A Study on Merten's conjecture

Kailash Meiyappan, Srikar Varadaraj

November 15, 2017

1 Introduction

Definition 1 *Möbius function is defined for positive integers n as*

$$\mu(n) = \left\{ \begin{array}{ll} 1 & n = 1 \\ 0 & a^2|n \text{ for some } a > 1 \\ (-1)^r & n \text{ has } r \text{ distinct prime factors} \end{array} \right\}$$

Definition 2 *Merten's Function is defined as*

$$M(n) = \sum_{k=1}^{n} \mu(k)$$

These functions are the central objects of study in the statement of Merten's conjecture. Before we state the conjecture itself, we look at the main algorithm we will be using to study number theoretic results.

1.1 LLL algorithm

The LLL algorithm is a polynomial time lattice reduction algorithm invented by Lenstra, Lenstra and Lovasz (1982). Given a basis B of size d with $n-$dimensional integer coordinates for a lattice L with $d \leq n$, LLL finds a "reduced basis", or basis vectors that are short and "nearly orthogonal" in time poly(parameters).
Since 1982, number theorists have used LLL to prove results on Diophantine Approximations. Noam Elkies [Elk00] used LLL to find rational points near special curves. De Weger and Dokchitser [Dok04] have even suggested methods for tackling the long-standing ABC conjecture using LLL.

1.2 Merten's Conjecture

Definition 3 *Merten's conjecture states that* $\forall\ n > 1$,

$$m(n) = \frac{|M(n)|}{\sqrt{n}} < 1$$

n	$m(n)$
1	1
2	0
3	0.57735
4	0.5
5	0.894427
6	0.408248
7	0.755929
8	0.707107
9	0.666667
10	0.316228
11	0.603023
12	0.57735
13	0.83205
14	0.534522
15	0.258199
16	0.25
17	0.485071
18	0.471405
19	0.688247
20	0.67082
21	0.436436
22	0.213201
23	0.417029
24	0.408248
25	0.4
400	0.05
800	0.0353553
1200	0.11547
1600	0.175
2000	0.111803
2400	0.244949
2800	0.415761
3200	0.194454
3600	0
4000	0.142302
4400	0.13568
4800	0.129904
5200	0
5600	0.120268
6000	0
6400	0.3125
6800	0.194029
7200	0
7600	0.0458831
8000	0.0111803

1.3 Disproof of Merten's conjecture

The first disproof of Merten's conjecture came about from a clever application of LLL. Odlyzko and Riele [OtR85] showed that Merten's conjecture is false and in fact:

$$lim_{n\to\infty} inf(m(n)) \leq -1.009$$

and

$$lim_{n\to\infty} sup(m(n)) \geq 1.06$$

In 2006, the result was improved by Kotnik and Riele [KtR06] to show that:

$$lim_{n\to\infty} sup(m(n)) \geq 1.2184$$

The authors note in [OtR85] that the disproof is indirect and does not produce a single value for which $|M(n)| > n^{\frac{1}{2}}$. In [Pin87], Jacob Pintz proved that $\exists$ a counterexample to Merten's conjecture for some $n < exp(3.21 \times 10^{64})$, hence giving another disproof of the conjecture. However, to date, we do not know the first value $n = \alpha$ for which Merten's conjecture does not hold. We do know that $\alpha > 10^{14}$ [KtR06], but it remains computationally difficult with current methods.

1.4 Organization and Goals

This study is intended to be a clear exposition of the Merten's conjecture, its history and motivation, consequences and the first disproof by Odlyzko and Riele. Through this, we hope to gain a deep understanding of LLL and important number theoretic functions.
By analyzing a key application of LLL to a central problem in Number Theory, we hope to gain insight into how LLL can be used elsewhere on open problems. Finally, we look at current active open problems in number theory and see how LLL might be applied.
We would like this work to serve as a resource for further research.

