm}(s) := \sum_{x \in \mathrm{GL}_2(\mathbb{Z})\backslash V_{\mathbb{Z}}^{\pm}} \frac{1}{|\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(x)|}|\mathrm{Disc}(x)|^{-s} = \sum_{\pm\mathrm{Disc}(R)>0} \frac{1}{|\mathrm{Aut}(R)|}|\mathrm{Disc}(R)|^{-s}. \tag{5}$$

In the former sum, $V_{\mathbb{Z}}$ is the lattice defined in (2), the sum is over elements of positive or negative discriminant respectively, and $\mathrm{Stab}(x)$ is the stabilizer of x in $\mathrm{GL}_2(\mathbb{Z})$. The latter sum is over isomorphism classes of cubic rings.

Shintani proved [23] that all of these Dirichlet series converge absolutely for $\Re(s) > 1$, enjoy analytic continuation to all of $\mathbb{C}$ with poles only at $s = 1$ and $s = 5/6$, and satisfy the matrix functional equation

[1] The Shintani zeta functions are commonly defined in terms of $\mathrm{SL}_2(\mathbb{Z})$ rather than $\mathrm{GL}_2(\mathbb{Z})$, multiplying (5) by a factor of 2.

$$(6)\quad \begin{pmatrix} \xi^+(1-s) \\ \xi^-(1-s) \end{pmatrix} = \Gamma\Big(s - \frac{1}{6}\Big)\Gamma(s)^2\Gamma\Big(s + \frac{1}{6}\Big)2^{-1}3^{6s-2}\pi^{-4s} \times \begin{pmatrix} \sin 2\pi s & \sin \pi s \\ 3\sin \pi s & \sin 2\pi s \end{pmatrix} \begin{pmatrix} \widehat{\xi}^+(s) \\ \widehat{\xi}^-(s) \end{pmatrix}.$$

He also explicitly computed all of the residues.

In the above, the *dual zeta functions* $\widehat{\xi}^{\pm}(s)$ are defined identically to $\widehat{\xi}(s)$, except that in place of (2) one sums over the dual lattice $\widehat{V_{\mathbb{Z}}}$, where one insists that the middle two coefficients be divisible by 3.

Shintani didn't address the question of counting cubic fields; the first people to use zeta functions for this were Datskovsky and Wright [7]. Shintani, did, however prove an asymptotic density formula for the number of $x \in \mathrm{GL}_2(\mathbb{Z})\backslash V_{\mathbb{Z}}^{\pm}$ with $0 < \pm\mathrm{Disc}(x) < X$, and he obtained an error term of $O(X^{3/5+\epsilon})$. This is in contrast with the geometry of numbers method, where a natural barrier is $O(X^{3/4})$ (and, indeed, BST achieved precisely this). It is not out of the question that one could do better, but one would need something stronger than Davenport's lemma: given any 4-dimensional blob whose volume is X, it will certainly have some lower dimensional projection whose volume is at least $X^{3/4}$.

A major difference between Shintani's method and the geometry of numbers method is that Shintani's zeta function equally counts both the reducible points and the irreducible points. In Bhargava's work, a beautiful simplification was made possible by the fact that the set of reducible points essentially coincides with the cusp. In other words, the points that one doesn't want to count for algebraic reasons like in the the part of the fundamental domain that one doesn't want to count for analytic reasons. So, the algebraist and the analyst agree to cut off the cusp from the very beginning, and everyone is happy.

This isn't perfect: although every point with $a = 0$ is reducible; it is not the case that every point in the main body is irreducible. Bhargava proved that there are $\ll X^{3/4}$ reducible points in the main body of the fundamental region — certainly good enough for sharp quantitative results, but it does highlight that geometry of numbers methods are, in some sense, somewhat inexact.

In contrast, let $\phi : [0, \infty)$ be a suitably smooth test function ($\phi(t) = e^{-t}$ is the canonical example); then, Shintani's theory, when combined with standard techniques in analytic number theory, yields an *exact* formula for $\sum_R |\mathrm{Aut}(R)|^{-1}\phi(|\mathrm{Disc}(R)|/X)$. The disadvantage, however, is that it is quite difficult to entangle the reducible from the irreducible rings. This was accomplished (again by Shintani, with later work due to Taniguchi), but only after a severe effort, and not in a way that seems to generalize nicely to more complicated prehomogeneous vector spaces.

The remainder of these notes will illustrate a few of the tools which into leveraging Shintani's theory (with extensions due to Datskovsky and Wright, F. Sato, Taniguchi and the author, and others) and obtaining the results described above.

2. Are zeta functions completely different from Bhargava's method?

Should I leave this in? It is cool, but I think it might be the subject of a paper in progress by Bhargava, Taniguchi, and/or others.

