ON COUNTEREXAMPLES TO THE MERTENS CONJECTURE

SEUNGKI KIM AND PHONG Q. NGUYEN

Abstract. We use state-of-art lattice algorithms to improve the upper bound on the lowest counterexample to the Mertens conjecture to $\approx \exp(1.96 \times 10^{19})$, which is significantly below the conjectured value of $\approx \exp(5.15 \times 10^{23})$ by Kotnik and van de Lune [KvdL04].

1. Introduction

The Mertens conjecture [M97], dating back to 1897, is a statement about the growth rate of the Mertens function

$$M(x) := \sum_{1 \leq n \leq x} \mu(n),$$

where $\mu(n)$ is the Möbius function

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is squarefree, and has } k \text{ distinct prime factors,} \\ 0 & \text{if } n \text{ is not squarefree.} \end{cases}$$

The size of $M(x)$ is of interest in number theory, since it is closely related to the size of the real parts of the zeroes of the Riemann zeta function $\zeta(s)$. For example, a short argument (See e.g. [OtR85, Sec. 2]) shows that if $M(x) = O(x^\theta)$, then the Riemann zeta function $\zeta(s)$ has no zeroes on the half-plane $\operatorname{Re} s > \theta$. For $\theta = 1/2 + \varepsilon$ for arbitrarily small $\varepsilon > 0$, the latter statement is the famous Riemann hypothesis. The Mertens conjecture is a much bolder claim that

$$|M(x)| < x^{1/2} \text{ for all } x > 1. \tag{1.1}$$

It took nearly a century for this conjecture to be disproved, by Odlyzko and te Riele [OtR85] in 1985. Their argument consisted of certain insights from classical analytic number theory, and, perhaps surprisingly, the use of a lattice reduction algorithm. No alternative (dis)proof that does not rely on lattice reduction is known to this day.

A natural follow-up is to ask about the size of the smallest counterexample $\mathfrak{x}$ to the Mertens conjecture (1.1). This is also related to the estimation of the growth rate of $M(x)$, for which several different conjectures exist. For example, the experimental work by Kotnik and van de Lune [KvdL04] suggests that

$$|M(x)x^{-1/2}| \approx 1/2 \cdot \sqrt{\log\log\log x} \tag{1.2}$$

along the local extrema of $M(x)/x^{1/2}$, from which they derive the conjecture that

$$\mathfrak{x} \approx \exp(5.15 \times 10^{23}), \tag{1.3}$$

since $1/2 \cdot \sqrt{\log\log(5.15 \times 10^{23})} \approx 1$.

It is also possible to give rigorous upper bounds on $\mathfrak{x}$, thanks to Pintz [P87]. The original data of Odlyzko and te Riele [OtR85], under the theorem of Pintz [P87], translates to the statement that $\mathfrak{x} < \exp(3.21 \times 10^{64})$. The later more extensive set of experiments by

Submitted to Algor. Num. Th. Symp.

Kotnik and te Riele [KtR06] led to the improvement $\mathfrak{x} < \exp(1.59 \times 10^{40})$. Saouter and te Riele [StR14] refined the estimates given in [P87], and also ran more experiments of the same kind, which resulted in $\mathfrak{x} < \exp(1.004 \times 10^{33})$.

[OtR85] states that their achievement was possible thanks to a then-breakthrough in lattice reduction, the LLL algorithm due to Lenstra, Lenstra, and Lovász [LLL82]. [KtR06] and [StR14] also used LLL, and so did the relatively recent works such as Hurst [H18, Theorem 6.1]. However, over the last decade, there has been a huge amount of progress in lattice reduction techniques — largely motivated by the emergence of post-quantum cryptography based on computationally hard lattice problems. Noticing this, one would naturally be inclined to apply them to the context of the Mertens conjecture. Kim and Rozmarynowycz [KR23] was the first to point this out and to partially implement this idea, by simply replacing LLL with the BKZ algorithm ([SE94], [CN11]). As a result, they obtained a further improvement that $\mathfrak{x} < \exp(1.017 \times 10^{29})$. [KR23] was, however, still far from taking the full advantage of the recent advances in lattice problems.

In this work, we employ state-of-art lattice point enumeration techniques [LN13, GNR10] in search of a tighter upper bound on the lowest counterexample to the Mertens conjecture (1.1). As explained in Section 3 below, this is a much more natural and efficient strategy than running lattice reduction hundreds to thousands of times, which has been the method chosen by all the previous works mentioned above. We also give a careful consideration to the family of the lattices under question, since such an understanding is important for both predicting the outcome and improving the performance of the algorithm. As a result, we improve the choice of the lattice made by [OtR85] in a couple of ways, for the first time in the past 38 years. Moreover, we notice that our lattices are extremely orthogonal, with one unusually short vector; we adapt the enumeration algorithm accordingly to exploit this special shape, effectively speeding up our search by a factor of a few millions.

Thereby we were able to find a number of data points that beat the previous record and even the conjecture (1.3) by several orders of magnitude. The best we found, which took us barely half a day on a single core, implies that

$$\mathfrak{x} < \exp(1.96 \times 10^{19}), \tag{1.4}$$

which is below (1.3) by a factor of ≈ 26276 in the exponent. This would substantially impact the credibility of (1.2). A more extensive application of the methods of the present paper may help search for an alternative estimate on the growth rate of $M(x)$ — see Section 4.2 below.

There exist also numerous other applications of the computational aspect of lattices to number theory than the Mertens conjecture, see e.g. Simon [S10]. We hope that our work helps inform the community of the recent advances in lattice computations, and encourages revisiting some of the old problems with the new arsenal.

Organization. In Section 2, we briefly review the method of Odlyzko-te Riele [OtR85] and other works in the literature, in order to help the reader understand and put our work into perspective. In Section 3, we describe our experiment in detail. We conclude the paper in Section 4 with a discussion on further research directions to which the methods introduced herein may be applied.

Reproducibility. The present work involves significant data. Data files and/or source codes allowing to reproduce the data are available on `https://zenodo.org/records/10775723`.

Acknowledgments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885394). In addition, S.K. was supported by NSF grant CNS-2034176. We thank the anonymous referees for helpful comments and suggestions.

2. A review of the previous works

2.1. **The original argument of** [OtR85]**.** Resorting to proof by contradiction, let us assume the truth of the Mertens conjecture. As explained in [OtR85, Sec. 2], this implies in particular that all nontrivial zeroes ρ of $\zeta(s)$ satisfy $\operatorname{Re}\rho = 1/2$ (the Riemann hypothesis), and that they are all simple zeroes. For each such ρ with $\operatorname{Im}\rho > 0$, let us write $\gamma := \operatorname{Im}\rho$, $\alpha := |\rho\zeta'(\rho)|^{-1}$, $\psi := \arg(\rho\zeta'(\rho))$; conversely, whenever we write γ, α, or ψ, we are referring to the corresponding zero ρ of $\zeta(s)$. Then, for a certain increasing sequence $\{T_n\}$ with $n \le T_n \le n+1$, it holds that (see e.g. [KtR06, (3) and (4)] or [KvdL04, Sec. 2 and 3])

$$q(x) := \frac{M(x)}{x^{1/2}} = 2\lim_{n\to\infty} \sum_{0<\gamma<T_n} \alpha\cos(\gamma y - \psi) + O(x^{-1/2}), \tag{2.1}$$

where we write $y := \log x$ for short. Recall that the Mertens conjecture states that $|q(x)| < 1$ for $x > 1$.