2 Consequences of Merten's Conjecture and History

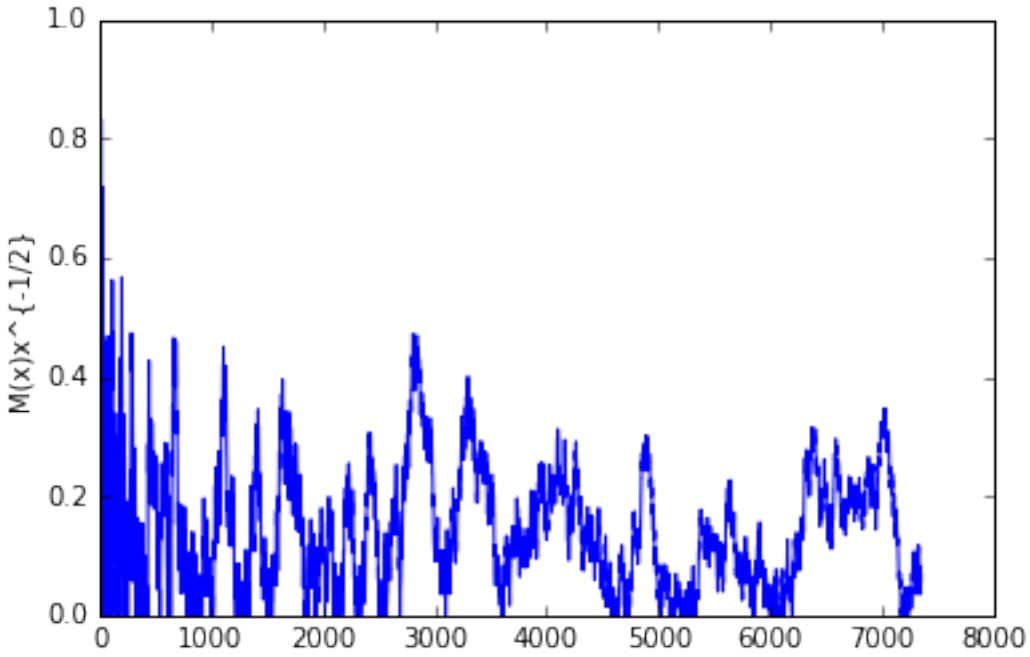

Figure 1: The value of $M(x)x^{-1/2}$. Clearly it is not going over 1.

2.1 The Riemann Hypothesis

Definition 4 *A Dirichlet Series is any series of the form:*

$$D(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

where s and a_n are complex.

Definition 5 *The Gamma function is given by the analytical continuation of the factorial function given by:*

$$\Gamma(n) = (n-1)!$$

where n is a positive integer, to the domain of complex numbers except for non-positive integers. It is defined as

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$$

for complex z with $Re(z) > 0$. And

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}$$

for z except non-positive integers.

Definition 6 *The Reimann Zeta function is given by the analytical continuation of the Dirichlet Series:*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for complex s with Re(s) > 1 to complex numbers.

The definition above converges only at $Re(s) > 1$. For other values of s, $\zeta(s)$ is computed as:

$$\zeta(s) = 2^s\pi^{s-1}sin(\frac{\pi s}{2})\Gamma(1-s)\zeta(1-s) \tag{1}$$

From the functional equation above, it is clear that at negative even integers, $\zeta(-2n) = 0$ because $sin(n\pi) = 0$. These are called the trivial zeros of the Reimann Zeta function. It is unclear if there are other zeros of the Zeta function and the problem of finding zeros of the zeta function is in fact the Reimann hypothesis.

Reciprocal of Reimann Zeta Function

The reciprocal of the Reimann Zeta function can be expressed in terms of the Mobius function and a Dirichlet series as

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} \tag{2}$$

for complex s with $Re(s) > 1$. To show that this is true, we first show the following theorem (Euler's product formula):

$$\zeta(s) = \prod_{p\ prime} \frac{1}{1-p^{-s}} \tag{3}$$

Proof:

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \dots$$

So,

$$\frac{1}{2^s}\zeta(s) = \frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \dots$$

Subtracting, we get rid of terms with 2^s in the denominator

$$(1 - \frac{1}{2^s})\zeta(s) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \dots$$

So,

$$\frac{1}{3^s}(1 - \frac{1}{2^s})\zeta(s) = \frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \frac{1}{21^s} + \dots$$

Subtracting, we get rid of terms with 3^s in the denominator

$$(1 - \frac{1}{3^s})(1 - \frac{1}{2^s})\zeta(s) = 1 + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{11^s} + \frac{1}{13^s} + \frac{1}{17^s} + \frac{1}{19^s} + \frac{1}{23^s} + \frac{1}{25^s} + \dots$$