We recall from Taniguchi's lecture that the **global zeta function** is

$$Z(f, s) := \int_{\mathrm{GL}_2(\mathbb{R})/\mathrm{GL}_2(\mathbb{Z})} (\det\ g)^{2s} \sum_{x \in V_{\mathbb{Z}} - S} f(gx)dg.$$

Here f is a test function satisfying suitable properties, and S is the zero locus. The 2 is present because we have

$$\mathrm{Disc}(gx) = (\det g)^2\mathrm{Disc}(x).$$

The key relation is the following:

Proposition 2.1. *We have*

$$Z(f, s) = \frac{1}{4\pi}\xi^+(L, s)\int_{V+} |\mathrm{Disc}(x)|^{s-1}f(x)dx + \frac{1}{12\pi}\xi^-(L, s)\int_{V-} |\mathrm{Disc}(x)|^{s-1}f(x)dx.$$

The proof is largely formal: interchange the summation and integration, and do a Jacobian change of variables. We obtain that the global zeta function is equal to a sum of two terms, each of which is a Dirichlet series ('the' Shintani zeta function) times a 'local zeta function'.

Shintani then proves the functional equation. The key idea is to complete the sum to $x \in V_{\mathbb{Z}}$ and then use the Poisson summation formula. It transpires that the sum over $x \in S$ which must be added contains all the subtlety, and in particular this is the sum which gives the zeta function its poles.

We compare to Bhargava's averaging method. Bhargava took a ball $B \subseteq V^{\pm}$ and averaged it, to obtain, e.g.,

$$N^+(X) = C \cdot \frac{\int_{g \in \mathcal{F}} \{\#x \in gB \cap V(\mathbb{Z})^{\mathrm{irr}}\ : |\mathrm{Disc}(x)| < X\}dg}{\int_{v \in B} |\mathrm{Disc}(v)|^{-1}dv.}$$

If one replaces the sharp count $|\mathrm{Disc}(x)| < X$ with a sum over all discriminants, weighted by $|\mathrm{Disc}(x)|^{-s}$ and also their automorphism groups, and the characteristic function of the ball B with $f(x)$, and loses the restriction to irreducible points, one recovers Proposition 2.1.

3. A fly and a sledgehammer

In this section we will wield a sledgehammer against the following fly:

Question 3.1. *How many positive integers are there less than X?*

The asymptotic formula $X + o(X)$ can be proved by either the geometry of numbers or zeta functions. The geometry of numbers method, via Davenport's lemma, establishes that the number of positive integers is equal to the length of the interval $[0, X]$, plus an error term $O(\max(V, 1))$, where V is the maximum volume of any projection of the interval $[0, X] \subset \mathbb{R}^1$ onto any subset of the coordinate axes. In other words, it is equivalent to the observation that the problem is trivial.

The zeta function method is more entertaining: it allows us to introduce a finite differencing method due to Landau; it illustrates the principle that smoothing is useful in analytic number theory, and demonstrates the genesis of this principle in Fourier analysis; it showcases the importance of upper bounds for the zeta function; it admits an improvement via a nontrivial study of Bessel functions. In short, it is everything an analytic number theorist could look for.

Jokes aside, it allows us to introduce all of our analytic tools in a setting where the problem is well understood; these will be the same tools that allow us to obtain the best error terms in the Davenport-Heilbronn theorems. Let us begin.

3.1. The Riemann zeta function. The *Riemann zeta function* is given by the infinite series

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s}, \tag{7}$$

absolutely convergent for $\Re(s) > 1$. We could write

$$\zeta(s) := \sum_{n=1}^{\infty} a(n) n^{-s}, \tag{8}$$

with $a(n) := 1$ for all n; we will not do this, but this would emphasize that the Riemann zeta function is indeed a generating series for a sequence of arithmetic interest.

We will presume that the reader is familiar (see [8], among very many other sources) with its various analytic properties, in particiular the *functional equation*

$$\zeta(s)\pi^{-s/2}\Gamma(s/2) = \zeta(1-s)\pi^{-\frac{1-s}{2}}\Gamma\Big(\frac{1-s}{2}\Big). \tag{9}$$

What will be important for us is that it has an analytic continuation to the whole complex plane, for which

$$\zeta(-\sigma + it) \ll (1 + |t|)^{1/2+\sigma} \tag{10}$$

for each fixed $\sigma > 0$, as $|t| \to \infty$. It has a simple pole at $s = 1$ with residue 1, and we will also use that $\zeta(0) = -\frac{1}{2}$.

Exercise 1. *Using the functional equation* (9) *and Stirling's formula, prove* (10)*.*

There is not much content to this exercise, but working it out will make clear the relationship between functional equations for zeta functions and their asymptotic growth in $\Re(s) < 0$.

Now, we have by *Perron's formula* that

$$\sum_{n<X} 1 = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s)\frac{X^s}{s}ds. \tag{11}$$

More on this later. We would *like* to evaluate this integral by shifting it to some vertical line $-\sigma + it$ with $\sigma > 0$, obtaining

$$\sum_{n<X} 1 = \mathrm{Res}_{s=1}\zeta(s)\cdot\frac{X^1}{1} + \frac{1}{2\pi i}\int_{-\sigma-i\infty}^{-\sigma+i\infty} \zeta(s)\frac{X^s}{s}ds. \tag{12}$$

The first term is equal to X, as expected. However, the shifted contour *does not converge.* Therefore, we must work harder to make this method work.

