On the Convex Closure of the Graph of Modular Inversions

Mizan R. Khan, Igor E. Shparlinski, and Christian L. Yankov

CONTENTS

In this paper we give upper and lower bounds as well as a heuristic estimate on the number of vertices of the convex closure of the set

$$G_n = \{(a, b) : a, b \in \mathbb{Z},\ ab \equiv 1 \,(\mathrm{mod}\, n)\,,\ 1 \le a, b \le n-1\}\,.$$

The heuristic is based on an asymptotic formula of Renyi and Sulanke. After describing two algorithms to determine the convex closure, we compare the numeric results with the heuristic estimate, and find that they do not agree—there are some interesting peculiarities, for which we provide a heuristic explanation. We then describe some numerical work on the convex closure of the graph of random quadratic and cubic polynomials over $\mathbb{Z}_n$. In this case the numeric results are in much closer agreement with the heuristic, which strongly suggests that the curve $xy = 1 \,(\mathrm{mod}\, n)$ is "atypical."

1. INTRODUCTION

Let G_n be the set

$$G_n = \{(a, b) : a, b \in \mathbb{Z},\ ab \equiv 1 \,(\mathrm{mod}\, n)\,,\ 1 \le a, b \le n-1\}\,,$$

whose cardinality is given by the Euler function $\varphi(n)$. If we scale by a factor of $1/n$, we get the set of points $n^{-1}G_n$, which is uniformly distributed in the unit square. More precisely, if $\Omega \subseteq [0, 1]^2$ has piecewise smooth boundary and $N(\Omega, n)$ is the cardinality of the intersection $\Omega \cap n^{-1}G_n$, then it is natural to expect, and in fact it can be proved using the bounds of Kloosterman sums, that

$$\left| |\Omega| - \frac{N(\Omega, n)}{\varphi(n)} \right| \to 0 \quad \text{as } n \to \infty, \qquad (1–1)$$

where $|\Omega|$ is the area of Ω. Figure 1, generated by MAPLE, illustrates this property.

Quantitative forms of (1–1) have been given in a number of works; see [Cobeli and Zaharescu 01, Granville et al. 05, Vajaitu and Zaharescu 02, Zhang 96, Zheng 96]

2000 AMS Subject Classification: Primary 11A07, 11H06, 11K38, 11N25

Keywords: Modular inversion, convex hull, distribution of divisors

1058-6458/2008 $ 0.50 per page

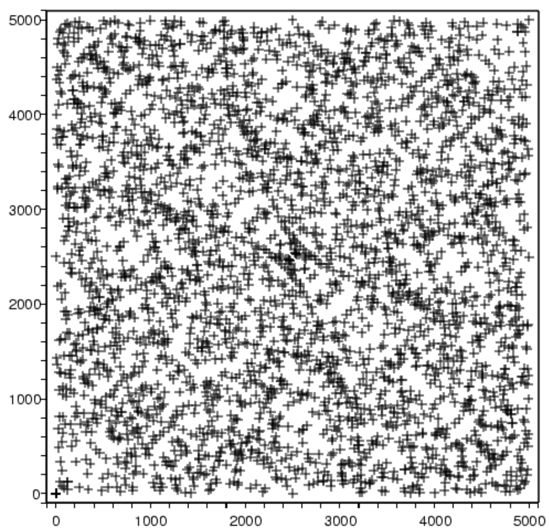

FIGURE 1. The graph G_{5001}.

and references therein. For example, it follows from more general results of [Granville et al. 05] that for primes p,

$$\left| |\Omega| - \frac{N(\Omega,p)}{p-1} \right| = O\left(p^{-1/4}\log p\right), \qquad (1\text{–}2)$$

where the implied constant depends only on Ω.

Here we continue to study some geometric properties of the set G_n and in particular, concentrate on the convex closure C_n of G_n. One of our questions of interest is the behavior of $v(n)$ and $V(N)$, where $v(n)$ denotes the number of vertices of C_n, and $V(N)$ denotes the average,

$$V(N) = \frac{1}{N-1}\sum_{n=2}^{N} v(n).$$

We demonstrate that the theoretic and algorithmic study of $v(n)$ has surprising links with various areas of number theory, such as bounds of exponential sums, distribution of divisors of "typical" integers, and integer factorization. On the other hand, we present heuristic estimates $h(n)$ and $H(N)$ for $v(n)$ and $V(N)$, respectively. These estimates arise by viewing G_n as a set of points that are randomly distributed and then using Satz 1 of [Rényi and Sulanke 63]. On comparing the estimates with our numeric results, we see that although the heuristic prediction $H(N)$ gives an adequate idea about the type of growth of $V(N)$, there is a deviation that behaves quite regularly and thus probably reflects certain hidden effects. We suggest an explanation. We also examine numerically some other interesting peculiarities in the behavior of $v(n)$, which lead us to several open questions.

Finally, we present some numerical evidence suggesting that the above effects do not arise for sets of points on other curves that behave more like truly random sets of points, which makes the study of G_n even more interesting.

The set G_n is a special case of the modular hyperbola

$$\mathcal{H}_{a,m} = \{(x,y) : xy \equiv a \,(\mathrm{mod}\; m)\}.$$

There are many interesting geometric questions that one can ask about such hyperbolas. The survey paper [Shparlinski 07] discusses some recent results on $\mathcal{H}_{a,m}$ and poses several open problems.

Throughout this paper, we use the order symbols O, o, $\ll$, $\gg$, $\asymp$, $\sim$ with their usual meanings in analytic number theory, where all implied constants are *absolute*. (We recall that the notations $A \ll B$, $B \gg A$ and $A = O(B)$ are equivalent, and $A \asymp B$ is equivalent to $A \ll B \ll A$).

2. SOME PRELIMINARY OBSERVATIONS

2.1 General Structure of C_n

We begin with a simple (but useful) remark on two lines of symmetry of G_n.

Proposition 2.1. *The points of G_n are symmetrically distributed about the lines $y = x$ and $x + y = n$.*

Therefore, if $(a,b) \in G_n$, then (b,a), its reflection in $y = x$, and $(n-b, n-a)$, its reflection in $x+y=n$, are elements of G_n. Consequently, (a,b) is a boundary point of C_n if and only if (b,a), $(n-b,n-a)$, and $(n-a,n-b)$ are boundary points of C_n.

Our next result shows that C_n is always a convex polygon with nonempty interior, except when $n = 2,3,4,6,8,12,24$.

Proposition 2.2. *The area of C_n is equal to* 0 *if and only if* $n = 2,3,4,6,8,12$*, or* 24.

Proof: Since $(1,1), (n-1,n-1) \in G_n$, the line $x = y$ has non-empty intersection with G_n. Consequently, the area of C_n equals 0 if and only if all the elements of G_n lie on the line $x = y$; that is, all of the elements of $\mathbb{Z}_n^*$ satisfy the congruence $x^2 \equiv 1 \,(\mathrm{mod}\; n)$. Elementary calculations show that this occurs only for $n = 2,3,4,6,8,12$ and 24. $\square$

From now on, we typically exclude the cases $n = 2, 3, 4, 6, 8, 12$, and 24.