And extending this argument to all primes,

$$...(1 - \frac{1}{5^s})(1 - \frac{1}{3^s})(1 - \frac{1}{2^s})\zeta(s) = 1$$

Or,

$$\frac{1}{...(1-\frac{1}{5^s})(1-\frac{1}{3^s})(1-\frac{1}{2^s})} = \zeta(s)$$

which is equivalent to

$$\zeta(s) = \prod_{p\ prime} \frac{1}{1-p^{-s}}$$

□ Now to prove that

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$$

We use the result 3

$$\frac{1}{\zeta(s)} = \prod_{p\ prime} (1-p^{-s}) = (1-\frac{1}{2^s})(1-\frac{1}{3^s})(1-\frac{1}{5^s})\dots$$

Expanding this, we get

$$\frac{1}{\zeta(s)} = 1 + \sum_{nprime} \frac{-1}{n^s} + \sum_{n=p_1p_2} \frac{-1}{p_1^s}\frac{-1}{p_2^s} + \sum_{n=p_1p_2p_3} \frac{-1}{p_1^s}\frac{-1}{p_2^s}\frac{-1}{p_3^s} + \dots$$

which is

$$= \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$$

This concludes the proof of the relation between an infinite series containing the mobius function and the reciprocal of the zeta function.

Reimann Hypothesis

The Reimann Hypothesis was introduced by Bernhard Reimann in 1859. The hypothesis is that all non-trivial zeros of the Reimann Zeta function are located at complex numbers with real part equal to half. In other words, if a complex number s is not a negative even integer, then:

$$\zeta(s) = 0 \Rightarrow Re(s) = \frac{1}{2}$$

Proof of Reimann's Hypothesis assuming Merten's Conjecture

Let $\sigma = Re(z) > 1$ for a complex number z. We also note that $M(x)$ is defined for positive reals as $M(\lfloor x \rfloor)$. Then, from 2,

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \sum_{n=1}^{\infty} \frac{M(n)-M(n-1)}{n^s} = \sum_{n=1}^{\infty} M(n)\left[\frac{1}{n^s} - \frac{1}{(n+1)^s}\right]$$

$$= \sum_{n=1}^{\infty} M(n) \int_n^{n+1} \frac{s dx}{x^{s+1}} = s\sum_{n=1}^{\infty} \int_n^{n+1} \frac{M(x)dx}{x^{s+1}} = s\int_1^{\infty} \frac{M(x)dx}{x^{s+1}} \tag{4}$$

This is true because $M(x)$ is a constant in the interval $[n, n+1)$. If Merten's conjecture were true, then since $M(x) \leq x^{1/2}$, this would imply that the last integral in 4 converges for $\sigma = Re(s) > 1/2$, giving the reciprocal of the Reimann Zeta function a continuation from $\sigma > 1$ to $\sigma > 1/2$. This would imply that $\zeta(s)$ has no zeros in the interval $\sigma > 1/2$. From the functional equation 1, this is equivalent to the Reimann Hypothesis. It can also be proved from 4 that all zeros of the Reimann zeta function must then be simple zeros, i.e. with multiplicity 1 for $\sigma > 1/2$ as follows:

$$|\frac{1}{\zeta(s)}| \leq |s| \int_1^{\infty} \frac{x^{1/2}dx}{x^{\sigma+1}} = \frac{|s|}{\sigma - (1/2)} \tag{5}$$

Now suppose that

$$s = \frac{1}{2} + i\gamma$$

is a such that it is a zero of multiplicity k. Then, for some $a > 0$,

$$|\zeta(s+u)| \approx au^k \text{ as } u \to 0^+$$

However, this is inconsistent with 5.

3 Key Techniques

3.1 LLL algorithm

Definition 7 *Given a set of vectors $V = (v_1, \ldots, v_n)$, a lattice Λ is defined as*

$$\Lambda = \sum_{i=1}^{n} a_i v_i$$

where $a_i \in Z$

The problem of finding the shortest vector in the lattice, i.e. the vector with the lowest norm is not known to be in P.

As a special case, the problem of finding the shortest vector in a one dimensional lattice is simply the gcd, which can be solved in polynomial time using the Euclidean algorithm.