3.2. Perron's formula and Mellin transforms. A more general form of Perron's formula says that if

$$\xi(s) = \sum_{n=1}^{\infty} a(n) n^{-s}, \tag{13}$$

then

$$\sum_{n=1}^{X} a(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \xi(s)\frac{X^s}{s}ds, \tag{14}$$

presuming that X is not an integer, and that the Dirichlet series converges absolutely for $\Re(s) = c$.

Exercise 2. *Prove first that*

$$\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} Y^s \tag{15}$$

is equal to 1 *if* $Y > 1$ *and* 0 *if* $0 < Y < 1$*. Then, use this fact to prove Perron's formula.*

If $Y = 1$, the integral equals $\frac{1}{2}$, but we can get away with not caring about this.

We can vary this as follows, and obtain a *smoothed* version. We have

$$\sum_{n=1}^{X} a(n)(X-n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \xi(s)\frac{X^{s+1}}{s(s+1)}ds. \tag{16}$$

Exercise 3. *Prove this.*

Exercise 4. *Prove that*

$$\frac{1}{2}\sum_{n=1}^{X} a(n)(X-n)^2 = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \xi(s)\frac{X^{s+2}}{s(s+1)(s+2)}ds. \tag{17}$$

Exercise 5. *For each positive integer c, formulate and prove a generalization with* $(X-n)^c$ *on the left and an appropriate expression on the right.*

All of these are special cases of a more general theorem, that

$$\sum_{n=1}^{\infty} a(n)\phi(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \xi(s)\widehat{\phi}(s)ds, \tag{18}$$

where the *Mellin transform* $\widehat{\phi}(s)$ is

$$\widehat{\phi}(s) := \int_0^\infty \phi(x)x^{s-1}dx. \tag{19}$$

See Iwaniec and Kowalski (or the Wikipedia article on the Mellin inversion theorem) for conditions on which this theorem is valid. This is just a change of variables from the Laplace or Fourier transform. In particular, we can extrapolate from Fourier analysis that *if* ϕ *is smoother,* $\widehat{\phi}(s)$ *decays faster as* $|\Im(s)| \to \infty$.

3.3. Landau's method. We can now obtain an estimate for $\sum_{n=1}^{X}(X-n)$. (Of course we can do so by trivial means; the point of all this is that the same method will work for $\sum_{n=1}^{X} a(n)(X-n)$.)

We have, by (16),

$$\sum_{n=1}^{X}(X-n) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s)\frac{X^{1+s}}{s(s+1)}ds. \tag{20}$$

Shifting the contour to $\Re(s) = -\sigma$ for some small $\sigma > 0$, we obtain poles from the integrand at $\Re(s) = 1$ and $\Re(s) = 0$, and thereby obtain

$$\sum_{n=1}^{X}(X-n) = \frac{X^2}{2} - X + \frac{1}{2\pi i}\int_{-\sigma-i\infty}^{\sigma+i\infty} \zeta(s)\frac{X^{1+s}}{s(s+1)}ds. \tag{21}$$

Exercise 6. *Rigorously justify the previous step, in the process determining* $\sigma_0 > 0$ *such that the integral above is absolutely convergent for* $\sigma < \sigma_0$.

The previous exercise shows that we may take $\sigma = 1/2 - \epsilon$, and one may verify that the integral is then $\ll_\epsilon X^{1/2+\epsilon}$, so that

$$\sum_{n=1}^{X}(X-n) = \frac{X^2}{2} - X + O(X^{1/2+\epsilon}). \tag{22}$$

Now, following Landau, we introduce a combinatorial trick to estimate $\sum_{n=1}^{X}$. For any $Y > 0$, we also have that

$$\sum_{n=1}^{X+Y}(X+Y-n) = \frac{(X+Y)^2}{2} - (X+Y) + O((X+Y)^{1/2+\epsilon}). \tag{23}$$

Exercise 7. *Assuming that* $Y < X$*, subtract the first equation from the second, divide through by* Y*, and choose* $Y = X^{1/4}$ *to obtain that*

$$\sum_{n=1}^{X} 1 = X + O(X^{1/4+\epsilon}). \tag{24}$$

Remark. *In estimating* (21) *we made no attempt to exploit the oscillation coming from the term* X^s*. See the work of Chandrasekharan and Narasimhan* [5] *for more on how to exploit this oscillation and thereby obtain better error terms.*

3.4. Finite differencing using combinatorics. We would like to apply this finite differencing procedure more generally. However, in place of (10) we will obtain bounds of the shape

$$\xi(-\sigma+it) \ll (1+|t|)^{d\cdot(1/2+\sigma)}, \tag{25}$$

where d is the analytic degree (essentially, the number of gamma factors) of the zeta function $\xi(s)$. Therefore, (16) will not in general be smooth enough, and we may need (17) or its generalization. We must therefore also generalize our finite differencing procedure, and this is done as follows.

Exercise 8. *Prove, for each positive integer j, the combinatorial identity*

$$\frac{1}{j!}\sum_{i=0}^{j}\binom{j}{i}(-1)^{j-i}(X+iY-n)^j = Y^n. \tag{26}$$

Now let $a(n)$ be any arithmetic sequence, for which we write $S_j(X) = \sum_{n=1}^{X} a(n)(X-n)^j$.