2.2 Points in the Triangle $\mathcal{T}_n$

By Proposition 2.1, we need to know only the vertices of C_n that lie in the triangle $\mathcal{T}_n$ with vertices $(0,0)$, $(0,n)$, and $(n/2, n/2)$ to determine C_n. We denote such vertices of C_n by $(a_0, b_0), (a_1, b_1), \dots, (a_s, b_s) \in C_n \cap \mathcal{T}_n$, where $a_0 < a_1 < \cdots < a_s$.

Proposition 2.3. *We have the following:*

1. $(a_0, b_0) = (1, 1)$;
2. $a_i < b_i$ *for* $i = 1, \dots, s$;
3. $b_0 < b_1 < \cdots < b_s$.
4. $b_i - a_i < b_{i+1} - a_{i+1}$ *for* $i = 0, \dots, s-1$.

Proof: Assertions 1 and 2 are clear. Assertions 3 and 4 follow from the following observation: The line through (a_i, b_i) and its symmetric counterpart $(n - b_i, n - a_i)$ intersect the line $x + y = n$ at the point $((n - b_i + a_i)/2, (n + b_i - a_i)/2)$. Since $a_i < a_{i+1}$ and (a_{i+1}, b_{i+1}) is a vertex of C_n, it follows that (a_{i+1}, b_{i+1}) must actually lie inside the smaller triangle with vertices (a_i, b_i), $(a_i, n - a_i)$, and $((n - b_i + a_i)/2, (n + b_i - a_i)/2)$. □

2.3 On the Difference $b_s - a_s$

The inequalities in Proposition 2.3 may seem obvious, but they play a key role in our algorithms to compute the vertices of C_n. The vertex (a_s, b_s) has an important property. Let $M(n)$ denote the quantity

$$M(n) = \max\{\, |a - b| : 1 \le a, b \le n - 1 \text{ and } ab \equiv 1 \pmod{n}\}.$$

An immediate consequence of Proposition 2.3 is that

$$b_s - a_s = M(n).$$

The quantity $M(n)$ has been studied in [Ford et al. 05, Khan 01, Khan and Shparlinski 03]. It is shown in [Khan and Shparlinski 03] that

$$n - M(n) \ll n^{3/4 + o(1)}. \tag{2–1}$$

On the other hand, by [Ford et al. 05, Theorem 3.1], for almost all n,

$$n - M(n) \gg n^{1/2} (\log n)^{\delta/2} (\log\log n)^{3/4} f(n),$$

where

$$\delta = 1 - \frac{1 + \log\log 2}{\log 2} = 0.086071\dots$$

and $f(x)$ is any positive function tending monotonically to zero as $x \to \infty$. We recall that it has been proposed in [Ford et al. 05, Conjecture 4.1] that the above bound is quite tight.

Conjecture 2.4. *For almost all* n,

$$n - M(n) \ll n^{1/2} (\log n)^{\delta/2} (\log\log n)^{3/4} g(n),$$

where $g(x)$ *is any function tending monotonically to* ∞ *as* $x \to \infty$.

In support of Conjecture 2.4 we make the following observation. For a fixed $\varepsilon > 0$, define the set

$$\mathcal{N}(\varepsilon) = \{n \in \mathbb{N} : \exists d \mid (n-1) \text{ such that } n^{1/2-\varepsilon} \le d \le n^{1/2}\}.$$

By [Hall and Tenenbaum 88, Theorem 22], $\mathcal{N}(\varepsilon)$ has positive asymptotic density. Since

$$d\left(n - \frac{n-1}{d}\right) \equiv 1 \pmod{n},$$

we see that

$$n - M(n) \le n - \left(n - \frac{n-1}{d} - d\right) = \frac{n-1}{d} + d \ll n^{1/2+\varepsilon},$$

for every n with this property. This immediately implies that for any $\varepsilon > 0$,

$$n - M(n) \le n^{1/2+\varepsilon}$$

for a set of values of n of positive density, which is a weaker form of what is assumed in Conjecture 2.4. In [Ford et al. 05], one can also find more developed heuristic arguments supporting Conjecture 2.4.

We make one further remark about the vertex (a_s, b_s). Following [Tenenbaum 76], we introduce the quantities

$$\rho_1(m) = \max_{d \mid m,\, d \le \sqrt{m}} d \quad \text{and} \quad \rho_2(m) = \min_{d \mid m,\, d \ge \sqrt{m}} d.$$

We note that

$$a_s = \rho_1(kn - 1) \quad \text{and} \quad (n - b_s) = \rho_2(kn - 1),$$

where k is the integer such that $a_s(n - b_s) = kn - 1$.

2.4 Heuristics

Our heuristic attempt to approximate $v(n)$ makes use of a probabilistic model. Specifically, we view the points of $n^{-1}G_n$ as being randomly distributed in the unit square (which is supported by theoretic results of [Cobeli and Zaharescu 01, Granville et al. 05, Vajaitu and Zaharescu 02, Zhang 96, Zheng 96]) and then appeal to [Rényi and Sulanke 63, Satz 1]. Let $\mathcal{R}$ be a convex polygon in the plane with r vertices and let P_i, $i = 1, \ldots, n$, be n points chosen at random in $\mathcal{R}$ with uniform distribution. Let X_n be the number of sides of the convex closure of the points P_i, and let $E(X_n)$ be the expectation of X_n. Then

$$E(X_n) = \frac{2}{3} r(\log n + \gamma) + c_{\mathcal{R}} + o(1), \qquad (2\text{–}2)$$

where $\gamma = 0.577215\ldots$ is the Euler constant, and $c_{\mathcal{R}}$ depends on $\mathcal{R}$ and is maximal when $\mathcal{R}$ is a regular r-gon or is affinely equivalent to a regular r-gon. In particular, for the unit square $\mathcal{R} = [0, 1]^2$, we have

$$c_{\mathcal{R}} = -\frac{8}{3} \log 2.$$

More precise results are given in [Buchta and Reitzner 97], but they do not affect our arguments.

Using (2–2) with $r = 4$, it is plausible to conjecture that for most n,

$$v(n) \approx h(n), \qquad (2\text{–}3)$$

where

$$h(n) = \frac{8}{3}(\log \varphi(n) + \gamma - \log 2).$$

A portion of our work has been to generate numerical data to test this conjecture.

3. BOUNDS ON $v(n)$

3.1 Lower Bounds

Here we give a lower bound on $v(n)$ in terms of the number of divisors function $\tau(n)$. We begin by establishing some notation and making a couple of pertinent observations.

For a fixed n, let us consider the curves $\alpha_j(n)$ and $\beta_j(n)$ defined by

$$\alpha_j(n) : x(n - y) = jn - 1, \quad 1 \le x \le y \le n - 1,$$

and

$$\beta_j(n) : y(n - x) = jn - 1, \quad 1 \le y \le x \le n - 1.$$

A key observation used repeatedly is that for each point of G_n, there is a j in the range $1, \ldots, \lceil n/4 \rceil$ such that the point lies on the curve $\alpha_j(n)$ or $\beta_j(n)$. We denote the region bounded by the curves $\alpha_1(n)$ and $\beta_1(n)$ by $\mathcal{R}_n$. Figure 2 is an illustrative example. We note that the outermost curves are $\alpha_1(41)$ and $\beta_1(41)$.

