Detecting square-free numbers via the explicit formula

Ghaith Hiary
(with Andrew Booker and Jon Keating)

Partial information

Given an integer d, we would like "partial information" about it.

Examples.

- Certify the compositeness of d without knowing its factors.
  – Solved in poly-log time by Agrawal-Kayal-Saxena (AKS).
- d has an odd or even number of (distinct) prime factors?
  – Doable in $d^{1/3+o(1)}$ time: AKS + check for factors $\leq d^{1/3}$.
  – Can be improved to $d^{1/6+o(1)}$ time (Pollard-Strassen).
  – However, in practice, faster to simply factor d using heuristically subexponential time algorithms.
- Does d have a simple (i.e. multiplicity one) prime factor?

  … etc.

Testing square-freeness

Question. How fast can the square-freeness of d be checked? Can it be done in subexponential time without having to factor?

Besides trial division, here's what's available:

- The Pollard-Strassen algorithm: Can find all factors of d less than B in $B^{1/2}d^{o(1)}$ time/space. $\Longrightarrow d^{1/6+o(1)}$ time/space.

  But slow, large memory requirements.

- Subexponential factoring algorithms; e.g. *The General number field sieve* (GNFS) expected to work in $\exp((\log d)^{1/3+o(1)})$ time (fastest available in this class).

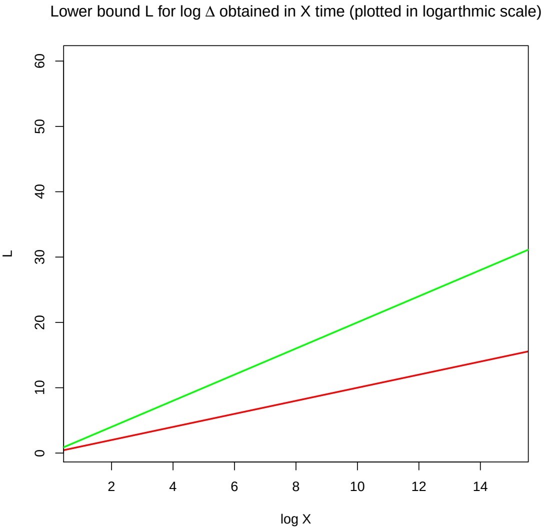
  Very successeful in practice. Best bet to learn about d.

  Unfortunately, GNFS does not yield partial information about d. Either find a factor or no info.

  Is there a more economical fast way?

Framing the question in terms of a lower bound

$d = m^2\Delta$, where Δ is square-free. How good a lower bound L on $\log \Delta$ can be obtained in time X?



Would like a method such that $L = (\log X)^\eta$ for some $\eta > 1$.

Then, a lower bound L for $\log|\Delta|$ costs $\exp(L^{1/\eta})$ time to obtain if true. This is subexp if $\eta > 1$.

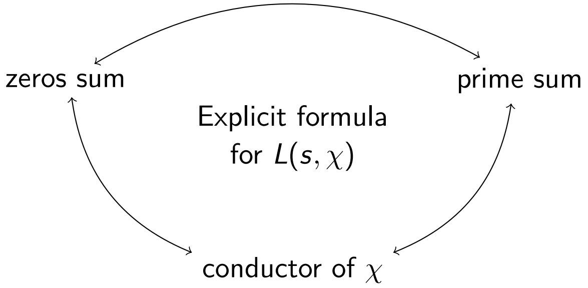
Notice that GNFS takes takes $\exp((\log d)^{1/3+o(1)})$ time regardless of the desired lower bound L.

Using the explicit formula

Let χ be a primitive Dirichlet character of conductor Δ.

Explicit formula for $L(s,\chi)$: Let $g(x)$ be a real even continuous piecewise differentiable compactly supported, and let $h(x) = \int_{\mathbb{R}} g(y) e^{ixy}\, dx$, then

$$g(0) \log|\Delta| = \sum_{\gamma} h(\gamma) + 2\Re \sum_{n \geq 1} \frac{\chi(n)\Lambda(n) g(\log n)}{\sqrt{n}} + \underbrace{\textit{Gamma contrib}}_{\textit{can be computed easily}}$$



- Explicit formula relates the zeros, the coefficients, and the conductor.
- Proved using the Euler product and the func. equation.