Exercise 9. *Using the above, prove that*

$$\frac{1}{j!}\sum_{i=0}^{j}\binom{j}{i}(-1)^{j-i}S_j(X) = Y^j\cdot\sum_{n=1}^{X} a(n) + E, \tag{27}$$

where the error term E satisfies

$$E = O_j\left(Y^j\sum_{n=X}^{X+Y} a(n)\right), \tag{28}$$

and if the $a(n)$ are nonnegative,

$$0 \le E \le Y^j\sum_{n=X}^{X+Y} a(n). \tag{29}$$

We arrive now at the payoff.

Exercise 10. *Consider the k-divisor function $d_k(n)$, for which we have $d_k(n) \ll_{k,\epsilon} n^{\epsilon}$ and*

$$\sum_{n=1}^{\infty} d_{k-1}(n)n^{-s} = \zeta(s)^k. \tag{30}$$

Generalize the technique described above to obtain asymptotic estimates for $\sum_{n=1}^{X} d_{k-1}(n)$, with lower order terms and power saving error terms.

This will be rather involved, but quite rewarding to work out. You will use (17) *and its further generalization.*

The method works for any zeta function $\xi(s)$ with meromorphic continuation to $\mathbb{C}$, provided that:

- We have an estimate of the shape

$$\xi(-\sigma+it) \ll_{\sigma} (1+|t|)^{d\cdot(1/2+\sigma)}, \tag{31}$$

  as is known for every zeta function in the 'extended Selberg class'.
- We can somehow bound the expression $\sum_{n=X}^{X+Y} a(n)$.

In particular, the method can be used to obtain asymptotic formulas, with an error term $\ll X^{3/5+\epsilon}$, for the number of cubic *rings*. (Shintani's functional equation (6) gives $d=4$ in (31)). To count *fields*, a sieve is required.

4. A simple sieve

Here we introduce the inclusion-exclusion sieve and illustrate when it can be used to obtain power-saving error terms.

We will count the number $N^{\flat}(X)$ of *squarefree integers less than X*. To parallel our proof for cubic fields we define some rather obtuse terminology and notation. We will define an integer to be *squarefree at q* if it is not divisible by q^2, so that an integer is squarefree if and only if it is squarefree at every prime.

Moreover, we write $N(q,X)$ (in this section only) for the number of positive integers $< X$ and divisible by q^2, for which we trivially[2] have $N(q,X) = \frac{X}{q^2} + O(1)$.

By inclusion-exclusion, we have

[2]If you'd like to have more fun than this, note that the Dirichlet series associated to integers divisible by q^2 is $q^{-2s}\cdot\zeta(s)$. Now, apply the Landau method to get an estimate for $N(q,X)$. How does your error term depend on q?

$$
\begin{aligned}
N^\flat(X) &= \sum_{q\geq 1}\mu(q)N(q,X) = \sum_{q\leq\sqrt{X}}\mu(q)N(q,X)\\
&= \sum_{q\leq\sqrt{X}}\mu(q)\Big(X/q^2+O(1)\Big) = X\cdot\sum_{q\leq\sqrt{X}}\mu(q)q^{-2}+O(\sqrt{X})\\
&= X\sum_{q=1}^{\infty}\mu(q)q^{-2}+O(\sqrt{X})\\
&= X\prod_p\Big(1-\frac{1}{p^2}\Big)+O(\sqrt{X})\\
&= X\zeta(2)^{-1}+O(\sqrt{X}) = \frac{6}{\pi^2}X+O(\sqrt{X}).
\end{aligned}
$$

It is somewhat atypical that we can obtain such good error terms in analytic number theory with so little effort. For example, an error term of $O(X^{1/2+\epsilon})$ for the prime counting function is equivalent to the Riemann Hypothesis; worth a million dollars and (if you are under forty) a probable Fields Medal as well. It is instructive to attempt to obtain such rewards in five lines:

Exercise 11. *Adapt the above sieve to sieving the integers for indivisibility, and attempt to obtain an asymptotic formula for the prime counting function $\pi(x)$ with a power saving error term.*

Fail. Then, discuss what feature of the previous problem allowed for an easy proof of such good error terms.

Interestingly, the above line of reasoning allows us to determine the asymptotic density of *quadratic* fields. This can be seen as follows.

The set of *fundamental discriminants* is equal to $\{1\}$ plus the set of discriminants of quadratic fields. It is a classical exercise in algebraic number theory (exercise: work out the details) that this set is equal to the set of squarefree integers $\equiv 1 \pmod 4$, and 4 times the set of squarefree integers $\equiv 3 \pmod 4$.

Exercise 12. *Prove that*

$$
\sum_{D>0} D^{-s} = \frac{1}{2}\bigg[\big(1-2^{-s}+2\cdot 4^{-s}\big)\frac{\zeta(s)}{\zeta(2s)}+\big(1-4^{-s}\big)\frac{L(s,\chi_4)}{L(2s,\chi_4)}\bigg], \tag{32}
$$

where the Dirichlet series extends over all fundamental discriminants. Then prove a similar formula for negative fundamental discriminants.