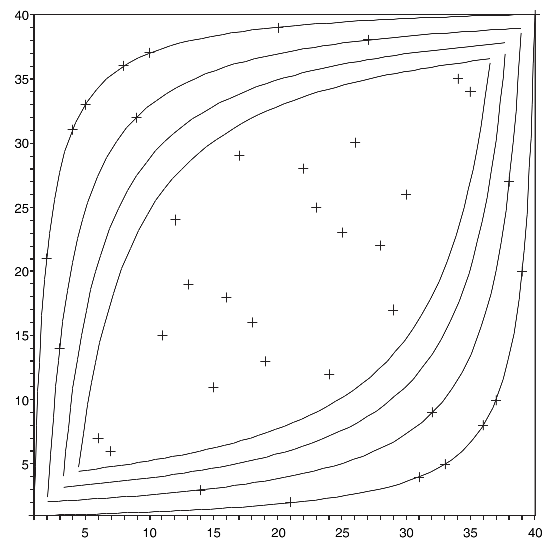

FIGURE 2. The graph G_{41} and the curves $\alpha_j(41)$, $\beta_j(41)$, $j = 1, 2, 3, 4$.

For an integer $s \ge 1$, we define

$$T(s) = \max_{i=1,\ldots,\tau(s)-1} \frac{d_{i+1}}{d_i},$$

where $1 = d_1 < \cdots < d_{\tau(s)} = s$ are the positive divisors of s.

Clearly,

$$T(s) \le P(s), \qquad (3\text{–}1)$$

where $P(s)$ denotes the largest prime divisor of s.

Let D_n be the convex closure of the points

$$(d_i, n - (n - 1)/d_i), (n - (n - 1)/d_i, d_i),$$

for $i = 1, \ldots, \tau(n - 1)$. Clearly, we have the inclusions $D_n \subseteq C_n \subseteq \mathcal{R}_n$. We remark that if $n - 1$ is prime, the set D_n is simply the line segment connecting the points $(1, 1)$ and $(n - 1, n - 1)$.

The purpose of our next proposition is to give a criterion to determine which of the $\alpha_j(n)$, $2 \le j \le \lceil n/4 \rceil$, lie strictly in the interior of D_n and hence strictly in the interior of C_n. We denote by Γ_n the set of boundary points (x, y) of D_n such that $y \ge x$; that is, $\Gamma_n = \{(x, y) : (x, y) \in \partial D_n, y \ge x\}$.

Proposition 3.1. *Let $1 = d_1 < \cdots < d_{\tau(n-1)} = n - 1$ be the positive divisors of $n - 1$. Then for any integer $m \ge 2$, $i = 1, \ldots, \tau(n - 1) - 1$,*

$$\Gamma_n \cap \alpha_m(n) = \varnothing \Leftrightarrow \frac{d_{i+1}}{d_i} + \frac{d_i}{d_{i+1}} < 4m - 2 + \frac{4(m - 1)}{n - 1}.$$

Proof: This is a routine computation, and so we only sketch an outline. The polygonal curve Γ_n is the union of line segments

$$L_i : (1-t)(d_i, n-(n-1)/d_i) + t(d_{i+1}, n-(n-1)/d_{i+1}),$$

$0 \le t \le 1$, with $i = 1, \ldots, (\tau(n-1)-1)$. Now $L_i \cap \alpha_m(n) = \varnothing$ if and only if the quadratic equation

$$(d_{i+1} - d_i)\left(\frac{n-1}{d_{i+1}} - \frac{n-1}{d_i}\right)t^2 - (d_{i+1} - d_i)\left(\frac{n-1}{d_{i+1}} - \frac{n-1}{d_i}\right)t + (1-m)n = 0$$

in t has no real solutions. $\square$

A useful consequence of Proposition 3.1 is that if

$$m \ge \left\lfloor \frac{T(n-1)+3}{4} \right\rfloor, \qquad (3\text{–}2)$$

with $m \in \mathbb{Z}$ and $m \ge 2$, then $\Gamma_n \cap \alpha_m(n) = \varnothing$.

Theorem 3.2. *For all $n \ge 2$,*

$$v(n) \ge 2\left(\tau(n-1) - 1\right),$$

and for sufficiently large x,

$$\#\left\{n \le x : v(n) = 2\left(\tau(n-1)-1\right)\right\} \gg \frac{x}{\log x}.$$

Proof: Since $C_n \subseteq \mathcal{R}_n$, any $(x,y) \in G_n \cap (\alpha_1(n) \cup \beta_1(n))$ is a vertex of C_n, and either x or y is a divisor of $(n-1)$. Therefore, $v(n) \ge 2\left(\tau(n-1)-1\right)$.

By (3–2), we have $\Gamma_n \cap \alpha_2(n) = \varnothing$ for every n with $T(n-1) \le 5$. Consequently, for such n, all of the vertices of C_n lie on $\alpha_1(n) \cup \beta_1(n)$, and thus $v(n) = 2\left(\tau(n-1)-1\right)$. On the other hand, by [Saias 97, Theorem 1], we know that for any fixed t and sufficiently large x,

$$\#\left\{n \le x : T(n-1) \le t\right\} \asymp \frac{x \log t}{\log x}.$$

Applying this result with $t = 5$, we conclude the proof. $\square$

It is easy to construct explicit examples of n with $v(n) = 2\left(\tau(n-1)-1\right)$. For instance, it follows from (3–1) and (3–2) that this holds for $n = 2^r 3^s 5^t + 1$, where r, s, t are nonnegative integers.

Since for any $\delta > 0$, we have

$$\limsup_{k\to\infty} \tau(k) 2^{-(1-\delta)\log k/\log\log k} = \infty$$

(see [Hardy and Wright 79, Theorem 317]), the same holds true for $v(n)$, and so we can infer that the heuristic estimate (2–3) is sometimes exponentially smaller than $v(n)$.

Corollary 3.3. *For any $\delta > 0$,*

$$\limsup_{n\to\infty} v(n) 2^{-(\frac{3}{8}-\delta)h(n)/\log h(n)} = \infty.$$

We have that $v(n) \ge 2(\tau(n-1)-1)$, and it is natural to ask when one has strict inequality. Our next result gives a partial answer to this question. Specifically, we exhibit a set of positive density for which we have strict inequality. Furthermore, if we assume Conjecture 2.4, then we have strict inequality for almost all n.

Theorem 3.4. *The strict inequality*

$$v(n) > 2(\tau(n-1)-1)$$

holds

(i) *for a set of n of positive density;*

(ii) *for almost all n, provided that for almost all n, we have $n - M(n) \le n^{1/2+o(1)}$.*

Proof: (i) Let

$$\mathcal{E} = \{n \ : \ v(n) = 2(\tau(n-1)-1)\},$$

$$\mathcal{I} = \{n \ : \ a_s\,(n-b_s) = n-1\}, \text{ and } \mathcal{I}(x) = \mathcal{I} \cap [1,x].$$

It is important to note that the values of s, a_s, and b_s all depend on n. We remind the reader of the following properties of the point (a_s, b_s) used in the proof below: it is the highest vertex of C_n that lies on or below the line $x+y=n$, $M(n) = b_s - a_s$, and $a_s \le n - b_s$. Clearly, $\mathcal{E} \subseteq \mathcal{I}$.