(2.1) suggests one possible strategy for disproving (1.1): for a large $N > 0$, find y such that the sum

$$q_N(x) := 2\sum_{0<\gamma<N} \alpha\cos(\gamma y - \psi) \tag{2.2}$$

is large. This can be interpreted as a problem in simultaneous Diophantine approximation, as follows. Let us denote by $|a|_{2\pi}$ the representative of $a \pmod{2\pi}$ in $(-\pi, \pi]$. If we can find y such that all $|\gamma y - \psi|_{2\pi}$ are small, then we can expect that

$$\begin{aligned} 2\sum_{0<\gamma<N} \alpha\cos(\gamma y - \psi) &\approx \sum_{0<\gamma<N} \alpha(2 - |\gamma y - \psi|_{2\pi}^2) \\ &\approx \sum_{0<\gamma<N} 2\alpha. \end{aligned}$$

Now it is known that the last sum diverges as $N \to \infty$. Hence, for a sufficiently large N, if we can indeed find such y and thereby not lose too much in the estimates above, we can hope to be able to demonstrate that (2.1) is greater than 1, and — at least in principle — even arbitrarily large. This is precisely the approach taken by [OtR85].

(Or alternatively, by finding y such that all $|\gamma y - \psi - \pi|_{2\pi}$ are small, we can try to show $q(x)$ can be large in the negative direction.)

[OtR85] converts the problem of the (weighted inhomogeneous simultaneous) Diophantine approximation to the problem of reducing a certain lattice. They consider the lattice in $\mathbb{R}^{N+2}$, say L_0, consisting of the integer linear combinations of the rows of

$$\begin{pmatrix} \lfloor 2\pi\sqrt{\alpha_1}2^\nu \rfloor & 0 & \cdots & 0 & 0 & 0 \\ 0 & \lfloor 2\pi\sqrt{\alpha_2}2^\nu \rfloor & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lfloor 2\pi\sqrt{\alpha_N}2^\nu \rfloor & 0 & 0 \\ -\lfloor \sqrt{\alpha_1}\psi_1 2^\nu \rfloor & -\lfloor \sqrt{\alpha_2}\psi_2 2^\nu \rfloor & \cdots & -\lfloor \sqrt{\alpha_N}\psi_N 2^\nu \rfloor & 2^\nu N^4 & 0 \\ \lfloor \sqrt{\alpha_1}\gamma_1 2^{\nu-10} \rfloor & \lfloor \sqrt{\alpha_2}\gamma_2 2^{\nu-10} \rfloor & \cdots & \lfloor \sqrt{\alpha_N}\gamma_N 2^{\nu-10} \rfloor & 0 & 1 \end{pmatrix}, \tag{2.3}$$

where α_i's are the α's ordered in descending order so that $\alpha_1 > \alpha_2 > \ldots$, and the ψ_i, etc., are those corresponding to the zero ρ_i of $\zeta(s)$ associated to α_i; and the role of 2^ν is to approximate the entries of (2.3) to ν most significant base 2 digits. Given a basis of a lattice such as this one, lattice reduction algorithms such as LLL compute another basis of the same lattice consisting of vectors of reasonably short length (depending on the strength of the algorithm), called a *reduced basis* of that lattice.

Let us consider a reduced basis of L_0. We claim that it must contain a vector of the form

$$\begin{aligned}(p_1\lfloor 2\pi\sqrt{\alpha_1}2^\nu\rfloor + z\lfloor\sqrt{\alpha_1}\gamma_1 2^{\nu-10}\rfloor - \lfloor\sqrt{\alpha_1}\psi_1 2^\nu\rfloor, \ldots \\ \ldots, p_N\lfloor 2\pi\sqrt{\alpha_N}2^\nu\rfloor + z\lfloor\sqrt{\alpha_N}\gamma_N 2^{\nu-10}\rfloor - \lfloor\sqrt{\alpha_N}\psi_N 2^\nu\rfloor, \pm 2^\nu N^4, z)\end{aligned} \tag{2.4}$$

for some integers $p_1, \ldots, p_N$ and z. There certainly must be a vector whose penultimate entry is nonempty, since a reduced basis is, in particular, a basis. But then, since $2^\nu N^4$ is huge compared to the rest of the entries of (2.3), one can argue from the performance guarantee of LLL [LLL82, Prop. 1.12] that the penultimate entry must be as small as possible.

Now set $y = \pm 2^{10} z$, the sign being that of $2^\nu N^4$ in (2.4). Then it can be seen, from the "size-reducedness" property of a reduced basis [LLL82, (1.4)], that each of the first N entries of (2.4) are approximately equal to $2^\nu\sqrt{\alpha}\cdot|\gamma_i y - \psi|_{2\pi}$.

One may still be (rightfully) curious as to how this would result in a good Diophantine approximation y. As [OtR85, Sec. 3] writes, some part of it is a miracle: LLL is famously known to perform much better than the theoretical guarantee given by [LLL82, Prop. 1.12]. However, as pointed out in [KR23], much of it can be adequately explained in the language of lattice problems. What the reduction of (2.3) really achieves is the resolution of the *approximate closest vector problem* (aCVP) of finding a point on the lattice in $\mathbb{R}^{N+1}$ generated by the rows of

$$\begin{pmatrix} \lfloor 2\pi\sqrt{\alpha_1}2^\nu\rfloor & 0 & \cdots & 0 & 0 \\ 0 & \lfloor 2\pi\sqrt{\alpha_2}2^\nu\rfloor & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lfloor 2\pi\sqrt{\alpha_N}2^\nu\rfloor & 0 \\ \lfloor\sqrt{\alpha_1}\gamma_1 2^{\nu-10}\rfloor & \lfloor\sqrt{\alpha_2}\gamma_2 2^{\nu-10}\rfloor & \cdots & \lfloor\sqrt{\alpha_N}\gamma_N 2^{\nu-10}\rfloor & 1 \end{pmatrix} \tag{2.5}$$

that is reasonably close to the "target" vector

$$(-\lfloor\sqrt{\alpha_1}\psi_1 2^\nu\rfloor, -\lfloor\sqrt{\alpha_2}\psi_2 2^\nu\rfloor, \cdots, -\lfloor\sqrt{\alpha_N}\psi_N 2^\nu\rfloor, 0) \in \mathbb{R}^{N+1}$$

via *Babai's nearest plane algorithm* [B86], one of the standard approaches to aCVP to this day, and implicitly used as a subroutine inside LLL itself. This interpretation allows one to predict the outcome heuristically, that matches the actual output rather well — see [KR23, Sec. 2] for details.

2.2. **Results on the smallest counterexample.** As in the introduction, let us continue to denote by $\mathfrak{x}$ the smallest real number for which (1.1) does not hold. Giving a rigorous upper bound on $\mathfrak{x}$ became possible thanks to the following result of Pintz [P87].

Theorem 2.1 (Pintz [P87])**.** *Let*

$$h_P(y) := 2\sum_{\gamma<14000} \alpha\exp(-1.5\cdot 10^{-6}\gamma^2)\cos(\gamma y - \psi). \tag{2.6}$$

If there exists $y \in [e^7, e^{50000}]$ *with* $|h_P(y)| > 1 + e^{-40}$*, then* $\mathfrak{x} < \exp(y + \sqrt{y})$.

For the value of y found in [OtR85], Theorem 2.1 implies that $\mathfrak{x} < \exp(3.21 \times 10^{64})$. Later, [KtR06] repeated the experiment of [OtR85], that we described in the previous section, over a broader range of parameters N and ν, and found a lower working value of y, which corresponds to the improved bound $\mathfrak{x} < \exp(1.59 \times 10^{40})$.