Remark. *If one combines positive and negative fundamental discriminants, one obtains the beautiful equation*

$$
\sum |D|^{-s} = \Big(1-2^{-s}+2\cdot 4^{-s}\Big)\frac{\zeta(s)}{\zeta(2s)} = \prod_p\bigg(\frac{1}{2}\sum_{[K_v:\mathbb{Q}_p]\leq 2}|\mathrm{Disc}(K_v)|_p^s\bigg). \tag{33}
$$

Work of Wright [30] *proves this using prehomogeneous vector spaces, in a way that very nicely generalizes.*

Exercise 13. *Prove that the number of positive or negative fundamental discriminants $0<\pm D<X$ is $\frac{3}{\pi^2}X+O(X^{1/2})$, for either choice of sign. This gives an asymptotic formula for the counting function of quadratic fields.*

This problem is still easy, but involves a little bit of kludge, of the flavor that occurs quite frequently when adapting general analytic methods to specific situations.

It is interesting to note that one can also write the Dirichlet series of squarefree integers as $\frac{\zeta(s)}{\zeta(2s)}$ and estimate $N^\flat(X)$ directly by Landau's method (or by contour integration, anyway). In this case it is easiest to move the contour only to $\Re(s)=\sigma$ for some fixed $\sigma\geq\frac{1}{2}$, to avoid complications coming from the zeroes of $\zeta(2s)$. One can then use the *convexity bound*

$$
\zeta(\sigma+it) \ll (1+|t|)^{\frac{1-\sigma}{2}} \tag{34}
$$

to estimate the integrand in the critical strip. One also has *subconvexity bounds* better than the above; see Iwaniec and Kowalski's book [12] for an introduction to this fascinating topic.

5. Shintani's zeta function: local specifications and their Fourier transforms

Recall that we are interested in counting cubic fields, or equivalently their maximal orders. Maximal orders are cubic rings, and recall that there are three properties which distinguish maximal orders $\mathcal{O}$ from general cubic rings $\mathbb{R}$.

- Maximality: $\mathcal{O}$ is not properly contained in any other cubic ring.
- Nondegeneracy: $\mathrm{Disc}(\mathcal{O})\neq 0$.
- Irreducibility: $\mathcal{O}$ is an integral domain; equivalently, the corresponding cubic form is irreducible over $\mathbb{Z}$.

The *nondegeneracy* condition is, in some sense, automatic: the Shintani zeta function only counts nondegenerate cubic rings, and so we never see the degenerate rings when we count. (Note, however, that the residue formulas are described in terms of the degenerate rings, so that in some deep sense we very much have to understand the degenerate rings to make the zeta function method work.)

The *maximality* condition is, in some sense, simultaneously the most involved and the most straightforward to deal with. We will spend most of this section describing how to detect this condition.

The *irreducibility* condition is the wild card, and it is the one condition that is treated very differently in the geometry of numbers and zeta functions methods. In the GON method, for a certain well-chosen fundamental domain one 'cuts off' the cusp, which can easily be seen to contain only reducible points: so, conveniently, the points which are most difficult to count are precisely the points which we don't *want* to count.

In the zeta function method, we do not *yet* have any good *a priori* way to separate the reducible from the irreducible points. We will therefore end up counting all field extensions of degree ≤ 3 (indeed: all *etale algebras* of degree ≤ 3, which in the this case are in bijection with field extensions, but which in higher degrees include nontrivial direct sums of fields) and subtracting everything of lower degree.

In the cubic case, we understand quite well what to subtract: there is only one number field of degree 1, and you counted the quadratic fields in Exercise 13.

5.1. Nonmaximality and the sieve. The key to counting maximality is that it satisfies an appropriate local-to-global principle. Namely, a cubic ring R is maximal if and only if it is maximal at each prime p, where the latter condition is defined by the following equivalent conditions:

Proposition 5.1 ([9, 4])**.** *A cubic ring R is maximal at p if and only if the following equivalent conditions hold:*

- *R is not contained in any other cubic ring with index divisible by p.*
- *R is not contained in any other cubic ring with p-power index.*
- *$R \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is maximal as a cubic ring over $\mathbb{Z}_p$.*
- *(The Davenport-Heilbronn correspondence) The cubic form f corresponding to R is not a multiple of p, and there is no $\mathrm{GL}_2(\mathbb{Z})$-transformation of $f(u,v) = au^3 + bu^2v + cuv^2 + dv^3$ such that a is a multiple of p^2 and b is a multiple of p.*

Exercise 14. *Prove the local-to-global principle and the equivalence of the above conditions. Rather than working from scratch, you might refer to the briefly sketched proofs in* [4] *and give an elaboration.*

Define $N^{\pm}(q, X)$ to be the number of cubic rings R of discriminant $0 < \pm D < X$ which are nonmaximal at every prime dividing q. We recall the inclusion-exclusion sieve introduced previously to count squarefree integers. In this case, the number of maximal cubic rings is given by the simple formula

$$N^{\pm}_{\max}(X) = \sum_{q \geq 1} \mu(q) N^{\pm}(q, X). \tag{35}$$

This extremely simple formula was overlooked by Davenport and Heilbronn, Datskovsky and Wright, Belabas (in his initial work on error estimates in counting cubic fields), and others. Exercise 11 illustrates why this was so easy to overlook. Indeed, this exercise illustrates the extreme usefulness of the bound

$$N^{\pm}(q, X) \ll X 3^{\omega(q)}/q^2, \tag{36}$$

uniformly in q, which is proved in [1] by reasonably elementary methods. (It amounts to bounding how many cubic rings R' can be contained in a fixed cubic ring R with fixed index.) Such a bound is very atypical in applications of sieve methods.