Let

$$\mathcal{A} = \{n \ : \ \exists p \text{ prime with } p|(n-1) \text{ and } p \ge n^{0.76}\}.$$

Using Mertens's formula (see [Hardy and Wright 79, Theorem 427]), we obtain

$$\#(\mathcal{A} \cap [1,x]) \ge \sum_{x^{0.76} \le p \le x} \left\lfloor \frac{x-1}{p} \right\rfloor \sim (\log(25/19))x.$$

We complete our proof by showing that aside for possibly a finite number of exceptions, the elements of $\mathcal{A}$ do not belong to $\mathcal{E}$. Since $a_s \le n - b_s \le n - M(n)$ and, by (2–1), $n - M(n) \ll n^{3/4+o(1)}$, we conclude that for $n \in I$, with n large, any prime divisor of $(n-1)$ is less than $n^{0.76}$. Consequently, $\mathcal{A} \cap \mathcal{I}$ is finite and hence $\mathcal{A} \cap \mathcal{E}$ is finite.

(ii) We now prove the following conditional statement. If for almost all n, $n - M(n) \le n^{1/2} g(n)$ with some function $g(n) = n^{o(1)}$, then $\#\mathcal{I}(x) = o(x)$.

Without loss of generality, we may assume that $g(n)$ is monotonically increasing. This time we write $\mathcal{I}(x)$ as the disjoint union of three sets: $\mathcal{J}_1(x)$, $\mathcal{J}_2(x)$, and $\mathcal{J}_3(x)$, where

$$\mathcal{J}_1(x) = \left\{ n \in \mathcal{I}(x) : n - b_s \le \frac{\sqrt{x}}{g(x)} \right\},$$
$$\mathcal{J}_2(x) = \left\{ n \in \mathcal{I}(x) : \frac{\sqrt{x}}{g(x)} < n - b_s \le \sqrt{x} g(x) \right\},$$
$$\mathcal{J}_3(x) = \left\{ n \in \mathcal{I}(x) : \sqrt{x} g(x) < n - b_s < x^{0.76} \right\}.$$

Now $\#\mathcal{J}_1(x) \le x g(x)^{-2} = o(x)$, and by our assumption we also have $\#\mathcal{J}_3(x) = o(x)$. So to conclude we need to show that $\#\mathcal{J}_2(x) = o(x)$. This follows by the following observation. Let

$$H(x, y, z) = \{ n \le x : \exists d | n \text{ with } y < d \le z \}.$$

Then

$$\#\mathcal{J}_2(x) \le H\left(x, \sqrt{x}/g(x), \sqrt{x} g(x)\right),$$

and by [Ford 08, Theorem 1],

$$H\left(x, \sqrt{x}/g(x), \sqrt{x} g(x)\right) = o(x),$$

which concludes the proof. □

We remark that the assumption of Theorem 3.4 (ii) is weaker than Conjecture 2.4, whose bound probably holds for almost all primes. This would then imply that

$$v(p) > 2\left(\tau(p-1) - 1\right)$$

for almost all primes p. On the other hand, it is reasonable to expect that there are infinitely many primes of the form $n = 2^r 3^s 5^t + 1$ (in fact even of the form $p = 3 \cdot 2^r + 1$), and therefore equality would occur infinitely often as well. We conclude this section by proving that $v(n)$ can be substantially larger than $\tau(n-1)$.

Theorem 3.5. *There is an infinite sequence of integers n_j with*

$$v(n_j) \ge \exp\left(\left(\frac{2\log 2}{11} + o(1)\right) \frac{\log n_j}{\log\log n_j}\right)$$

and

$$\tau(n_j - 1) = 2.$$

Proof: Let n be a shifted prime, that is, $n = p+1$, where p is prime. We first show that for such integers,

$$v(n) = v(p+1) \ge 2(\tau(2p+1) - 3).$$

Let ℓ be the line through $(1,1)$ that is tangent to $\alpha_2(n)$. Since $(1,1)$ and (p,p) are the only points of G_n on $\alpha_1(n)$, all of the points of G_n lie on or below ℓ. A straightforward calculation shows that ℓ meets $\alpha_2(n)$ at the point (x,y) where the x-coordinate is

$$x = \frac{1}{1 - ((p+1)/(2p+1))^{1/2}} < 4.$$

Hence every divisor d of $2p+1$ with $3 < d < (2p+1)/3$ gives rise to a vertex on $\alpha_2(n)$. Consequently, the number of vertices on $\alpha_2(n)$ is at least $\tau(2p+1) - 4$. By symmetry there is an equal number of vertices on $\beta_2(n)$, and since $(1,1)$ and (p,p) are also vertices of C_n, we obtain the desired inequality.

We now let Q_j denote the product of the first j odd primes and set p_j to be the smallest prime satisfying the congruence $2p_j \equiv -1 \pmod{Q_j}$. By the prime number theorem, $\log Q_j \sim j \log j$, and by Heath-Brown's version of Linnik's theorem [Heath-Brown 92] we have $p_j < c Q_j^{11/2}$, for an absolute constant $c \ge 1$. On combining $p_j < c Q_j^{11/2}$ with the asymptotic $\log Q_j \sim j \log j$, we obtain

$$\begin{aligned} \tau(2p_j + 1) &\ge \tau(Q_j) = 2^j \\ &\ge \exp\left(\left(\frac{2\log 2}{11} + o(1)\right) \frac{\log p_j}{\log\log p_j}\right). \end{aligned}$$

Setting $n_j = p_j + 1$, we conclude the proof. □

In particular, we see from Theorem 3.5 that

$$\limsup_{n\to\infty} \frac{\log v(n)}{\log \tau(n-1)} = \infty.$$

Furthermore, we can replace the terms $\log v(n)$ and $\log \tau(n-1)$ by the k-fold iteration of the logarithm for any $k \in \mathbb{N}$. Unfortunately, we do not see any approaches to establishing the following conjecture.

Conjecture 3.6. *We have*

$$\liminf_{n\to\infty} v(n) = \infty.$$

3.2 Upper Bounds

We have the following upper bound on $v(n)$.

Theorem 3.7. *For $n \to \infty$,*

$$v(n) \le n^{3/4+o(1)}.$$

Proof: In Section 2.2, we labeled the highest vertex of C_n in the triangle $\mathcal{T}_n$ by (a_s, b_s). Trivially, $s \le a_s$ and $a_s \le n - b_s$. Hence

$$\begin{aligned} v(n) &\le 4s + 2 \le 4a_s + 2 \le 2(n - b_s + a_s + 1) \\ &= 2(n - M(n) + 1), \end{aligned}$$

and the bound (2–1) concludes the proof. □

Most certainly the bound of Theorem 3.7 is not tight. If we assume Conjecture 2.4, then

$$v(n) \le n^{1/2+o(1)}$$

for almost all n. This still seems too high, and the actual order of $v(n)$ is almost certainly much smaller. A different upper bound for $v(n)$ can be derived from (3–2). For integers n such that $n-1$ has only small prime factors, this upper bound is significantly better than that given in Theorem 3.7.