For example, if $v_1 = (10, 0)$ and $v_2 = (16, 0)$, then the shortest vector is given by $(2, 0) = -3(10, 0) + 2(16, 0)$

The Lenstra Lenstra Lovasz algorithm[LLL82] or L^3 algorithm is a lattice basis reduction algorithm that allows for finding an approximate solution to the shortest vector problem. Given an input basis B = $b_1, b_2, ..., b_n$ with integer coordinates in n dimensional space, the LLL algorithm computes n linearly independent vectors V = $v_1, v_2, ..., v_n$ such that they obey the following properties for a certain fixed value of δ in the range (0.25,1]: If B* is the basis obtained after performing a Gram Schmidt orthogonalization (without normalization) and the Gram Schmidt coefficient μ_{ij} = $\frac{<b_i, b_j^*>}{b_j^* . b_j^*}$ then

- For 1 <= i < j <= n, μ_{ij} <= 0.5
- (Lovasz condition) For k = 2, 3, ..., n $|b_k^*|^2 >= (\delta - \mu_{k,k-1}^2)|b_{k-1}^*|^2$

Algorithm 1: LLL Algorithm

function LLL(B) ▷ Where B = $(b_1, ..., b_n)$ is the basis
    Compute B* = $b_1^*, b_2^*, ..., b_n^*$ ▷ where B* is the Gram-Schmidt basis but not normalized
    for $i = 2$ to n **do**
        for $j = i - 1$ to n **do**
            b_i = b_i - $c_{i,j} b_j$, where $c_{i,j} = \lfloor \frac{<b_i, b_j^*>}{b_j^* . b_j^*} + 0.5 \rfloor$
        end for
    end for
    if $\exists k$ such that Lovasz condition is not satisfied **then**
        swap b_k and b_{k-1}
        return LLL(B)
    end if
    return B
end function

It is easy to show the following lemma[LLL82] follows from the two LLL conditions.

Lemma 1 *If $v_1, v_2, ..., v_n$ is a LLL reduced basis of a lattice L, then*

$$|v_1| \leq \left(\frac{4}{4\delta - 1}\right)^{(n-1)/2} \lambda_1(L)$$

where $\lambda_1(L)$ is the shortest vector of L; and

$$\prod_{i=1}^{n} |v_i| \leq \left(\frac{4}{4\delta - 1}\right)^{n(n-1)/4} det(L)$$

where $det(L)$ is the determinant of the lattice L and $|v_i|$ is the Euclidean norm of the vector v_i.

3.2 LLL Runtime

The LLL algorithm as presented does not necessarily look like it runs in polynomial time.
To prove that LLL algorithm in fact does run in polynomial time, we must show three things:

1. Number of steps per function call is polynomial.

2. Number of function calls is polynomial.

3. The length of the vectors does not blow up.

The first item is obvious. We run n^2 iterations and all we do is to perform vector operations. All of these are clearly polynomial time.
The second is not so trivial. To prove this, we define a new variable D as

$$D = \prod_{k=1}^{n} det(L(b_1, \ldots, b_k))^2$$

Where $L(b_1, \ldots b_k)$ is the lattice defined by the first k vectors. We will show that in each function call, this term D decreases by a factor of δ.
Clearly, the first n^2 iterations do not change D, since we are only adding and subtracting integer combinations of vectors from each other, this does not change the Gram-Schmidt reduced basis B^*, and D can be represented in terms of vectors of B^*.
Now suppose that vectors b_i and b_{i+1} were swapped. For $j < i$, clearly $L(b_1 \ldots b_j)$ is unchanged. For $j > i$, the only change is that the two vectors i and i+1 are swapped and again the lattice is unchanged. For $j = i$, however, we have that b_i is replaced by b_{i+1}. And so, if the new product is D', we get:

$$\frac{D}{D'} = \frac{\prod_{j \leq i} |b_j^*|^2}{(\prod_{j<i} |b_j^*|^2)|b_{i+1}^*|^2} = \frac{|b_i^*|^2}{|b_{i+1}^*|^2}$$

We know that the swap only happens when the Lovasz condition is not satisfied, i.e.

$$|b_{i+1}^*|^2 < (\delta - \mu_{i+1,i})|b_i^*|^2 < \delta |b_i^*|^2$$

Or,

$$\frac{|b_i^*|^2}{|b_{i+1}^*|^2} \geq \frac{1}{\delta}$$

And so,

$$D' \leq \delta D$$

This is after one swap. After k swaps, we get:

$$D^{(k)} \leq \delta^k D$$

Since D is a positive number, and is polynomial, we know that the algorithm must terminate at some point, say when $D = 1$.