Therefore, one can *bound the tail* and obtain the formula

$$N^{\pm}_{\max}(X) = \sum_{q \leq Q} \mu(q) N^{\pm}(q, X) + O(X/Q^{1-\epsilon}). \tag{37}$$

Therefore, if we can obtain estimates for $N^{\pm}(q, X)$, with power saving estimates and with the q-dependence controlled, we can prove a power saving error estimate for $N^{\pm}_{\max}(X)$. This is the same for both the GON and the zeta function methods, with one exception: the zeta function method carries along the reducible rings, where in the GON method they were excluded from the start.

Remark. *Counting squarefree integers, we remarked that we could divide our Dirichlet series through by $\zeta(2s)$ instead of introducing a sieve. One may check that no similar such technique applies here.*

5.2. Estimation of $N^{\pm}(q, X)$. Define the *q-nonmaximal Shintani zeta functions* by

$$\xi^{\pm}_q(s) := \xi^{\pm}(\Phi_q, s) := \sum_{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash V^{\pm}_{\mathbb{Z}}} \frac{1}{|\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(x)|} \Phi_q(x) |\mathrm{Disc}(x)|^{-s}, \tag{38}$$

where $\Phi_q(x)$ is the characteristic function of 'nonmaximal at q'. By the Davenport-Heilbronn correspondence (Proposition 5.1), this function depends only on $x \pmod{q^2}$, and is $\mathrm{GL}_2(\mathbb{Z}/q^2\mathbb{Z})$-invariant (required for the above sum is well defined).

It follows by work of Datskovsky-Wright and F. Sato that these zeta functions also have analytic continuation and a functional equation. The functional equation is identical if one takes the dual zeta functions to be defined by

$$\widehat{\xi}_q^{\pm}(s) := (q^2)^{4s-4} \sum_{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash \widehat{V}_{\mathbb{Z}}} \frac{1}{|\mathrm{Stab}(x)|} \widehat{\Phi}_q(x) |\mathrm{Disc}(x)|^{-s} \tag{39}$$

where

$$\widehat{\Phi}_q(x) := \sum_{y \in V_{\mathbb{Z}/q^2\mathbb{Z}}} \Phi_q(y) \exp(2\pi i [x,y]/q^2), \tag{40}$$

for the alternating bilinear form

$$[x,y] = x_4 y_1 - \frac{1}{3} x_3 y_2 + \frac{1}{3} x_2 y_3 - x_1 y_4 \tag{41}$$

used to identify V with $\widehat{V}$, where x_i and y_j are the coordinates of x and y respectively.

Remark. *The above definitions can also be normalized differently. For example, it is arguably more natural to incorporate the q^{8-8s} appearing in (39) into the functional equation, so that the dual zeta function is of the form $\sum_{m \in \mathbb{Z}^+} \widehat{a_q}(m) m^{-s}$, where in particular the m are all integers.*

In any case, the choice of normalization does not make any substantive difference to the subsequent analysis.

Indeed, all of the above is true in greater generality. If $\Psi : V_{\mathbb{Z}/mZ} \to \mathbb{C}$ is any $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$-invariant function, then the above equations all hold with Φ_q replaced with Ψ, and q^2 replaced with m. Moreover, the Fourier transforms above enjoy the following multiplicativity property:

Exercise 15. *Prove that $\widehat{\Phi}_q$ satisfies the multiplicativity property $\widehat{\Phi}_q(x)\widehat{\Phi}_{q'}(x) = \widehat{\Phi}_{qq'}(x)$for all $(q,q') = 1$, and formulate and prove a generalization of this property to a wide class of functions Ψ (which you should explicitly describe).*

5.3. Analytic bounds for modified zeta functions.

Exercise 16. *Taking (44) for granted, prove that for any $\Psi \pmod m$ as above we have*

$$\xi^{\pm}(\Psi, -\sigma + it) \ll_\sigma (1 + |t|)^{4 \cdot (1/2 + \sigma)} m^{4 + 4\sigma}. \tag{42}$$

You will use the trivial bound

$$|\widehat{\Phi}_m(x)| \leq m^4. \tag{43}$$

Prove in addition that if the bound m^4 is replaced with some $C < m^4$ for which (43) holds uniformly in x, we can replace the bound in (44) with

$$\xi^{\pm}(\Psi, -\sigma + it) \ll_\sigma (1 + |t|)^{4 \cdot (1/2 + \sigma)} C m^{4\sigma}. \tag{44}$$