Theorem 3.8. *For $n \to \infty$,*

$$v(n) \le T(n-1)n^{o(1)}.$$

Proof: From (3–2) we see that only points from the curves $\alpha_j(n)$ and $\beta_j(n)$ for which

$$j \le m_n = \left\lfloor \frac{T(n-1)+3}{4} \right\rfloor$$

contribute to $v(n)$. Since every curve $\alpha_j(n)$, $\beta_j(n)$ contains at most $\tau(jn-1)$ points of G_n, we derive

$$v(n) \le \sum_{j=1}^{m_n} 2\tau(jn-1).$$

We conclude by invoking the asymptotic inequality $\tau(r) \ll r^{o(1)}$; see [Hardy and Wright 79, Theorem 315]. □

4. COMPUTING C_n

4.1 Systematic Search Algorithm

We now describe a deterministic algorithm to construct the vertices of C_n that lie in the triangle $\mathcal{T}_n$. It is a variant of the famous algorithm given in [Graham 72] known as the "Graham scan." The main virtue of our algorithm compared to some other convex closure algorithms is that we do not need to generate and store all of the points of G_n before determining the convex closure. Instead, we generate the points one by one, discard most of them along the way, and halt in a reasonable amount of time.

Algorithm 4.1.

1. Set $a_0 := 1$; $b_0 := 1$.
2. For $i = 0, 1, \ldots$:
   (a) Set $a_{i+1} :=$ to be the smallest integer $a \in \mathbb{Z}_n^*$ satisfying the inequalities

   $$a_i < a \le \frac{n + a_i - b_i}{2} \quad \text{and} \quad b_i - a_i < a^{-1} - a.$$

   If either of the above conditions cannot be met, the algorithm terminates.

   (b) Set $b_{i+1} := a^{-1}$.

   (c) Convexity check:

      i. If $i = 1$ goto Step 2(a).

      ii. If $i \ge 2$ and the angle between the points (a_{i-1}, b_{i-1}), (a_i, b_i), and (a_{i+1}, b_{i+1}) is reflex, then return to Step 2(a); otherwise, discard the point (a_i, b_i) and set

      $$a_i := a_{i+1}, \quad b_i := b_{i+1}, \quad i := i-1,$$

      and return to Step 2(c).

We note that the inequalities in Step 2(a) are motivated by Proposition 2.3. Clearly, Algorithm 4.1 is deterministic, and it immediately follows from (2–1) that its complexity is $O(n^{3/4+o(1)})$.

4.2 A Factorization-Based Algorithm

The observation that the points in $G_n \cap \alpha_1(n)$ are vertices of C_n combined with (3–2) allows us to devise a variation on Algorithm 4.1. The idea is to use factorization to create a smaller input set and then run the algorithm.

Let $\mathcal{P}_n$ be the polygonal region with vertices

$$\begin{aligned} &(1, n-1),\ (1,1),\ (d_1, n-(n-1)/d_1),\ \ldots, \\ &\quad (d_k, n-(n-1)/d_k), \\ &\quad \big(((n-1)/d_k + d_k)/2, n - ((n-1)/d_k + d_k)/2\big), \\ &\quad \big(\sqrt{n-1}, n - \sqrt{n-1}\big), \end{aligned}$$

where $1 = d_0 < d_1 < \cdots < d_k$ are the factors of $n-1$ that are less than or equal to $\sqrt{n-1}$. Since the vertices of C_n can lie only on the curves $\alpha_j(n)$, $\beta_j(n)$ such that

$$j \le m_n = \left\lfloor \frac{T(n-1)+3}{4} \right\rfloor,$$

we need only determine which of the points of the union

$$U_n = \bigcup_{j=1}^{m_n} S_{j,n}$$

are vertices of C_n, where $S_{j,n} = \alpha_j(n) \cap G_n \cap \mathcal{P}_n$. It is useful to keep in mind that

$$\#U_n \le \sum_{j=1}^{m_n} \#S_{j,n} \le \sum_{j=1}^{m_n} \tau(jn-1) = m_n n^{o(1)};$$

see [Hardy and Wright 79, Theorem 315]. We now apply the following algorithm.

Algorithm 4.2.

1. Factorization:
   (a) Find all of the factors $1 = d_0 < d_1 < \cdots < d_k \le \sqrt{n-1}$ of $n-1$.
   (b) Set $S_1 := \{(1,1), (d_1, n-(n-1)/d_1), \ldots, (d_k, n-(n-1)/d_k)\}$.
   (c) Compute $t := T(n-1)$.
   (d) Set $m_n := \lfloor (t+3)/4 \rfloor$.
   (e) For $j = 2, \ldots, m_n$, factor $jn-1$ and construct the set $S_{j,n}$.
   (f) Set $U_n := \cup_{j=1}^{m_n} S_{j,n}$.
2. Determining the vertices:
   (a) Order the points of U_n by increasing first coordinate.
   (b) Apply the appropriate versions of Steps 2(a) and 2(c) of Algorithm 4.1 to the elements of U_n.

The complexity of Algorithm 4.2 depends on the type of algorithm we use for the factorization step. If we use any subexponential probabilistic factorization algorithm that runs in time $n^{o(1)}$ [Crandall and Pomerance 05, Chapter 6], then the complexity of Step 1 of Algorithm 4.2 is at most

$$\#U_n n^{o(1)} = m_n n^{o(1)}.$$

Furthermore, the complexity of Step 2 of Algorithm 4.2 is of the same form as well. So the overall complexity of Algorithm 4.2 is at most

$$m_n n^{o(1)} = T(n-1) n^{o(1)}.$$

This is lower than that of Algorithm 4.1 if $T(n-1) \le n^{3/4}$. For any fixed $\lambda \ge 0$, the proportion of the positive integers k with $T(k) \le k^{\lambda}$ is given by a certain continuous function $\psi(\lambda) > 0$ [Tenenbaum 79]. Using [Saias 97, Corollary A], we conclude that

$$\psi\left(\frac{3}{4}\right) = \int_0^{7/8} \rho\left(\frac{1}{x} - 1\right) \frac{dx}{x} = \int_{1/7}^{\infty} \rho(y) \frac{dy}{1+y}$$
$$= 0.866468\ldots,$$

where $\rho(u)$ is the Dickman function (see [Dickman 30] or [Tenenbaum 95, Section III.5.4]). Thus the proportion of the positive integers n with $T(n-1) \le n^{3/4}$ is $\psi(3/4) = 0.866468\ldots$. (The bound in Step 1(d) of Algorithm 4.2 is certainly not tight. It can probably be replaced by a bound of order $n^{o(1)}$ or even possibly a power of $\log n$, but unfortunately, we have not been able to prove such a result.)