$$D^{(k)} \leq 1$$
$$\Rightarrow \delta^k D \leq 1$$

which gives us

$$k \leq -\log(D)/\log(\delta)$$

or

$$k \leq \log_{\frac{1}{\delta}}(D)$$

Since δ is a constant, this is just $O(\log(D))$ recursive calls.
The last thing we need to show is that these vector magnitudes do not blow up in size.

$$|b_i^*|^2 = \frac{D_i}{D_{i+1}} \leq D$$

$$|b_i|^2 = |b_i^*|^2 + \sum_{j<i} \mu_{i,j}^2 |b_j^*|^2 \leq D + (n/4)D \leq nD$$

since the μ terms are less than half after the first n^2 iterations. We also need to show that the rational numbers can be written precisely. The only rational numbers we use with a denominator not equal to one are the μ terms. The μ terms have a denominator of $|b_i|^2$, which must clearly divide D_i. So, the denominator is smaller than D. Since the μ terms are less than half, the numerator must also be smaller than D. So every μ term can be written as a fraction with both numerator and denominator using $\log(D)$ bits.
This proves that the vectors have polynomially bounded size.
The total time taken by the algorithm is

$$time = t_1 t_2 t_3$$

where t_1 is the number of operations taken per recursive call, t_2 is the time taken per operation i.e. the size of the terms and t_3 is the number of recursive calls. We have proved that each of these three terms is polynomial and so the entire algorithm terminates in polynomial time.

4 Disproof of Merten's Conjecture

4.1 Initial Ideas

In 1942, [Ing42] showed that

$$\frac{M(x)}{\sqrt{x}}$$

might take large values and his idea did not involve actually computing $M(x)$. As we saw before, one of the main reasons Mertens and Stieljes believed the Merten's conjecture to be true was because it holds for large values of x and any counterexample would have to be extremely big. In particular, defining:

$$m(y) = \frac{M(x)}{\sqrt{x}} = \frac{M(y)}{e^{y/2}}, x = e^y$$

and

$$m_1 = lim_{n\to\infty} \inf(m(y)), m_2 = lim_{n\to\infty} \sup(m(y))$$

and

$$h(y,T) = 2 \sum_{0<\gamma<T} [(1 - \frac{\gamma}{T}) cos(\frac{\pi\gamma}{T}) + \frac{1}{\pi} sin(\frac{\pi\gamma}{T})] \frac{cos(y\gamma - \psi_\gamma)}{|\rho\zeta'(\rho)|}$$

with $\rho = \beta + i\gamma$ as the simple complex zeros of the Riemann-zeta function, $\beta = \frac{1}{2}$, $\psi_\gamma = arg(\rho\zeta'(\rho))$. Ingham [Ing42] proved that assuming the Riemann hypothesis and the simple-zero conjecture, then:

$$\forall y_0 \in \mathbb{R}, m_1 \leq h(y_0, T) \leq m_2$$

and any $h(y,T)$ is approximated arbitrarily closely and infinitely often by $\frac{M(x)}{\sqrt{x}}$. It can also be shown that Ingram's theorem is optimal in the sense that $\{h(y,T)\}$ is the set of accumulation points of the sequence $\frac{M(n)}{\sqrt{n}}$.
We now have an outline of the disproof:

- Merten's conjecture implies the Riemann hypothesis. It also implies that the zeros of the Riemann zeta function are simple.
- $h(y,T)$ is a good approximation to $\frac{M(x)}{\sqrt{x}}$.
- We want to find large values of $h(y,T)$ and y, T corresponding to these large values. This is the part where LLL is used.
- Finding large values of $h(y,T)$ implies the existence of counterexamples to the Merten's conjecture.
- However, RH $\implies$ large values of $h(y,T)$ $\implies$ Merten's conjecture does not hold. Since Merten's conjecture $\implies$ RH, we then have a contradiction!

- Therefore, Merten's conjecture cannot hold.

We now need to understand how the seemingly complicated function $h(y,T)$ behaves and how to go about finding large values of $|h(y,T)|$.