Real characters and positivity

Let Δ be a fundamental discriminant. Apply the explicit formula with $\chi(.)$ the Kronecker symbol $\left(\frac{\Delta}{.}\right)$, so

$$g(0)\log|\Delta| = \sum_{\gamma} h(\gamma) + 2\sum_{n\geq 1}\left(\frac{\Delta}{n}\right)\frac{\Lambda(n)g(\log n)}{\sqrt{n}} + \underbrace{\textit{Gamma contrib}}_{\textit{can be computed easily}}$$

Assume the generalized Riemann hypothesis for $L(s,(\Delta|.))$.

Use a test Fourier-pair (g,h) such that $h(x)\geq 0$. For example,

$$g(y) = \frac{1_{|y|<Y}}{Y}\left(1-\frac{|y|}{Y}\right), \quad h(x) = \int_{\mathbb{R}} g(y)e^{ixy}\,dx = \frac{\sin(xY/2)^2}{(xY/2)^2}.$$

Since $h(x)\geq 0$, zeros contribution $\sum_{\gamma} h(\gamma)\geq 0$.

So can simply drop $\sum_{\gamma} h(\gamma)$, and still get a lower bound on $|\Delta|$.

Therefore, we can get a lower bound without knowing the zeros γ.

A lower bound from the prime sum

If $g(x)$ is supported on $[-X, X]$, we therefore have

$$g(0)\log|\Delta| \geq 2\sum_{1\leq n\leq X}\left(\frac{\Delta}{n}\right)\frac{\Lambda(n)g(\log n)}{\sqrt{n}} + \textit{Gamma contrib.}$$

Now, let $d = m^2\Delta$, where Δ is square-free. Assume $d \equiv 1 \bmod 4$, so d is a fundamental discriminant. Assume $\left(\frac{d}{n}\right) \neq 0$, $1 \leq n \leq e^X$, so $\left(\frac{d}{n}\right) = \left(\frac{m^2}{n}\right)\left(\frac{\Delta}{n}\right) = \left(\frac{\Delta}{n}\right)$. (If $\left(\frac{d}{n}\right) = 0$, then it's even better, we find a factor!) Then we have

$$g(0)\log|\Delta| \geq 2\sum_{1\leq n\leq X}\left(\frac{d}{n}\right)\frac{\Lambda(n)g(\log n)}{\sqrt{n}} + \textit{Gamma contrib.}$$

Last, use quadratic reciprocity or Euler's criterion for fast computation of $\left(\frac{d}{n}\right)$ for $n = p^k$.

That is, we can compute $\left(\frac{d}{n}\right)$ fast without knowing its conductor.

Good and bad news

Explicit formula yields a lower bound on the *least period* of $\left(\frac{d}{n}\right)$:

$$g(0)\log|\Delta| \geq 2 \sum_{1\leq n\leq X} \left(\frac{d}{n}\right) \frac{\Lambda(n)g(\log n)}{\sqrt{n}} + \textit{Gamma contribution}$$

where $g(x)$ is supported on $[-X, X]$.

Is this a good lower bound? In general no!

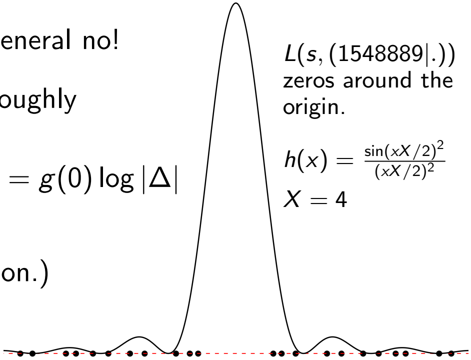
Zeros sum typically dominates, roughly

$$\sum_{\gamma} h(\gamma) \approx \frac{\log|\Delta|}{2\pi} \int_{\mathbb{R}} h(x)\,dx = g(0)\log|\Delta|$$

(view it as Monte Carlo integration.)

Unless possibly if big zero gap.

Example of a big zero gap



$L(s, (1548889|.))$ zeros around the origin.

$h(x) = \frac{\sin(xX/2)^2}{(xX/2)^2}$

$X = 4$

Large zero gaps

If there is a large zero gap, then we have a chance.