On the other hand, if we use a deterministic factoring algorithm in Step 1, then Algorithm 4.2 is of complexity at most

$$m_n (m_n n)^{1/4+o(1)} = T(n-1)^{5/4} n^{1/4+o(1)}$$

unconditionally, and of complexity at most

$$m_n (m_n n)^{1/5+o(1)} = T(n-1)^{6/5} n^{1/5+o(1)}$$

under the extended Riemann hypothesis; see [Crandall and Pomerance 05, Section 6.3]. Accordingly, this is better than Algorithm 4.1 for $T(n-1) < n^{2/5}$ and $T(n-1) < n^{11/24}$, respectively. The corresponding proportions of the positive integers n satisfying these inequalities are $\psi(2/5)$ and $\psi(11/24)$. Since [Saias 97, Corollary A] expresses both $\psi(2/5)$ and $\psi(11/24)$ as double integrals, it is easier to compute $\psi(3/4)$ than either of these two values.

5. COMPUTATIONAL RESULTS

5.1 Expected value of $V(N)$

Let

$$\eta = \sum_p \frac{\log(1-1/p)}{p} = -0.580058\ldots,$$

where the sum runs over all prime numbers p. Surprisingly, this quantity has already appeared in various, seemingly unrelated, number-theoretic questions; see [Finch 03, p. 122].

Proposition 5.1. *We have*

$$\frac{1}{N} \sum_{n=1}^{N} \log \varphi(n) = \log N + \eta - 1 + O\left(\frac{\log \log N}{N}\right).$$

Proof: Obviously,

$$\frac{1}{N}\sum_{n=1}^{N}\log\varphi(n) = \frac{1}{N}\sum_{n=1}^{N}\log n + \frac{1}{N}\sum_{n=1}^{N}\sum_{p\mid n}\log\left(1-\frac{1}{p}\right),$$

where the last sum is taken over prime divisors $p \mid n$. The first sum on the right-hand side is $\log N - 1 + o(1)$ by Stirling's formula. By changing the order of summation in the second sum, we derive

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\sum_{p\mid n}\log(1-1/p) &= \frac{1}{N}\sum_{p\le N}\log(1-1/p)\sum_{\substack{n\le N\\ p\mid n}}1\\ &= \frac{1}{N}\sum_{p\le N}\log(1-1/p)\left(\frac{N}{p}+O(1)\right)\\ &= \sum_{p\le N}\frac{\log(1-1/p)}{p} + O\left(\frac{1}{N}\sum_{p\le N}\frac{1}{p}\right)\\ &= \sum_{p\le N}\frac{\log(1-1/p)}{p} + O\left(\frac{\log\log N}{N}\right),\end{aligned}$$

where the last step follows by Mertens's formula [Hardy and Wright 79, Theorem 427].

Observing that

$$\sum_{p\le N}\frac{\log(1-1/p)}{p} = \eta - \sum_{p>N}\frac{\log(1-1/p)}{p} = \eta + O\left(\frac{1}{N}\right),$$

we conclude our proof. □

Combining heuristic (2–3) with Proposition 5.1 for the average $V(N)$, we get the heuristic $V(N) \sim H(N)$, where

$$\begin{aligned}H(N) &= \frac{8}{3}(\log N + \gamma + \eta - 1 - \log 2)\\ &\approx 2.66666 \cdot \log N - 4.52264.\end{aligned}$$

In Figure 3, we compare the graphs of $H(N)$ and the least-squares approximation

$$L(N) = 3.551166 \cdot \log N - 9.610899 \qquad (5\text{–}1)$$

to $V(N)$, where N ranges over the interval $[2, 5770001]$. The values of $V(N)$ are represented by diamonds along the graph of $L(N)$, while $H(N)$ is the lower curve.

We see that although $V(N)$ behaves like a logarithmic function and thus resembles $H(N)$, the two functions clearly deviate from each other. This deviation seems to be of a regular nature, which suggests that there should be a natural explanation for the behavior of $V(N)$. In an attempt to understand this, we computed $v(n)$, $h(n)$,

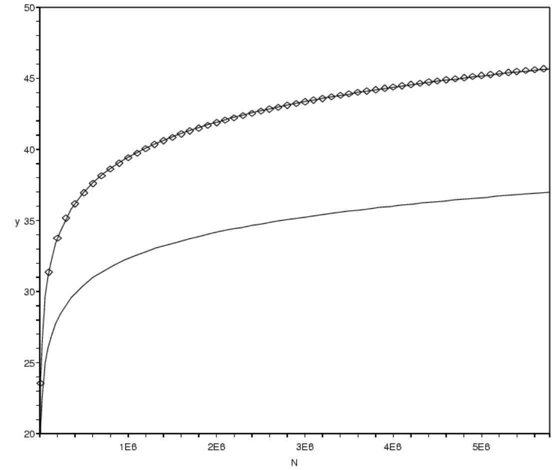

FIGURE 3. $V(N)$, $H(N)$, and $L(N)$ for $2 \le N \le 5{,}770{,}001$.

and $\tau(n-1)$ for 50,000 random integers in the interval $[10^6, 10^8]$ and did some comparisons. We present the individual data in the histograms in Figures 4 and 5, and the comparisons in Figures 6 through 9 and 11. In several histograms, the extreme values on the right are not visible. Hence, for visual clarity we have truncated them on the right. Under each histogram we state in the caption the minimum value, the maximum value, and the number of values that are not shown.

The histograms in Figures 6, 8, and 9 provide evidence that for most values of n, $h(n)$ is a good approximation to $v(n)$. This leads to the main peak. After comparing the histograms in Figures 6 and 7, it is plausible to speculate that some of the secondary peaks of $(v(n) - h(n))$ to the right of 0 correspond to large values of $\tau(n-1)$ that are quite "popular." It would be interesting to find (at

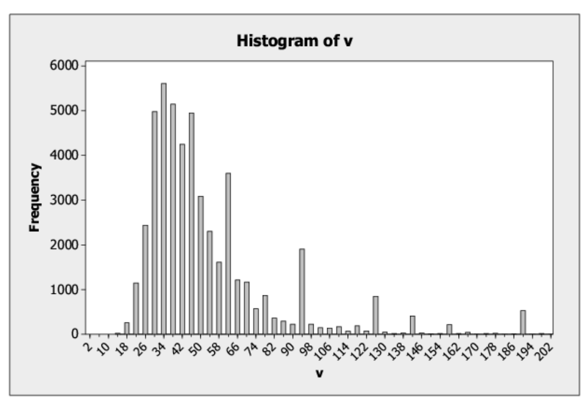

FIGURE 4. Frequency histogram of $v(n)$, min $= 14$, max $= 766$ (645 values omitted).

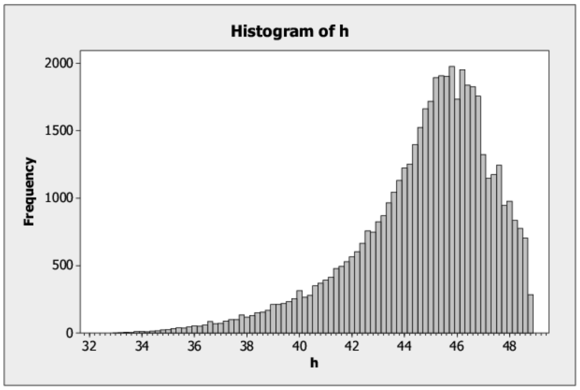

FIGURE 5. Frequency histogram of $h(n)$, min $= 33.01$, max $= 48.81$.