4.2 Large values of $h(y,T)$

To begin with, note that:

$$(1-\alpha)cos(\alpha\pi) + \frac{sin(\pi\alpha)}{\pi}$$

is non-negative for $0 \leq \alpha \leq 1$. Further, note that:

$$\sum_{\rho} \frac{1}{\rho\zeta'(\rho)}$$

diverges, so it is reasonable to expect that if we can choose a value of y such that:

$$y\gamma - \psi_\gamma \approx 2n\pi$$

for integral n, and for all possible γ, then,

$$cos(y\gamma - \psi_\gamma) \approx 1$$

which implies that $h(y,T)$ can be made arbitrarily large. These statements can be made mathematically precise, but they obscure the more important ideas. We will now need a theorem from the theory of Diophantine Approximations to see why it may be possible to find a y to satisfy a series of linear systems as described above.

4.3 Kronecker's Theorem

Kronecker's theorem is a result in diophantine approximations applying to several real numbers x_i, for $1 \leq i \leq m$, that generalizes Dirichlet's approximation theorem to multiple variables.
The classical Kronecker approximation theorem is formulated as follows:

Theorem 1 *Given real n-tuples* $\alpha_i = (\alpha_{i1}, ..., \alpha_{in}) \in \mathbb{R}^n$ *and* $\beta = (\beta_1, ..., \beta_n) \in \mathbb{R}^n$*, then the condition:*

$$\forall\epsilon\exists q_i, p_j \in \mathbb{Z} : |\sum_{i=1}^{m} q_i\alpha_{ij} - p_j - \beta_j| < \epsilon, 1 \leq j \leq n$$

holds iff $\forall r_1, ..., r_n \in \mathbb{Z}, i = 1, .., m$ with

$$\sum_{j=1}^{n} \alpha_{ij} r_{ij} \in \mathbb{Z}, i = 1, ..., m \tag{6}$$

the number:

$$\sum_{j=1}^{n} \beta_j r_j \tag{7}$$

is also an integer. In other words, the vector β can be approximated to arbitrary accuracy by linear combinations of the α_i vectors and integer vectors if the equation 6 $\implies$ equation 7 is an integer.
Applying the theorem to our case,

$$\forall\epsilon > 0, \exists y, m_\gamma \in \mathbb{Z} \text{ such that } |y\gamma - \psi_\gamma - 2\pi m_\gamma| < \epsilon \tag{8}$$

for all $\gamma \in (0,T)$, allowing us to make $|h(y,T)|$ arbitrarily large, as required. However, the above only holds if the γ we consider are linearly independent over the rationals.
Based on computational evidence, there is no reason to suspect that γ values are dependent. Hence, it seems reasonable to suspect that we can make $|h(y,T)|$ reasonably large.
Further evidence for why Merten's conjecture might be false came from studies which noted that if $\frac{M(x)}{\sqrt{x}}$ is bounded then there would be infinitely many linear relations over the integers between γ_i, where γ_i refers to the imaginary part of the non-trivial roots of the zeta function.

4.4 Application of LLL and Finding large $h(y,T)$

The early efforts of mathematicians led to good guesses of y based on heuristic analysis, which allowed them to find large values of $|h(y,T)|$. In this way, one of the first promising results was:

$$m_2 \geq 0.779, m_1 \leq -0.638 \tag{9}$$

These results were improved throughout the years by the exact same method of computing the first N zeros of the zeta function and then finding an approximate solution y to the Diophantine approximation equation (8), allowing us to maximize $|h(y,T)|$.
The general approach, then, was to find a y such that each of:

$$\eta_j = y\gamma_j - \psi_j - 2\pi m_j, 1 \leq j \leq n \tag{10}$$

is small, where

$$\psi_j = arg(\frac{1}{2} + i\gamma_j)\zeta'(\frac{1}{2} + i\gamma_j) \tag{11}$$

Here we consider n non-trivial zeros of the zeta function. It is precisely to this problem that LLL was applied by Odlyzko and Riele [OtR85]. They consider the following lattice:

$$\begin{bmatrix} -\alpha_1\psi_1 2^v & \alpha_1\gamma_1 2^{v-10} & 2\pi\alpha_1 2^v & \dots & 0 \\ -\alpha_2\psi_2 2^v & \alpha_2\gamma_2 2^{v-10} & 0 & 2\pi\alpha_2 2^v & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\alpha_n\psi_n 2^v & \alpha_n\gamma_n 2^{v-10} & 0 & \dots & 2\pi\alpha_n 2^v \\ 2^v n^4 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{bmatrix}$$

where $v \in \mathbb{Z}$ such that $2n \leq v \leq 4n$ and:

$$\alpha_j = |(\frac{1}{2} + i\gamma_j)\zeta'(\frac{1}{2} + i\gamma_j)|^{-1/2} \tag{12}$$