[StR14] improved Theorem 2.1 by refining Pintz's estimates on certain contour integrals involving the zeta function. Consequently, they were essentially able to replace h_P in Theorem 2.1 by

$$h_{StR}(y) := 2 \sum_{\gamma<74000} \alpha \exp(-3 \cdot 10^{-9}\gamma^2) \cos(\gamma y - \psi). \tag{2.7}$$

As can be seen by comparing (2.6) and (2.7), $|h_{StR}|$ tends to be somewhat larger than $|h_P|$, so that some of the "near-misses," i.e. those y for which $|h_P(y)| < 1$ but very close, may satisfy $|h_{StR}(y)| > 1$ and become valid "hits." For practical values of y, $|h_{StR}(y)| > 1 + 6 \cdot 10^{-8}$ implies that $\mathfrak{x} < \exp(y + \sqrt{y})$. With this, and some extra search for the candidate values, [StR14] attained $\mathfrak{x} < \exp(1.004 \times 10^{33})$.

Recently, [KR23] essentially repeated the experiment of [KtR06], except for the following few tweaks:

(i) They replaced LLL with the more powerful BKZ, which leads to a much better solution to the aCVP problem.
(ii) They ran tens of thousands of trials, and for each trial, they perturbed the basis (2.3) hoping that the randomization effect would help find a lower value of y for which Theorem 2.1 is applicable.

These were possible — with only the computational power of a personal laptop — thanks to the substantial advances in lattice reduction over the last decade. They succeeded in finding the value

$$y = 1017256208\ 7569945816\ 8018857216.806640625 \text{ with } h_P(y) = 1.0034372\ldots,$$

which implies $\mathfrak{x} < \exp(1.017 \times 10^{29})$, the best record to this date.

3. Our experiment

3.1. **Lattice point enumeration.** The method of lattice reduction for finding y satisfying the conditions of Theorem 2.1 has a few limitations. To maximize (2.6) or (2.7), it makes sense to account for as many summands as possible, that is, to take N as large as possible. However, high-quality lattice reduction becomes extremely time-consuming for $N \geq 100$, so the trial-and-error strategy of [KR23] would take too much computational resource: non-trivial lattice tasks such as finding a closest lattice point typically run in time exponential in N. Furthermore, maximizing (2.6) or (2.7) is not exactly a lattice problem: it is possible that the optimal solution does not arise from a particularly close lattice point. The terms on the right-hand side of (2.6) or (2.7) that are left out in the construction (2.3) turn out to be large enough to affect the outcome either in or against our favor, and they are essentially the matter of a coin toss.

The technique of lattice point enumeration provides a natural solution to this dilemma. When the lattice dimension is moderate, it is possible to enumerate all the lattice points within any small ball, or if the ball is larger, to enumerate many lattice points inside the ball. To do this, a standard algorithm is enumeration with pruning [LN13, GNR10], building upon work by Pohst [P81], Kannan [K83], and Schnorr-Euchner [SE94]. Gama, Nguyen and Regev [GNR10] showed how to speed up rigorously the Schnorr-Euchner [SE94] enumeration of lattice points within a zero-centered ball, which consists

of a depth-first search (DFS) of a carefully constructed tree. This technique was later adapted to any ball by Liu and Nguyen [LN13].

Let L be a full-rank lattice in $\mathbb{R}^n$. Given a target $\mathbf{t} \in \mathbb{Q}^n$, a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of L and a radius $R > 0$, enumeration [P81, K83] outputs $L \cap S$ where $S = \mathrm{Ball}_n(\mathbf{t}, R)$:. It performs a recursive search using projections, to reduce the dimension of the lattice: if $\|\mathbf{v}\| \leq R$, then $\|\pi_k(\mathbf{v})\| \leq R$ for all $1 \leq k \leq n$, where π_k denotes the orthogonal projection on $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{k-1})^\perp$. One can easily enumerate $\pi_n(L) \cap S$. And if one enumerates $\pi_{k+1}(L) \cap S$ for some $k \geq 1$, one derives $\pi_k(L) \cap S$ by enumerating the intersection of a one-dimensional lattice with a suitable ball, for each point in $\pi_{k+1}(L) \cap S$. Concretely, it can be viewed as a depth-first search of the following enumeration tree $\mathcal{T}$: the nodes at depth $n + 1 - k$ are the points of $\pi_k(L) \cap S$. The running-time of enumeration depends on R and B, but is typically super-exponential in n, even if $L \cap S$ is small.

Pruned enumeration [GNR10, SE94] uses a pruning set $P \subseteq \mathbb{R}^n$, and outputs $L \cap (\mathbf{t}+P)$. The advantage is that for suitable choices of P, enumerating $L \cap (\mathbf{t} + P)$ is much cheaper than $L \cap S$, yet under mild heuristics, $L \cap (\mathbf{t} + P)$ is expected to cover most of $L \cap S$. The pruning set P should be viewed as a random variable: it depends on the choice of basis B. In [GNR10], P is defined by a function $f : \{1, \ldots, n\} \to [0, 1]$, a radius $R > 0$ and a lattice basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ as follows:

$$P_f(B, R) = \{\mathbf{x} \in \mathbb{R}^n \text{ s.t. } \|\pi_{n+1-i}(\mathbf{x})\| \leq f(i)R \text{ for all } 1 \leq i \leq n\}, \tag{3.1}$$

This form of pruning is known as cylinder pruning, because $P_f(B, R)$ is an intersection of cylinders: each inequality $\|\pi_{n+1-i}(\mathbf{x})\| \leq f(i)R$ defines a cylinder. And [GNR10] provides an algorithm which, given as input (B, R, f), decides if $L \cap P_f(B, R)$ is non-empty, and if so, outputs one element of $L \cap P_f(B, R)$. It is easy to modify this algorithm to tackle more needs: for instance, [LN13] extended it to $L \cap (\mathbf{t} + P_f(B, R))$where $\mathbf{t}$ is an additional input. In our case, we are interested in enumerating the whole $L \cap (\mathbf{t} + P_f(B, R))$: specifically, we implemented Alg. 1, which is a slight variant of [LN13].

Gama *et al.* [GNR10] showed how to efficiently compute tight lower and upper bounds for $\mathrm{vol}(P_f(B, R))$ and also estimated of the cost of enumerating $L \cap S \cap P_f(B, R)$, using the Gaussian heuristic (that for a lattice L and a set S, $|L \cap S| \approx \mathrm{vol}\, S/\det L$) on projected lattices $\pi_i(L)$: these estimates are usually accurate in practice, and they can also be used in the CVP case [LN13].

In our experiments, we used a function f close to linear. It is well-known that if $\mathbf{x}$ denotes the projection of a random unit vector of $\mathbb{R}^n$ onto an i-dimensional subspace, then $\|\mathbf{x}\|^2$ has distribution $\mathrm{Beta}(i/2, (n-i)/2)$. Accordingly, we took $f^2(i) = \mu_i + 2\sigma_i$ where μ_i and σ_i are respectively the expectation and the standard deviation of the $\mathrm{Beta}(i/2, (n - i)/2)$ distribution. From the analysis of [GNR10], for this choice of f, $L \cap (\mathbf{t} + P_f(B, R))$ should cover most of $L \cap \mathrm{Ball}_n(\mathbf{t}, R)$.