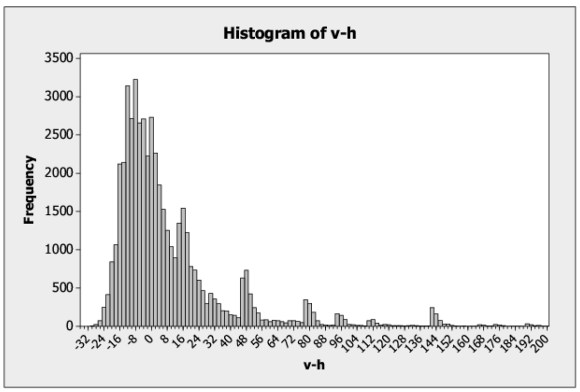

FIGURE 6. Frequency histogram of $(v-h)$, min $= -29.93$, max $= 714.41$ (458 values omitted).

least heuristically) a model that correctly describes these secondary peaks (their height, frequency, and so on).

Let X be a random variable. We say that X is *lognormally* distributed if $\log X$ is a normal distribution, and X is *loglogistically* distributed if $\log X$ is a logistic distribution. The probability density functions of the lognormal distribution is

$$f(x;\mu,\sigma) = \frac{\exp(-(\log x-\mu)^2/(2\sigma^2))}{\sqrt{2\pi}\sigma x},$$

where μ and σ^2 are the mean and variance of $\log(X)$. The probability density function of the loglogistic distribution is

$$f(x;\mu,\sigma) = \frac{\exp((\log x-\mu)/\sigma)}{\sigma x(1+\exp((\log x-\mu)/\sigma))^2},$$

where μ is the scale parameter and σ is the shape parameter.

In Figures 8 and 9 we have provided the scaled histograms of $(v-h)/h$ with the lognormal fit and the loglogistic fit respectively, since both of them seem to be reasonable approximations. Numerically, the loglogistic fit seems to be better. However, here is a heuristic

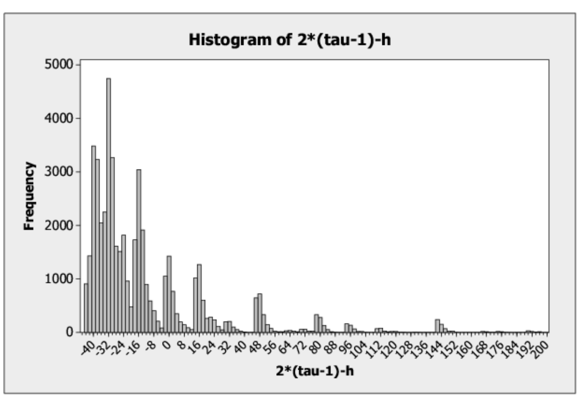

FIGURE 7. Frequency histogram of $2(\tau(n-1)-1)-h(n)$, min $= -44.96$, max $= 714.41$ (443 values omitted).

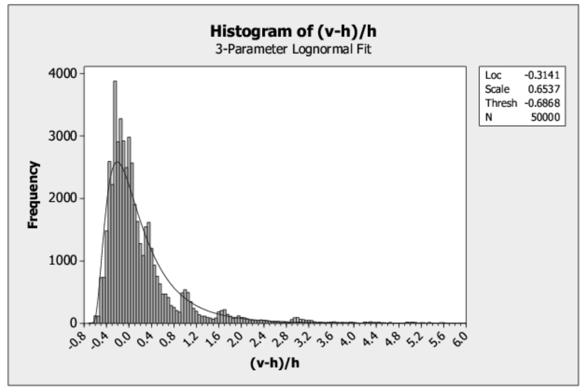

FIGURE 8. Frequency histogram of $(v-h)/h$ with a lognormal fit, min $= -0.68$, max $= 14.77$ (170 values omitted).

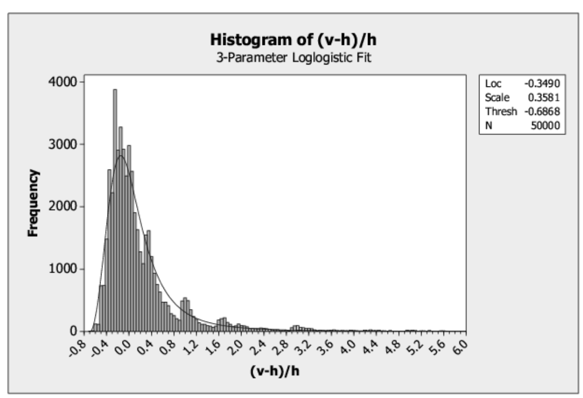

FIGURE 9. Frequency histogram of $(v-h)/h$ with a loglogistic fit, min $= -0.68$, max $= 14.77$ (170 values omitted).

argument (articulated by one of the referees) suggesting that the lognormal is more accurate. By the Erdős–Kac theorem [Tenenbaum 95, III.4.4, Theorem 8], $\omega(s)$ is normally distributed, and since $\tau(s) = 2^{\omega(s)+O(1)}$ for most integers s, we conclude that $\log \tau(s)$ is also normally distributed. Given the connection between $v(n)$ and the divisor functions, it seems reasonable to believe that a lognormal distribution is more accurate.

As a curiosity, we also mention that in the highly asymmetric histograms of Figures 6, 8, and 9, we still have $v(n) < h(n)$ in 25,057 out of 50,000 cases. It would be interesting to understand whether this is a coincidence or whether there is some regular effect behind this.

Our heuristic explanation for the difference between $V(N)$ and $H(N)$ is as follows. Overall, G_n behaves like a "pseudorandom" set, but as we observed in Theorem 3.2, there are some "regular points" on the convex closure arising from the divisors of $n-1$. For a typical integer n, these points have little effect, but for exceptional values of n, they make a substantial contribution to the value of $v(n)$, which is sufficient to interfere with the "pseudorandom" behavior of G_n.

To see this, it is useful to recall that although for most integers we have

$$\tau(n-1) = (\log n)^{\log 2 + o(1)} = h(n)^{\log 2 + o(1)}$$

[Hardy and Wright 79, Theorem 432], on average we have

$$\sum_{n=2}^{N} \tau(n-1) \sim N \log N \sim \frac{3N}{8} H(N)$$

[Hardy and Wright 79, Theorem 320]. Therefore, the contribution of $2\tau(n-1)$ from the points on the curves $\alpha_1(n)$ and $\beta_1(n)$ (see Theorem 3.2) is negligible compared to $h(n)$ for almost all n, but on average are of the same order as $0.75H(N)$. Thus it is plausible to assert that the values of $H(N)$ reflect only the "pseudorandom" nature of G_n, whereas the contribution of $2\tau(n-1)$ from the curves $\alpha_1(n)$, $\beta_1(n)$ reflects certain "regular" properties of the points of G_n.