Center $h(x)$ around the big zero gap $\Longrightarrow \sum_\gamma h(\gamma)$ is likely small. This can be quantified as

Theorem. Assume the GRH, and let χ be a real character of conductor $|\Delta|$. Suppose that $L(1/2+it,\chi)$ has no zeros with imaginary part $(t_0, t_0+\delta)$ for some $t_0 \geq 1$ and $\delta > 0$. Then there is a Fourier pair $g(x)$ and $h(x)$ such that $h(x) \geq 0$, $g(x)$ is supported on $|x| \leq \delta^{-1} \log\log|t\Delta|$, and

$$\sum_\gamma h(\gamma) \ll \frac{g(0)}{\delta\sqrt{\log\log|t\Delta|}}.$$

(So the larger the zero gap $\delta \Rightarrow$ the shorter the prime sum that we need to evaluate.)

Looking for large gaps by twisting

Let $\mathcal{F} := \mathcal{F}(X)$ be the set of fundamental discriminants $|q| \leq X$. Assume $X = \Delta^{o(1)}$ as $\Delta \to \infty$.

Consider the following family of Dirichlet L-functions $\{L(s, (q\Delta|.), \ q \in \mathcal{F}\}$.

Let $\gamma_1(q\Delta)$ be the first zero of $L(1/2 + it, (q\Delta|.))$.

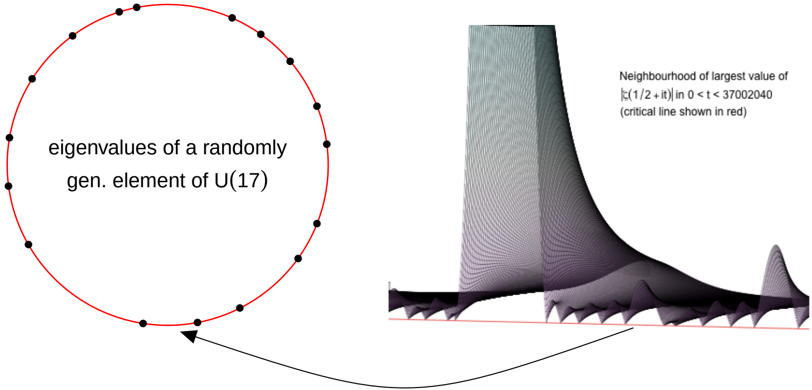
What do we expect the size of

$$\max_{q \in \mathcal{F}} \gamma_1(q\Delta)?$$

Note, on average, the zeros of $L(1/2 + it, (q\Delta|.))$, with $t \ll 1$ say, are spaced $\frac{1}{2\pi} \log(q|\Delta|) \sim \frac{1}{2\pi} \log(|\Delta|)$ apart.

Random matrix theory (RMT) and zero spacings

Suitably normalized zeros of an L-function, or a family of L-functions, have the same statistics (to leading order) as normalized eigenphases of random matrices from a compact matrix group (or matrix ensemble) for large but finite parameter; e.g. $A \in U(N) \longleftrightarrow \{\zeta(1/2+it),\ T \le t \le T+2\pi\}$, $N \leftrightarrow \log(T/2\pi)$.



How large a zero gap does RMT suggest?

Let $USp(2N)$ be the compact group of $2N \times 2N$ unitary matrices A satisfying $A^t JA = J$, where $J = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}$. Let $A \in USp(2N)$. The eigenvalues of A are $e^{\pm i\theta_1}, \ldots, e^{\pm i\theta_N}$.

The random matrix philosophy suggests that the lowest zero $\gamma_1(q\Delta)$, $q \in \mathcal{F}(X)$, $X = \Delta^{o(1)}$, is modeled by the lowest eigenphase θ_1 of matrices from $USp(2N)$ with $2N = \log(|\Delta|)$.