3.2. **Choice of lattice.** With optimizing (2.6) or (2.7) in mind, for each nontrivial zero ρ of $\zeta(s)$ with $\mathrm{Im}\,\rho > 0$, we define $\alpha^* = \alpha \exp(-1.5 \cdot 10^{-6}\gamma^2)$ if we want to find large values of $|h_P|$, or $\alpha^* = \alpha \exp(-3 \cdot 10^{-9}\gamma^2)$ if we want to find large values of of $|h_{StR}|$. We define the corresponding γ and ψ as earlier. We index the ρ's and other variables accordingly so that $\alpha_1^* > \alpha_2^* > \ldots$.

Algorithm 1 Pruned Enumeration for BDD of unbalanced lattices (slight variant version of [LN13, GNR10])

Input: A reduced basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_m)$ of a lattice L, a target vector $\mathbf{t} = \sum_{i=1}^{m} t_i \mathbf{b}_i$, a bounding function $R_1^2 \leq \cdots \leq R_m^2$, the Gram-Schmidt matrix $\mu = (\langle \mathbf{b}_i, \mathbf{b}_j^* \rangle / \|\mathbf{b}_j^*\|^2)$ and the (squared) norms $\|\mathbf{b}_1^*\|^2, \ldots, \|\mathbf{b}_m^*\|^2$, where the $\mathbf{b}_j^*$'s are the Gram-Schmidt orthogonalization of the $\mathbf{b}_j$'s.

Output: The basis coefficients of all $\mathbf{v} \in L$ such that the projections of $\mathbf{v} - \mathbf{t}$ have norms less than the R_i's, *i.e.* $\|\pi_{m+1-k}(\mathbf{v} - \mathbf{t})\| \leq R_k$ for all $1 \leq k \leq m$, and there is no $x_1 \in \mathbb{Z}$ such that $\|\mathbf{v} + x\mathbf{b}_1 - \mathbf{t}\| < \|\mathbf{v} - \mathbf{t}\|$. The latter constraint avoids returning too many vectors, when $\mathbf{b}_1$ is much shorter than $\det(L)^{1/m}$.

$\sigma \leftarrow (0)_{(m+1)\times m}$; $r_0 = 0; r_1 = 1; \cdots; r_m = m; \rho_{m+1} = 0$
for $k = m$ **downto** 1
  for $i = m$ **downto** $k+1$ **do** $\sigma_{i,k} \leftarrow \sigma_{i+1,k} + (t_i - v_i)\mu_{i,k}$ **endfor**
  $c_k \leftarrow t_k + \sigma_{k+1,k}$ // $c_k \leftarrow t_k + \sum_{i=k+1}^{m}(t_i - v_i)\mu_{i,k}$, centers
  $v_k \leftarrow \lfloor c_k \rceil$ // *current combination*;
  $w_k = 1$ // *jumps*;
  $\rho_k = \rho_{k+1} + (c_k - v_k)^2 \cdot \|\mathbf{b}_k^*\|^2$
endfor
$k = 1$;
while true **do**
  $\rho_k = \rho_{k+1} + (c_k - v_k)^2 \cdot \|\mathbf{b}_k^*\|^2$ // *compute norm squared of current node*
  if $\rho_k \leq R_{m+1-k}^2$ (*we are below the bound*) **then**
    if $k = 1$ **then**
      return $(v_1, \ldots, v_m)$; (*solution found; so we're going up the tree*)
      $k \leftarrow k + 1$ // *going up the tree*
      if $k = m + 1$ **then**
        return $\emptyset$ (*there is no solution)*
      end if
      $r_{k-1} \leftarrow k$ // *since v_k is about to change, indicate that (i, j) for $j < k$ and $i \leq k$ are not synchronized*
      // *update v_k*
      if $v_k > c_k$ **then** $v_k \leftarrow v_k - w_k$ **else** $v_k \leftarrow v_k + w_k$
      $w_k \leftarrow w_k + 1$
    else
      $k \leftarrow k - 1$ // *going down the tree*
      $r_{k-1} \leftarrow \max(r_{k-1}, r_k)$ // *to maintain the invariant for $j < k$*
      for $i = r_k$ **downto** $k+1$ **do** $\sigma_{i,k} \leftarrow \sigma_{i+1,k} + (t_i - v_i)\mu_{i,k}$ **endfor**
      $c_k \leftarrow t_k + \sigma_{k+1,k}$ // $c_k \leftarrow t_k + \sum_{i=k+1}^{m}(t_i - v_i)\mu_{i,k}$
      $v_k \leftarrow \lfloor c_k \rceil$; $w_k = 1$
    end if
  else
    $k \leftarrow k + 1$ // *going up the tree*
    if $k = m + 1$ **then**
      return $\emptyset$ (*there is no solution)*
    end if
    $r_{k-1} \leftarrow k$ // *since v_k is about to change, indicate that (i, j) for $j < k$ and $i \leq k$ are not synchronized*
    // *update v_k*
    if $v_k > c_k$ **then** $v_k \leftarrow v_k - w_k$ **else** $v_k \leftarrow v_k + w_k$
    $w_k \leftarrow w_k + 1$
  end if
end while

For parameters ν, ν_y, ν_t, we consider the integer lattice L spanned by the rows of the $(N+1) \times (N+1)$ matrix

$$\begin{pmatrix} \lfloor \sqrt{\alpha_1^*}\gamma_1 2^{\nu_y} \rfloor & \lfloor \sqrt{\alpha_2^*}\gamma_2 2^{\nu_y} \rfloor & \dots & \lfloor \sqrt{\alpha_N^*}\gamma_N 2^{\nu_y} \rfloor & 2^{\nu_t} \\ \lfloor \sqrt{\alpha_1^*} 2\pi 2^{\nu} \rfloor & 0 & \dots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & & 0 & 0 \\ 0 & \dots. & 0 & \lfloor \sqrt{\alpha_N^*} 2\pi 2^{\nu} \rfloor & 0 \end{pmatrix}. \tag{3.2}$$

We used α^* instead of α as in (2.5), since we are trying to optimize h_P (2.6) or h_{StR} (2.7), not q_N (2.2) as in [OtR85]. The new parameters ν_y and ν_t will be helpful in further analyzing and improving the outcome of the experiment, as explained in the next section. Our "target vector" is

$$\mathbf{t} = \left(\lfloor \sqrt{\alpha_1^*}\psi_1 2^{\nu} \rfloor, \lfloor \sqrt{\alpha_2^*}\psi_2 2^{\nu} \rfloor, \cdots, \lfloor \sqrt{\alpha_N^*}\psi_N 2^{\nu} \rfloor, 0 \right)$$

for positive values of h_P (and

$$\mathbf{t}' = \left(\lfloor \sqrt{\alpha_1^*}(\psi_1 + \pi) 2^{\nu} \rfloor, \lfloor \sqrt{\alpha_2^*}(\psi_2 + \pi)) 2^{\nu} \rfloor, \cdots, \lfloor \sqrt{\alpha_N^*}(\psi_N + \pi) 2^{\nu} \rfloor, 0 \right)$$

for negative values of h_P). For a parameter $\gamma > 0$, we apply lattice enumeration to find all points $\mathbf{u} \in L$, or equivalently all $x \in \mathbb{Z}$, such that

$$\begin{aligned} \mathbf{u} - \mathbf{t} = \Big(& \left| x \lfloor \sqrt{\alpha_1^*}\gamma_1 2^{\nu_y} \rfloor - \lfloor 2^{\nu} \sqrt{\alpha_1^*}\psi_1 \rfloor \right|_{\lfloor 2^{\nu} 2\pi \sqrt{\alpha_1^*} \rfloor}, \cdots, \\ & \cdots, \left| x \lfloor \sqrt{\alpha_N^*}\gamma_N 2^{\nu_y} \rfloor - \lfloor 2^{\nu} \sqrt{\alpha_N^*}\psi_N \rfloor \right|_{\lfloor 2^{\nu} 2\pi \sqrt{\alpha_N^*} \rfloor}, x 2^{\nu_t} \Big) \end{aligned} \tag{3.3}$$