Applying the LLL algorithm to this $(n+2)$ by $(n+2)$ lattice generates a reduced basis $v_1, ..., v_{n+2}$ with some short vectors. The remaining algorithm goes as follows:

- Let the basis vectors of the initial lattice be denoted by $u_1, ..., u_{n+2}$.
- Note that the $(n+1)$th coordinate of all the original vectors is a multiple of $2^v n^4$.
- After applying LLL and getting the reduced basis $\{v_i\}$, we still have the $(n+1)$th coordinate as integer multiples of $2^v n^4$.
- It is immediately seen that $2^v n^4$ contributes the most to the magnitude of the vector. In order for the largest magnitude of the reduced basis vector to be small, must have the $(n+1)$th coordinate to be $\pm 2^v n^4$.
- Experimentally, the LLL algorithm happens to agree with the above heuristic analysis.

From here, we consider that a single vector in the reduced basis has a nonzero $(n+1)$th coordinate and we take it to be $2^v n^4$ without loss of generality. The j'th coordinate for this vector is:

$$z(\alpha_j\gamma_j 2^{v-10}) - \alpha j\psi_j 2^v - m_j 2\pi\alpha_j 2^v \tag{13}$$

with the $(n+2)$th coordinate simply being z. To minimize the length of the vector with these coordinates, note that the magnitude of the term in (13) above has to be small. By minimizing the above, we find large values of $h(y,T)$, hence completing the story.
One curiosity is that there doesn't seem to be rigorous mathematical justification in literature for why LLL performs well on the problem above. It simply does well according to experiments. Riele himself states, " This choice was made on heuristic grounds and did not guarantee that they would disprove the Mertens conjecture, but in the end they did".
The details of the calculations are tedious, but the following steps lead to the final disproof of Merten's conjecture:

- The first 400 γ are ordered: $\gamma_1^*, \gamma_2^*, ...$ such that the values $|\rho_j^* \zeta'(\rho_j^*)|^{-1}$ are monotonically decreasing and maximal.
- Note that the sum $2\sum_{j=1}^{n} |\rho_j^* \zeta'(\rho_j^*)|^{-1} >= 1$ for $n \geq 54$.
- In [OtR85], the authors chose $T = \gamma_{2000}$ (essentially the first 2000 zeros of the zeta function were considered). LLL was used to find a value of y that made η_j small. For $n = 70$, two values of y were found that gave $|h(y,T)| > 1$. This was enough to disprove Merten's conjecture without directly finding a counter example for $m(n) \leq 1$.
- Since then, the values of m_1 and m_2 have been improved by using more zeros of the zeta function, more computational power and by summing more terms in the description of $h(y,T)$.

In summary, the way LLL was used for the Disproof of Merten's Conjecture exemplifies the power of LLL in general and its shortcomings. LLL allows one to "shortcut" computations by finding "approximate" solutions very quickly. However, it does not induce additional number theoretic structure to a question, or provide deeper insight into the workings of number theoretic functions. It also seems to be that while LLL performs well in practice and gives results that approximate the shortest vector quite well, there is no mathematically rigorous understanding for why LLL works so well on average.

5 Alternate Directions and Future Work

The Riemann Hypothesis is considered one of the most important unsolved problems in number theory and mathematics in general. Any deep insight into number theoretic functions related to the RH would be considered as huge progress.
One particular line of thought in approaching the RH has been to look at finite fields and elliptic curves over finite fields. There are clear analogies for the RH in these cases, and in particular, the RH over finite fields and for elliptic curves over finite fields have actually been proven. It takes a lot more definitions and mathematically machinery to introduce the RH in these settings, but one might wonder if a similar Merten's conjecture can be postulated for finite fields and elliptic curves over finite fields. In 2012, Peter Humphries proposed the following formulation:

Conjecture 1 *Let E be an elliptic curve over $\mathbb{F}_q$ and let $M_{E/\mathbb{F}_q}(X)$ be the summatory function of the Mobius function of $E/\mathbb{F}_q$. Then for all sufficiently large positive integers X,*

$$|M_{E/\mathbb{F}_q}(X)| \leq q^{X/2} \tag{14}$$

There are some results regarding this conjecture, but it would be an interesting exercise to see how LLL could be used to find counter-examples for the conjecture. Does the LLL approach generalize for elliptic curves over finite fields? What what a lattice for the above problem look like? Is there an analogous $h(y,T)$ function that we can consider for the above problem?
These are some questions that the authors might pursue in the future. The field is very fertile and exciting, and the open problems are indeed extremely tough to crack.