Theorem. Fix $\delta > 0$, Let $\delta < \beta < 2 - \delta$, $M = \lfloor \exp(N^\beta) \rfloor$. Suppose $A_1, \ldots, A_M \in USp(2N)$ are chosen indpendently and uniformly with respect to the Haar probability measure on $USp(2N)$. Let $\theta_1(m)$ denote the first eigenphase of A_m. Then for any $\epsilon > 0$, we have

$$\mathbb{P}_N \left(\max_{1 \le m \le M} \theta_1(m) \ge (2 - \epsilon)\, N^{\beta/2 - 1} \right) \to 1\,, \qquad as \quad N \to \infty\,.$$

Heuristic running time

Conjecture. Fix $0 \leq \beta < 1$. Let $\gamma_1(q\Delta)$ be the first zero of $L(1/2+it,(q\Delta|.))$, $X = \exp(\log\Delta)^{\beta}$, and $\mathcal{F} := \mathcal{F}(X)$ be the set of fundamental discriminants $|q| \leq X$. Then

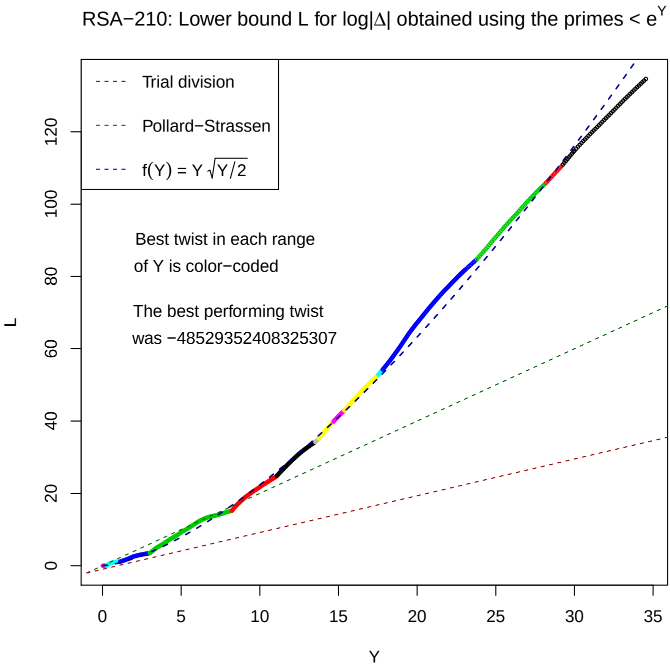
$$\log \max_{q\in\mathcal{F}} \gamma_1(q\Delta)/\log\log|\Delta| \sim \beta/2 - 1\,,$$

as $|\Delta| \to \infty$ through fundamental discriminants.

So if we sample the fundamental discriminants $|q| \leq \exp((\log|\Delta|)^{\beta})$, then by the conjecture we expect to find at least one q where there is a gap of size $(\log|\Delta|)^{\beta/2-1}$.

Want to ensure that $h(x)$ decays quickly outside of zero gap $\Longrightarrow$ take $g(y)$ to be supported on roughly $|y| \leq (\log|\Delta|)^{1-\beta/2}$.

Optimizing: sampling time = prime sum computation time, so $\beta = 1 - \beta/2 \Longrightarrow \beta = 2/3$. So by putting in effort $X = e^{Y}$, we expect a lower bound like $Y^{3/2}$.



RSA−210: Lower bound L for log|Δ| obtained using the primes < e^Y
Trial division
Pollard−Strassen
f(Y) = Y √(Y/2)
Best twist in each range
of Y is color−coded
The best performing twist
was −48529352408325307
L
Y
0
20
40
60
80
100
120
0
5
10
15
20
25
30
35

Example application

RSA challenge number `RSA-210` has 210 decimal digits (696 bits):

2452466449002782119765176635730880184670267876783327597434144517150616008300
3858721695220839933207154910362682719167986407977672324300560059203563124656
121846581790410013185929961993381701214933503487587055106 7

The GNFS has so far not been able to tell us any information about `RSA-210` (as it remains unfactored), but using the method I described we proved

<u>Theorem.</u> Assume the GRH for quadratic Dirichlet L-functions. Then the RSA challenge number `RSA-210` is not square-full; i.e. it has at least one prime factor of multiplicity 1.

Can we rescue part of the zeros contribution?

```
Using the primes < 1e7 and -65123121667 twist

v          point              w[v]
-----------------------------------------
45         0.3560000           4.0000000
46         0.3640000           1.0000000
71         0.5640000           1.0000000
98         0.7800000           1.5156296
99         0.7880000           2.5486078
146         1.1640000           4.4663347
-----------------------------------------
prime contr  :  44.65870
zeros contr  :  2.49460
improvement  :  5.59 %
logd lbound  :  47.15330
-----------------------------------------
# variables          :  500
# integer vars       :  45
interval covered     :  4.00000
grid spacing         :  0.00800
```