(note the similarity with (2.4)) is shorter than

$$K := \gamma \sqrt{\frac{N+1}{2\pi e}} \det L^{\frac{1}{N+1}},$$

where $\gamma \geq 1$ is a parameter to be chosen later, and $\sqrt{\frac{N+1}{2\pi e}}$ is the approximate radius of the ball in $\mathbb{R}^{N+1}$ of unit volume. Hence

$$\sum_{i=1}^{N} \left| x \lfloor \sqrt{\alpha_i^*}\gamma_i 2^{\nu_y} \rfloor - \lfloor 2^{\nu} \sqrt{\alpha_i^*}\psi_i^* \rfloor \right|^2_{\lfloor 2^{\nu} 2\pi \sqrt{\alpha_i^*} \rfloor} < K^2, \tag{3.4}$$

and thus we would expect

$$\sum_{i=1}^{N} \alpha_i^* |\gamma_i^* y - \psi_i^*|_{2\pi}^2 < \frac{K^2}{2^{2\nu}}, \tag{3.5}$$

where $y = x 2^{\nu_y - \nu}$.

We note that the lattice defined by (3.2) is unbalanced when $\nu \gg \nu_y$. Indeed, in such a case, the i-th coefficient of the first row is much smaller than the i-th diagonal coefficient and therefore, the first row has norm much smaller than $\det(L)^{1/(N+1)}$ and is likely to be a shortest vector of L: Fig. 1 shows the typical profile of a reduced basis of L, which differs from a reduced basis of a random lattice (in the sense of the Haar measure on $\mathrm{PGL}(N+1, \mathbb{Z}) \backslash \mathrm{PGL}(N+1, \mathbb{R})$). Here, the first vector is much smaller than $\det(L)^{1/(N+1)}$, and the Gram-Schmidt norms of the reduced basis first increase, before eventually decreasing geometrically as in a typical reduced basis [CN11, GNR10]: this means reduced bases of L are significantly more reduced than a reduced basis of a random

lattice, which makes enumeration faster for the same dimension. This property must also be taken into account when enumerating lattice points inside a ball. Indeed, whenever we have found $\mathbf{u} \in L$ such that $\|\mathbf{t} - \mathbf{u}\| \leq R$ then there are likely many integers $m \in \mathbb{Z}$ such that $\|\mathbf{t} - \mathbf{u} - m\mathbf{b}_1\| \leq R$, where $\mathbf{b}_1$ is the top row of (3.2), because $\|\mathbf{b}_1\|$ is much smaller than R and $\|\mathbf{t} - \mathbf{u}\|$. So if we want to enumerate $L \cap \mathrm{Ball}_n(\mathbf{t}, R)$, it is better to only enumerate the points $\mathbf{u} \in L \cap \mathrm{Ball}_n(\mathbf{t}, R)$ such that there is no nonzero $m \in \mathbb{Z}$ such that $\|\mathbf{t} - \mathbf{u} - m\mathbf{b}_1\| < \|\mathbf{t} - \mathbf{u}\|$: this is done by Alg. 1, which is a slight variant of [LN13]. In other words, we are actually enumerating over the projection of L onto the hyperplane orthogonal to $\mathbf{b}_1$. This phenomenon is inherent to our construction: if $y \approx z$, then $h_P(y) \approx h_P(z)$, so there is no need to return all the lattice points corresponding to $z \approx y$, whenever we have found a good y.

3.3. **Constraints on the parameters.** The above discussion leaves the five parameters $N, \nu, \nu_y, \nu_t, \gamma$ to be determined. In principle, N is the bigger the better, and the only constraint is the amount of computational power available. γ controls the number of the candidate points that are to be enumerated, since $|L \cap \mathrm{Ball}_n(\mathbf{t}, K)| \approx \gamma^{N+1}$. In our experiments, we chose $N \in \{120, 130, 140\}$, and $\gamma \in [1, 1.28]$.

ν, ν_y, ν_t have a strong influence on the entry sizes of (3.3), which helps us choose their values to some extent. We expect $\|\mathbf{u} - \mathbf{t}\| \approx K$, since most of the mass of a high-dimensional ball lies away from its center. Therefore, each entry would be of size around

$$\frac{K}{\sqrt{N+1}} = \frac{\gamma}{\sqrt{2\pi e}} \det L^{\frac{1}{N+1}} = \frac{\gamma}{\sqrt{2\pi e}} \cdot 2^{\frac{\nu_t}{N+1}} 2^{\frac{N\nu}{N+1}} \prod_{i=1}^{N} (2\pi\sqrt{\alpha_i^*})^{\frac{1}{N+1}}.$$

For $N = 120$ for instance, we have

$$\frac{1}{\sqrt{2\pi e}} \prod_{i=1}^{N} (2\pi\sqrt{\alpha_i^*})^{\frac{1}{N+1}} \approx 2^{-3.6},$$

from which we can predict

$$\sqrt{\alpha_i^*}|\gamma_i y - \psi_i|_{2\pi} \in 2^{-3.6-\frac{\nu-\nu_t}{N+1}} \cdot [1, \gamma] \tag{3.6}$$

for each i. Since we wish this to be small, $\nu - \nu_t$ needs to be at least of comparable size to N. By a similar computation, we also obtain the heuristic

$$y \in 2^{-3.6+\nu_y-\nu_t-\frac{\nu-\nu_t}{N+1}} \cdot [1, \gamma]. \tag{3.7}$$

Of course, these statements must be taken with a grain of salt, since we are looking at a ball of a relatively small radius, and our lattice has a rather unusual shape, as remarked earlier. Still, they can and do serve as useful guides in practice: in our experiments, the values of y found hardly differed from (3.7) by more than a factor of 2^4, as can be checked from Tables 1 and 2 below.

In addition, for the expectation (3.5) made from (3.4) to be reasonable, it is necessary that ν_t be not too small. The reason is that the difference between $x\lfloor\sqrt{\alpha_i^*}\gamma_i 2^{\nu_y}\rfloor$ and $\lfloor x\sqrt{\alpha_i^*}\gamma_i 2^{\nu_y}\rfloor$ can scale with x: for instance, $100 \cdot \lfloor 1.99\rfloor = 100$, whereas $\lfloor 100 \cdot 1.99\rfloor = 199$. From (3.7), we find $x \approx 2^{-3.6+\nu-\nu_t-\frac{\nu-\nu_t}{N+1}}$, dividing which by 2^ν gives $2^{-3.6-\nu_t-\frac{\nu-\nu_t}{N+1}}$. Hence (3.5) may diverge from our expectation by as much as $N2^{-3.6-\nu_t-\frac{\nu-\nu_t}{N+1}}$, which may be nontrivial for small values of ν_t. The previously used lattice construction (2.5) was problematic in this respect, since it fixes $\nu_t = 0$.

3.4. **Implementation details.** We used three software libraries: Arb [J17] for guaranteed interval arithmetic to compute h_P and h_{StR}, fplll [FPLLL] for lattice basis reduction (implementations of LLL and BKZ), and NTL [NTL] with which we implemented our variant of the pruned enumeration algorithm [LN13].