6 Number Theory, LLL and Other Applications

Till this point, we have primarily discussed the application of LLL to a very specific problem in Number Theory regarding the Riemann Hypothesis. However, LLL has been extremely useful in the computational theory of diophantine approximations over the past thirty years. For example, LLL can be used to find an approximation of a real number α using rational numbers.[NV09] This is important, because there is no way to represent a real number to arbitrary precision on a computer, while rational numbers can be represented as their numerator and denominator. The most popular way to do this is by using continued fractions. The LLL algorithm provides an alternative way to get good approximations.

Definition 8 *An n-ary quadratic form is a polynomial of n variables of degree two such that it is homogeneous (every monomial is of degree two).*

A binary quadratic form

$$q(x, y) = ax^2 + bxy + cy^2$$

is isomorphic to a matrix

$$\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$$

The original LLL paper [LLL82] suggests a quadratic form:

$$q(x, y) = M(\alpha' x - y)^2 + \frac{1}{M}x^2$$

Here, α' is a decimal approximation of α to precision of $\frac{1}{M}$. If M is large and we successfully find a short vector (x, y), then $q(x, y) = O(1)$, and so $x = O(\sqrt{M})$ and $a'x - y \approx \frac{1}{M}$

Definition 9 *A Gram Matrix or Gramian of a set of vectors $V = (v_1, ...v_n)$ is defined as*

$$G(V) = V^T V$$

such that each entry

$$G(V)_{i,j} = v_i.v_j$$

the dot product of v_i and v_j.

We consider the matrix corresponding to this quadratic form:

$$\begin{bmatrix} \alpha'^2 M + \frac{1}{M} & -\alpha' M \\ -\alpha' M & M \end{bmatrix}$$

This has a determinant of 1. If we run LLL on this matrix, we get a new matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

such that a short vector for (x, y) is (a, c). Using the bound from 1, we know that

$$|b_1| \leq 2^{(n-1)/4} det(L)^{1/n}$$

which, from [Coh93] gives us:

$$q(a, c) \leq 2^{(n-1)/2} det(q)^{1/n}$$

And so we have,

$$q(a, c) \leq \sqrt{2}$$

This implies that:

$$|\alpha' a - c| \leq \frac{2^{1/4}}{\sqrt{M}}$$

and

$$|a| \leq 2^{1/4}\sqrt{M}$$

And from the fact that α' is close to α, i.e.

$$|\alpha' - \alpha| \leq \frac{1}{M}$$

we get

$$|\alpha x - y| \leq \frac{2^{5/4}}{\sqrt{M}}$$

and so α is very close to y/x. We note that we start with an approximation α', which is usually a decimal approximation for a real number, say 3.14159 for π, and end up with a fraction $\frac{y}{x}$, both of which are usually $O(\sqrt{M})$. This does not gain or lose information, but rather spreads information from being primarily in the numerator for α' to being in both the numerator and denominator.

References

[Coh93] Henri Cohen. *A Course in Computational Algebraic Number Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1993.

[Dok04] Tim Dokchitser. Lll & abc. *Journal of Number Theory*, 107(1):161–167, 2004.

[Elk00] Noam D Elkies. Rational points near curves and small nonzero| x 3- y 2| via lattice reduction. In *International Algorithmic Number Theory Symposium*, pages 33–63. Springer, 2000.

[Ing42] AE Ingham. On two conjectures in the theory of numbers. *American Journal of Mathematics*, 64(1):313–319, 1942.

[KtR06] Tadej Kotnik and Herman te Riele. *The Mertens Conjecture Revisited*, pages 156–167. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[LLL82] Arjen Klaas Lenstra, Hendrik Willem Lenstra, and László Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[NV09] Phong Q. Nguyen and Brigitte Valle. *The LLL Algorithm: Survey and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[OtR85] A. M. Odlyzko and H. J. J. te Riele. Disproof of the mertens conjecture. *J. REINE ANGEW. MATH*, 357:357, 1985.

[Pin87] J Pintz. An effective disproof of the mertens conjecture. *Astérisque*, 147(148):325–333, 1987.