We also need the residue formulas for the q-nonmaximal zeta functions. The zeta functions are holomorphic away from $q = 1$ and $q = 5/6$, and at these points we have

$$\mathrm{Res}_{s=1} \xi_q^{\pm}(s) = \alpha^{\pm} \prod_{p|q} \left(\frac{1}{p^2} + \frac{1}{p^3} - \frac{1}{p^5} \right) + \beta \prod_{p|q} \left(\frac{2}{p^2} - \frac{1}{p^4} \right), \tag{45}$$

and

$$\mathrm{Res}_{s=5/6} \xi_q^{\pm}(s) = \gamma^{\pm} \zeta(1/3) \prod_{p|q} \left(\frac{1}{p^{5/3}} + \frac{1}{p^2} - \frac{1}{p^{11/3}} \right), \tag{46}$$

where $\alpha^+ = \pi^2/36$, $\alpha^- = \pi^2/12$, $\beta = \pi^2/12$, $\gamma^+ = \frac{\Gamma(1/3)^3}{4\sqrt{3}\pi} = \frac{2\pi^2}{9\Gamma(2/3)^3}$, and $\gamma^- = \sqrt{3}\gamma^+$.

At this point everything is in place for a proof of Davenport-Heilbronn:

Exercise 17. *Using the trivial bound and Landau's method, prove an asymptotic formula with a power saving estimate for $N_{\max}^{\pm}(X)$ and therefore the counting function of cubic fields $N_3^{\pm}(X)$.*

Currently, the best error term for counting $N_3^{\pm}(X)$ is $O(X^{2/3+\epsilon})$. What has been presented above is nowhere near good enough to get this error term. One requires the following in addition:

- A refinement of the Landau method, as given in a paper of Chandrasekharan and Narasimhan [5], whereby one treats the error term integrals in a highly nontrivial manner (using Bessel functions). In general, for a zeta function of degree d, this yields error terms on the order of $O(X^{\frac{d-1}{d+1}+\epsilon})$, in contrast to GON methods which typically yield (at best) $O(X^{\frac{d-1}{d}+\epsilon})$.
- An identity for nonmaximal cubic rings proved in [4], replacing nonmaximal cubic rings with the cubic rings containing them. We omit the details here, but this reduces consideration of the function $\Phi_q \pmod{q^2}$ to a related function $\pmod q$.

- A nontrivial treatment of the Fourier transforms, in place of (43). It is also not enough to prove a uniform bound for $|\widehat{\Phi}_m(x)|$; instead, one proves bounds which depend on the $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$-orbit of $x \pmod{m}$.

The last of these is perhaps the most interesting for a newcomer to the subject. Along these lines we propose a rather involved exercise.

Exercise 18. *Let p be a prime, and define $\Psi_p(x)$ to be 1 if $\mathrm{Disc}(p) = 0$ as an element of $\mathbb{F}_p$, and $\Psi_p(x) = 0$ otherwise. Investigate the Fourier transform $\widehat{\Psi_p}(x)$.*

The most interesting result is an explicit formula (which will depend on the $\mathrm{GL}_2(\mathbb{F}_p)$-orbit of x). Numerical experiments and upper bounds are also of interest.

Note that a detailed analysis of this last function Ψ_p eventually leads to asymptotic formulas (with secondary terms and explicit power-saving error terms) for the number of cubic rings or fields whose discriminants are divisible by p. These are useful in sieve methods; see for example papers of Belabas and Fouvry [2] (which does not use the zeta function method, but whose results could have been proved in this manner) or Martin and Pollack [16].

As of the writing of this article, investigations of related exponential sums are underway by the author and by his Ph. D. student Daniel Kamenetsky.

6. What's required for quartic and quintic fields?

To date, the Shintani zeta function method has not yet succeeded in giving the asymptotic density of quartic and quintic fields. This question was the subject of work of Yukie going back to the 90's. His first step was to analyze the relevant Shintani zeta function, this time associated to the prehomogeneous vector space $(\mathrm{GL}(3) \times \mathrm{GL}(2), \mathrm{Sym}^2 k^3 \otimes k^2)$ of pairs of ternary quadratic forms. After a Herculean effort, involving nontrivial results from geometric invariant theory, the theory of Lie algebras, and the theory of Eisenstein series, he obtained a principal part formula for the quartic Shintani zeta function. However, his formula was expressed in terms of multiple distributions associated to singular orbits, and explicitly evaluating all of these distributions was too complicated to be feasible.

It would be very desirable, and it should be possible, to simplify and streamline Yukie's work, with an eye towards completing it and then extending it to the quintic case. One interesting contribution along these lines is due to Taniguchi [26], based on ideas of Sato (unpublished) and Kogiso [15], which streamlines the proof of the principal part formulas for the cubic Shintani zeta function. It should be possible to translate these ideas to the quartic case, and we hope to see this carried out in the near future.

We explain one more simplification that might be possible, and that is also a good project for the near future. The desire for this simplification can already be seen in the cubic case.

One disagreeable redundancy can be described as follows. If you solved Exercise 17, you observed that the sieved α terms in (45) gave the asymptotic density of cubic fields, and that the β terms gave the density of quadratic fields. However, we already knew the density of quadratic fields.

This invites the following question: in our complete proof of the Davenport-Heilbronn theorem (incorporating also Shintani's very substantial work, which we have merely quoted here), we are required to compute β and then observe that it coincides precisely with the maximal rings we don't want to count. Is there a way to omit computing β in the first place?