For the values $\rho, \alpha, \gamma, \psi$ related to the zeroes of the Riemann zeta function up to height 14,000 we used those computed by Hurst [H18] using Mathematika with $\approx 10,000$ decimal digits of precision. This is sufficient to compute h_P. For h_{StR}, we needed heights up to 74,000: we used the arb library [J17] to compute the zeroes and the corresponding values with 300 decimal digits of precision, which took less than a core day.

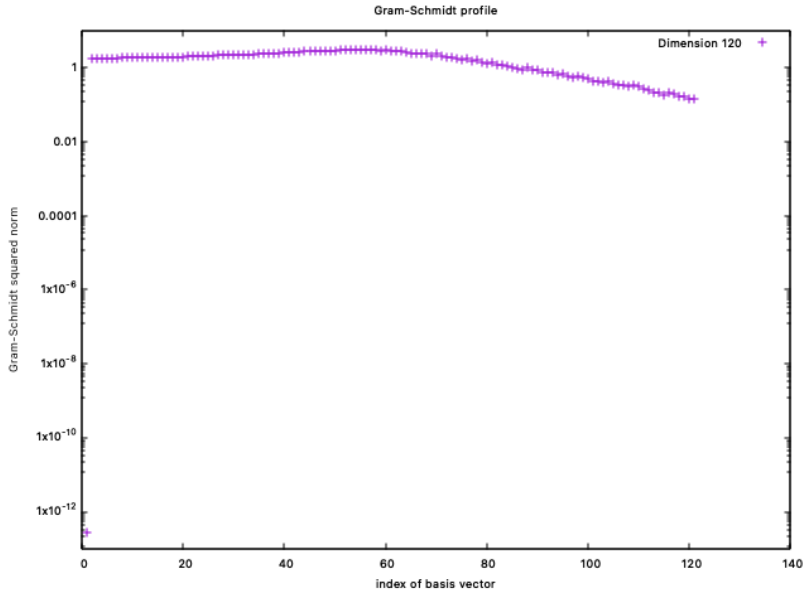

FIGURE 1. Profile of a 1-round BKZ-84 reduced basis of the Mertens lattice for $N = 120$, $\nu = 130$, $\nu_y = 100$, $\nu_t = 15$, and $\alpha^* = \alpha \exp(-1.5 \cdot 10^{-6}\gamma^2)$.

The best values of y which maximized or minimized h_P or h_{StR} which we found are provided in Tables 1 and 2: this means values for which $|h_P|$ or $|h_{StR}|$ are > 1 but also near-misses where it is slightly below 1. These tables were obtained using only 2 core days. We provide below more information for the best example for h_P and h_{StR}.

For h_P, our best y was found with the following process:

- We selected $N = 120$, $\nu = 130$, $\nu_y = 100$ and $\nu_t = 15$. The larger values $N = 130$ and $N = 140$ did not give better candidates.
- We computed a progressive 1-round BKZ-83 reduced basis of the lattice (3.2) using the fplll software [FPLLL]: this means that we ran the LLL algorithm, then one tour of BKZ-20 [SE94], then one tour of BKZ-21, and so on, until one tour of BKZ-83 [CN11]. This took about 2 hours on a single core.
- We ran pruned enumeration (Alg. 1) with scaling factor radius $\gamma = 1.23$ with the target $\mathbf{t}'$: the output was 37,937 lattice points. This took about 2 days on a single core. 37,937 is much less than $1.23^{121} \approx 7.6 \times 10^{10}$ suggested by the Gaussian heuristic: however, for each of these 37,937 points $\mathbf{u}$ there were about 2 million

points in $\mathbf{u} + \mathbb{Z}\mathbf{b}_1$ also in the ball, which means that the total number of points had order of magnitude 7.6×10^{10}.

- After trying all these 37,937 lattice points, we found that $h_P(y) \approx -1.012$ for

$$y \approx 23160\ 4645903103\ 2843375257.362502 \approx 2.32 \times 10^{24}.$$

This took a few hours on a single core.

TABLE 1. Best values of y for the function h_P

y	$h_P(y)$	$y + \sqrt{y}$	N	γ	ν	ν_y	ν_t
821801872381554552551865.064536	0.991	8.218×10^{23}	120	1.25	130	100	17
121701923526954856451053 4.246242	-0.993	1.217×10^{24}	120	1.23	130	100	15
231604645903103284337525 7.362502	-1.012	2.316×10^{24}	120	1.23	130	100	15
133551238704654603000494 97.114138	1.0019	1.336×10^{25}	120	1.28	120	100	12
150706585562099215360655 25.478881	1.0004	1.507×10^{25}	120	1.28	120	100	12

For h_{StR}, our best y was found with the following process:

- We selected $N = 140$, $\nu = 130$, $\nu_y = 100$ and $\nu_t = 30$. We also found the same candidate using $N = 120$ and different parameters.
- We computed a progressive 1-round BKZ-88 reduced basis of the lattice (3.2) using the fplll software [FPLLL]: this means that we ran the LLL algorithm, then one tour of BKZ-20 [SE94], then one tour of BKZ-21, and so on, until one tour of BKZ-88 [CN11]. This took a few hours on a single core.
- We ran pruned enumeration (Alg. 1) with scaling factor radius $\gamma = 1.19$ with the target $\mathbf{t}'$.
- Within a few hours, we obtained 17,406 lattice points, and one of them yielded $h_{StR}(y) \approx -1.007$ for

$$y \approx 1957187885\ 0562201959.215107 \approx 1.96 \times 10^{19}.$$

TABLE 2. Best values of y for the function h_{StR}

y	$h_{StR}(y)$	$y + \sqrt{y}$	N	γ	ν	ν_y	ν_t
8895437864289868028.044074	-0.974798	8.895×10^{18}	140	1.19	130	100	30
1385953971019784706 4.062257	-0.9949	1.386×10^{19}	140	1.19	130	100	30
1957187885056220195 9.215107	-1.007	1.957×10^{19}	140	1.19	130	100	30
4453369558095590279 0.827323	-0.9949	4.453×10^{19}	140	1.21	130	100	30
6417170555742045273 2.080835	-1.02	6.417×10^{19}	140	1.19	130	100	30
1338371855727955056 99.262652	-0.9998977	1.338×10^{20}	120	1.25	130	100	25
1555584886865681136 12.224656	1.025	1.555×10^{20}	140	1.19	130	100	30
1854156768208503755 95.577179	-0.997	1.854×10^{20}	120	1.25	130	100	25
1894712831494775402 26.654238	0.997	1.894×10^{20}	140	1.19	130	100	30
8340727722357591748 44.571429	1.0017	8.341×10^{20}	120	1.25	130	100	25
9554262640988679208 66.136509	-1.0002	9.554×10^{20}	120	1.25	130	100	25
8750553729179177422 74.218133	1.00057	8.751×10^{20}	120	1.25	130	100	25
1622648223749122520 779.415144	1.0079	1.623×10^{21}	120	1.25	130	100	25
1883922422293221654 459.096574	-1.00004	1.883×10^{21}	120	1.25	130	100	25

3.5. **Discussions.** If desired, there are a couple of ways to make small improvements on the values found in Tables 1 and 2 above. Something as simple as perturbing the values of y by a little could work: indeed, we were pointed out by an anonymous referee that

$$y = 1957187885\ 0562201959.21495,$$

a slight perturbation of our best y, yields $h_{StR} \approx -1.001$. Also, although so far we have been using the simplified bound $\mathfrak{x} < \exp(y + \sqrt{y})$ on the smallest counterexample, in fact we can take $\mathfrak{x} < \exp(y + 2\sqrt{ky})$, where $k = 1.5 \cdot 10^{-6}$ for h_P and $k = 3 \cdot 10^{-9}$ for h_{StR}, by Lemma E of [StR14]. Moreover, since

$$|M(x \pm n)| \geq |M(x)| - n$$

for any $n \in \mathbb{Z}_{>0}$,[1] if $|q(x)| = |M(x)x^{-1/2}| \geq 1 + \alpha$ for some $\alpha > 0$, then

$$|q(x \pm n)| \geq \left(1 \mp \frac{n}{x \pm n}\right)^{1/2} (1 + \alpha) - \frac{n}{(x \pm n)^{1/2}}.$$

From this, by a simple computation, it is possible to show that $|q(x \pm n)| \geq 1$ for $n < 0.99\alpha x^{1/2}$, say. This allows one to tighten the bound on $\mathfrak{x}$ a tiny bit further, to $\mathfrak{x} < \exp(y + 2\sqrt{ky}) - 0.99\alpha \exp(y/2 + \sqrt{ky})$; here, tracking the estimates in [StR14], one can set $\alpha = h_{StR}(y) - (1 + 6 \cdot 10^{-8})$.

Moving onto another topic, let us perform a small "sanity check" on our overall approach. When one transforms the problem of maximizing $|h_P|$ or $|h_{StR}|$ into a lattice problem, one makes several approximations, as discussed in the above sections. We would like to retrospectively check how sound these approximations are.

First, instead of considering the whole sum, we focus on the partial sum with the largest weights α_i^*: Figure 2 shows the correlations between h_{StR} and its value when restricted to the terms corresponding to the lattice. This confirms that high (resp. low) values of h_{StR} are indeed found among high (resp. low) values of the partial sum.

Second, when searching for high (resp. low) values of the partial sum, we enumerate a large number of lattice points close to some target. Figure 3 displays $h_{StR}(y)$ depending on the distance between the target and the lattice point: we see that the values maximizing $h_{StR}(y)$ are not necessarily those minimizing the distance, but $h_P(y) > 1$ looks unlikely to occur when the distance is bigger than some threshold. This means that we should enumerate a ball, but not a too large ball.

4. Further research topics

4.1. **More improvement on the lowest counterexample.** It seems plausible to us that, upon more extensive experiments and considerations, an even lower bound than our result (1.4) may be attainable. For instance, the choice of the parameters could be improved. The heuristics (3.6) and (3.7) suggest that perhaps one should increase ν considerably. However, one must also be conscious of the effect of such maneuver on the shape of the lattice (3.2). If ν is large, then ν_y must be of comparable size to ν, in order for the exponent of 2 in (3.7) to fall within a reasonable range. Hence, it may no longer be the case that $\nu \gg \nu_y$, which was a crucial condition for the efficiency of our approach. There may be a "sweet spot" choice of parameters balancing these and other factors to be taken into account, but it is currently unclear to us as to how to determine them, other than by trial and error.

[1] By carefully citing results on the distribution of the square-free numbers, this can be further improved to something close to $|M(x \pm n)| \geq |M(x)| - \pi^2 n/6$.

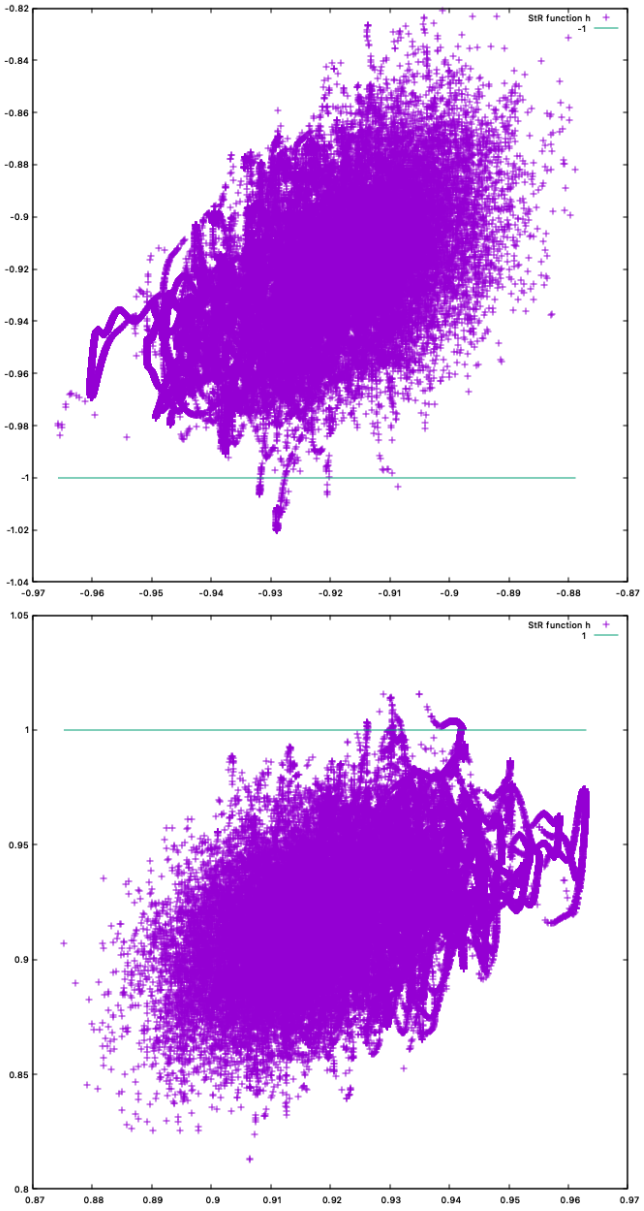

FIGURE 2. Correlations between $h_{StR}(y)$ and partial sums for $N = 120$.

A further improvement on the theoretical side may also lead to a better bound. Recently, Hathi [H23, Sec. 2.2] refined some of the crucial estimates given by [StR14], but it appears that he did not exploit it to the full extent.

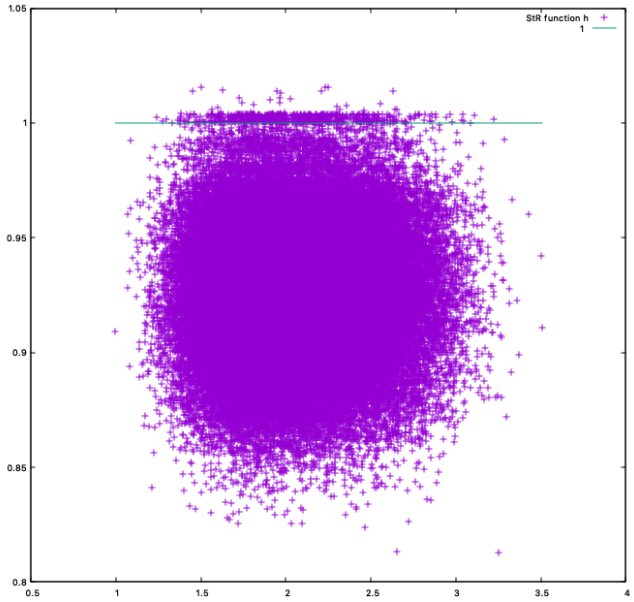

FIGURE 3. Correlations between $\|\mathbf{u} - \mathbf{t}\|^2$ and $h_{StR}(y)$ for $N = 120$, where $\mathbf{u}$ is the lattice point corresponding to y.