The answer is yes; see work of Shintani [24] and Taniguchi [25]. Nevertheless, their work is complicated, as indeed it presumably must be *if* one does not specialize to the case where essentially all nonmaximal rings have been sieved out. (The relationship between binary quadratic forms and reducible binary cubic forms is more complicated in the nonmaximal case.) Incorporating this specialization, one may ask if the answer can be distilled to a quick shortcut, suitable for generalizing and extending to the quartic case. We hope to investigate this question in the near future.

Bibliography to be updated. Will we have a separate bibliography for each chapter?

Bibliography

[1] K. Belabas, M. Bhargava, and C. Pomerance, *Error estimates in the Davenport-Heilbronn theorems*, Duke Math. J. **153** (2010), no. 1, 173–210.

[2] K. Belabas and E. Fouvry, *Sur le 3-rang des corps quadratiques de discriminant premier ou pseudo-premier*, Duke Math. J. **98** (1999), pp. 217-268.

[3] M. Bhargava, *The density of discriminants of quartic rings and fields*, Ann. of Math. (2) **162** (2005), no. 2, 1031–1063.

[4] M. Bhargava, A. Shankar, and J. Tsimerman, *On the Davenport-Heilbronn theorem and second order terms*, Invent. Math.

[5] K. Chandrasekharan and R. Narasimhan, *Functional equations with multiple gamma factors and the average order of arithmetical functions*, Ann. Math (2) **76** (1962), 93–136.

[6] B. Datskovsky and D. Wright, *The adelic zeta function associated to the space of binary cubic forms. II. Local theory*, J. Reine Angew. Math. **367** (1986), 27–75.

[7] B. Datskovsky and D. Wright, *Density of discriminants of cubic extensions*, J. Reine Angew. Math. **386** (1988), 116–138.

[8] H. Davenport, *Multiplicative number theory.*

[9] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic fields. II*, Proc. Roy. Soc. London Ser. A **322** (1971), no. 1551, 405–420.

[10] B. N. Delone and D. K. Faddeev, *The theory of irrationalities of the third degree (English translation)*, AMS, Providence, 1964.

[11] W. T. Gan, B. Gross, and G. Savin, *Fourier coefficients of modular forms on G_2*, Duke Math. J. **115** (2002), 105–169.

[12] H. Iwaniec and E. Kowalski, *Analytic number theory.*

[13] E. Landau, *Über die Anzahl der gitterpunkte in gewissen Bereichen*, Gött. Nachr. (1912), 687–771.

[14] E. Landau, *Über die Anzahl der gitterpunkte in gewissen Bereichen II*, Gött. Nachr. (1915), 209–243.

[15] Kogiso, ...

[16] G. Martin and P. Pollack, ...

[17] J. Nakagawa, *On the relations among the class numbers of binary cubic forms*, Invent. Math. **134** (1998), no. 1, 101–138.

[18] Y. Ohno, *A conjecture on coincidence among the zeta functions associated with the space of binary cubic forms*, Amer. J. Math. **119** (1997), no. 5, 1083–1094.

[19] D. Roberts, *Density of cubic field discriminants*, Math. Comp. **70** (2001), no. 236, 1699–1705.

[20] F. Sato, *On functional equations of zeta distributions*, Adv. Studies in Pure Math. **15** (1989), 465–508.

[21] M. Sato and T. Kimura, *A classification of irreducible prehomogeneous vector spaces and their relative invariants.* Nagoya Math. J. **65** (1977), 1–155.

[22] M. Sato and T. Shintani, *On zeta functions associated with prehomogeneous vector spaces*, Ann. of Math. (2) **100** (1974), 131–170.

[23] T. Shintani, *On Dirichlet series whose coefficients are class numbers of integral binary cubic forms*, J. Math. Soc. Japan **24** (1972), 132–188.

[24] T. Shintani, *On zeta-functions associated with the vector space of quadratic forms*, J. Fac. Sci. Univ. Tokyo Sect. I A Math. **22** (1975), 25–65.

[25] T. Taniguchi, *Distributions of discriminants of cubic algebras*, preprint, `arxiv.org/abs/math.NT/0606109`.

[26] T. Taniguchi, *A simple global theory of Shintani's zeta function*, forthcoming work.

[27] T. Taniguchi and F. Thorne, *Secondary terms in counting functions for cubic fields*, Duke Math. J.

[28] T. Taniguchi and F. Thorne, *Orbital L-functions for the space of binary cubic forms*, Can. Math. J.

[29] D. Wright, *The adelic zeta function associated to the space of binary cubic forms. I. Global theory*, Math. Ann. **270** (1985), no. 4, 503–534.

[30] D. Wright, *Twists of the Iwasawa-Tate zeta function*, Math. Z. **200** (1989), 209–231.

[31] D. Wright and A. Yukie, *Prehomogeneous vector spaces and field extensions*, Invent. Math. **110** (1992), no. 2, 283–314.

[32] A. Yukie, *Shintani zeta functions*, London Mathematical Society Lecture Note Series, 183, Cambridge University Press, Cambridge, 1993.

[33] Y. Zhao, *forthcoming work.*