4.2. **On the growth order of $q(x)$.** There exist several different conjectures for the growth rate of $q(x) = M(x)x^{-1/2}$, some of which have been ruled out by concrete numerical works. The surviving ones so far all have the form

$$|q(x)| = \Omega((\log\log\log x)^{\theta}) \tag{4.1}$$

for some $\theta > 0$: $\theta = 1/2$ by Kotnik and van de Lune [KvdL04], $\theta = 1$ by Kaczorowski [K07], and $\theta = 5/4$ by Ng [N04], which is attributed to Gonek.

The method of this paper may be applied to shed light upon this matter as well. Applying our strategy to $q_N(x)$ instead of $h_P(x)$ or $h_{StR}(x)$, one could enumerate candidate values for maximizing $|q_N(x)|$ within an interval that can be more or less controlled by the heuristic (3.7). Collecting these data points over various intervals, and then extrapolating as in [KvdL04, Fig. 4] or [KtR06, Fig. 3], one would obtain a heuristic lower bound on θ. Recall that our motivation for introducing lattice-point enumeration (cf. Section 3.1) was that the tails of the series $h_P(x)$ or $h_{StR}(x)$ fluctuate wildly depending on x; enumeration helps us efficiently search for x for which the tails become especially large. Since the tail of $q_N(x)$ would fluctuate even more wildly, as can be seen by comparing (2.2) with (2.6) and (2.7), there is a chance that it may be even more effective under this scenario, and lead to some large value of θ.

Another quantity of interest is the maximum known size of $q(x)$. As of this moment, the record is held by Hurst [H18], who found

$$\limsup_{x\to\infty} q(x) \geq 1.826054, \ \liminf_{x\to\infty} q(x) \leq -1.837625$$

using the LLL algorithm on (2.3) with $\nu = 17000$ and $N = 800$. [H18] also provides an estimate on the time complexity needed to improve this bound; for example, it would take

11 months to find x demonstrating $|q(x)| \geq 2.00$. However, with the method of our work, it seems likely that a larger value can be attained in a much shorter time.

4.3. **Linear relations among the zeroes of $\zeta(s)$.** Best and Trudgian [BT15] presented a remarkable alternative proof to the Mertens conjecture, along the lines of the idea first suggested by Ingham [I42]. According to [I42], if the Mertens conjecture were true, there would exist infinitely many linear relations of the form

$$\sum_{i=1}^{N} c_i \gamma_i = 0, c_i \in \mathbb{Z} \text{ not all zero}, \tag{4.2}$$

among the imaginary values of the zeroes of $\zeta(s)$. Later efforts weakened the condition (4.2) to the existence of one such relation with bounded N and c_i's — see [BT15, Theorem 2], attributed to Anderson and Stark. Using the LLL algorithm differently from [OtR85], [BT15] was able to prove the nonexistence of such a relation, the associated lattice problem being to find a nearly orthogonal basis of a given lattice. The reduced basis they found implies in particular that (4.2) is false for $N = 500$ and $|c_i| \leq 4976$, and that $\limsup_{x\to\infty} |q(x)| \geq 1.6383$. It would be natural to apply the advanced toolkits on lattice problems available today to improve on these numbers.

It is widely believed that relations of the form (4.2) do not exist. This statement is called the linear independence conjecture, and has far-reaching consequences in number theory — see [N04] and the references therein for details.

References

[B86] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. Combinatorica, 6(1):1-13, 1986.

[BT15] D. Best and T. Trudgian. Linear relations of zeroes of the zeta-function. Math. Comp., 84:2433-2446

[CN11] Y. Chen, P. Nguyen. BKZ 2.0: better lattice security estimates. Advances in cryptology — ASIACRYPT 2011, 1-20, Lecture Notes in Comput. Sci., 7073, Springer, Heidelberg, 2011.

[FPLLL] The FPLLL development team. `https://github.com/fplll/fplll`.

[GNR10] N. Gama, P. Nguyen and O. Regev. Lattice Enumeration Using Extreme Pruning. Advances in cryptology — EUROCRYPT 2010, 257–278, Lecture Notes in Comput. Sci., 6110, Springer, Heidelberg, 2010.

[H23] S. Hathi. Some explicit results in analytic number theory. Ph.D. thesis, University of New South Wales Canberra, 2023.

[H18] G. Hurst. Computations of the Mertens function and improved bounds on the Mertens conjecture. Math. Comp., 87:1013-1028, 2018.

[I42] A. E. Ingham. On two conjectures in the theory of numbers. American Journal of Mathematics, 64(1):313-319, 1942.

[J17] F. Johansson. Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. IEEE Transactions on Computers, 66(8):1281-1292, 2017.

[K07] J. Kaczorowski. Results on the Möbius function. J. London Math. Soc. (2) 75 (2007) 509-521.

[K83] R. Kannan. Improved algorithms for integer programming and related lattice problems. In Proc. 15th ACM Symp. on Theory of Computing (STOC), pages 193-206, 1983.

[KR23] S. Kim, J. Rozmarynowycz. A new upper bound on the smallest counterexample to the Mertens conjecture. arXiv:2305.00345.

[KtR06] T. Kotnik and H. te Riele. The Mertens conjecture revisited. Proc. of ANTS 2006, pp. 156-167. Springer, Berlin, Heidelberg.

[KvdL04] T. Kotnik and J. van de Lune. On the order of the Mertens function. Exp. Math., 13:473-481, 2004.

[LLL82] A. Lenstra, H. Lenstra and L. Lovász. Factoring polynomials with rational coefficients. Math. Ann., 261(4):515-534, 1982.

[LN13] M. Liu and P. Nguyen. Solving BDD by Enumeration: An Update. Topics in Cryptology - CT-RSA 2013 - The Cryptographers' Track at the RSA Conference 2013, 293–309, Lecture Notes in Comput. Sci., 7779, Springer, Heidelberg, 2013.

[MO90] J. Mazo and A. Odlyzko. Lattice points in high-dimensional spheres. Monatsh. Math., 110:47-61, 1990.

[M97] F. Mertens. Über eine zahlentheoretische Funktion. Sitzungsberichte Akad. Wiss. Wien IIa, 106:761-830, 1897.

[N04] N. Ng, The distribution of the summatory function of the Möbius function, Proc. London Math. Soc. (3) 89 (2004), no. 2, 361-389.

[OtR85] A. Odlyzko and H. te Riele. Disproof of the Mertens conjecture. J. Reine Angew. Math., 357:138-160, 1985.

[P87] J. Pintz. An effective disproof of the Mertens conjecture. Astérisque, 147-148:325-333, 1987.

[P81] M. Pohst. On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications. SIGSAM Bull., 15(1):37–44, 1981.

[StR14] Y. Saouter and H. te Riele. Improved results on the Mertens conjecture. Math. Comp., 83:421-433, 2014.

[SE94] C. Schnorr and M. Euchner. Lattice basis reduction: improved practical algorithms and solving subset sum problems. Math. Programming 66 (1994), no. 2, Ser. A, 181-199.

[NTL] V. Shoup. NTL: A Library for doing Number Theory, `https://libntl.org/`.

[S10] D. Simon. Selected applications of LLL in number theory. Chapter 7 in The LLL Algorithm, Survey and applications, Information Security and Cryptography, Nguyen and Vallée (eds), Springer Verlag Berlin Heidelberg (2010), 265-282.

Department of Mathematical Sciences, University of Cincinnati, 4199 French Hall West, 2815 Commons Way, Cincinnati, OH 45221-0025, United States
Email address: seungki.math@gmail.com

Department of Computer Science, Ecole Normale Supérieure, 45 rue d'Ulm, 75005, Paris, France
Email address: phong.nguyen@inria.fr