5.2 Weighted Average Contribution of Divisors

The lower bound of Theorem 3.2 takes into account only the contribution from the divisors of $n-1$. It is plausible to assume that the divisors of $jn-1$ with "small" $j \geq 2$ also give some regular contribution to $v(n)$. This probably requires some completely new arguments, since the contribution from such divisors is certainly not additive.

Experimenting with some weighted averages involving $\tau(jn-1)$ for "small" values of j, we have found that $g_1(n)$ and $g_2(n)$, given by

$$g_1(n) = 2(\tau(n-1) - 1) + 2 \sum_{j=2}^{\lfloor \log n \rfloor} j^{-3/2} \tau(jn-1),$$

$$g_2(n) = 2(\tau(n-1) - 1) + 2e \sum_{j=2}^{\lfloor \log n \rfloor} e^{-j} \tau(jn-1),$$

appear to be "reasonable" numerical approximations to $v(n)$.

It is too early to make any substantiated conjecture about the true contribution from the divisors of $jn-1$ with $j \geq 2$. Numerical experiments for a much broader range as well as some new ideas are needed. Nevertheless, our calculation raises the following question.

Question 5.2. Are there "natural" coefficients c_j, $j = 2, 3 \ldots$, and function $J(n)$ such that if we define $g(n)$ to be

$$g(n) = 2\tau(n-1) + \sum_{j=2}^{J(n)} c_j \tau(jn-1),$$

then we have

$$V(N) \sim \frac{1}{N-1} \sum_{n=2}^{N} g(n)$$

as $N \to \infty$?

Clearly, if $V(N) \sim C \log N$, then the answer to Question 5.2 is positive, and one could then set $J(n) = 2$ and determine the value of c_2 by "reverse engineering." However, we are asking for coefficients c_j and a function $J(n)$ that can be explained by some intrinsic reasons, provided such reasons exist.

5.3 The difference $v(n) - 2(\tau(n-1) - 1)$

Another computer experiment that we ran on our random set of 50,000 integers was to check the values of the difference $v(n) - 2(\tau(n-1) - 1)$. The histogram of our experiment is given in Figure 10.

The graph of Figure 10 suggests that the most "popular" value of $v(n) - 2(\tau(n-1) - 1)$ is 0. There is some obvious regularity in the distribution of other values, which would be interesting to explain.

The way we have derived the lower bound of Theorem 3.2 on the frequency of the occurrence $v(n) = 2\,(\tau(n-1) - 1)$ from (3–2) raises the following question.

Question 5.3. Is $T(n-1) = O(1)$ for all (or nearly all) integers n with $v(n) = 2\,(\tau(n-1) - 1)$?

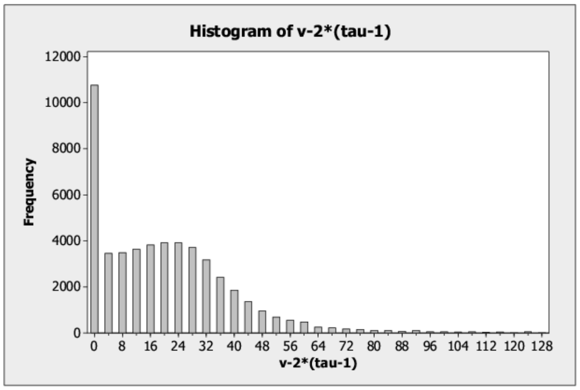

FIGURE 10. Frequency histogram of $v(n) - 2(\tau(n-1)-1)$, min $= 0$, max $= 484$ (199 values omitted).

An affirmative answer to this question would then allow us to conclude that

$$\#\{n \le x : v(n) = 2(\tau(n-1)-1)\} \asymp \frac{x}{\log x}.$$

In our random set of 50,000 integers we have 10,764 integers satisfying the equality $v(n) = 2(\tau(n-1)-1)$. For this latter set of integers we have computed the value of $t(n)$, where $t(n) = \lfloor (T(n-1)+3)/4 \rfloor$. We give this histogram in Figure 11. We remark that for 7198 integers of this sample, the value of $t(n)$ is 1, and for 2413 integers of this sample, the value of $t(n)$ is 2. Thus for at least 9611 integers out of 10,764 cases, we have $\Gamma_n \cap \alpha_2(n) = \varnothing$.

We have also found on examining the data that $v(n) - 2(\tau(n-1)-1)$ is invariably a multiple of 4, and this suggests the following conjecture.

Conjecture 5.4. *For almost all n,*

$$v(n) \equiv 2(\tau(n-1)-1) \pmod 4.$$

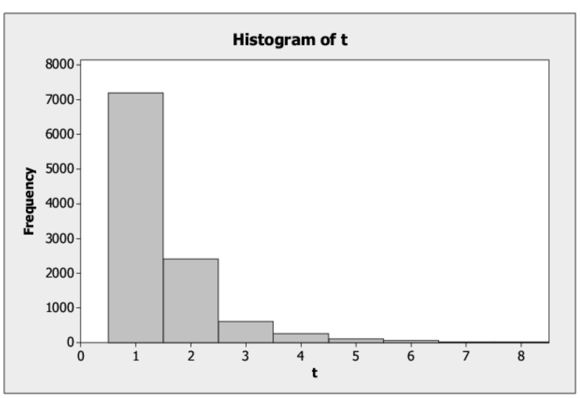

FIGURE 11. Frequency histogram of $t(n) = \lfloor (T(n-1)+3)/4 \rfloor$, min $= 1$, max $= 26$ (39 values omitted).

We have a simple heuristic argument for this conjecture. We know that $\tau(n-1)$ is odd if and only if $(n-1)$ is a square. Thus the conjecture reduces to the statement that for almost all n, $4 \nmid v(n)$. On invoking Propositions 2.1 and 2.3, we have that $4 \mid v(n)$ if and only if the vertex (a_s, b_s) lies on the line $x + y = n$. Intuitively, this seems to be a very rare occurrence (unfortunately, at present we are unable to put this key remark in a rigorous context); we typically see that $a_s + b_s = n$ only when n is the shifted square $m^2 + 1$.

6. OTHER CURVES

Studying the point sets

$$F_n(f) = \{(a,b) : a, b \in \mathbb{Z}, f(a,b) \equiv 0 \pmod n, \\ 0 \le a, b \le n-1\},$$

where $f(X,Y) \in \mathbb{Z}[X,Y]$, is certainly a natural line of inquiry, and this has been done in a number of works; see [Cobeli and Zaharescu 01, Granville et al. 05, Vajaitu and Zaharescu 02, Zheng 96] and references therein. In the case of prime modulus p, one can use the Bombieri bound of exponential sums [Bombieri 66] along a curve as a substitute for the bound of Kloosterman sums. In particular, for a prime $n = p$, under some mild assumptions on the polynomial f, one can easily obtain an analogue of Theorem 3.7 for sets $F_p(f)$. However, our other results are specific to the sets G_n and cannot be extended to other curves. It is worth remarking that for composite n, there are some analogues of the Bombieri bound [Stepanov and Shparlinski 89], but quite naturally, they are much weaker than the bound of [Bombieri 66]. So the Kloosterman sums give one of very few examples in which the strength of the bound remains almost unaffected by the arithmetic structure of the modulus.

Our preliminary tests show that the sets $F_n(f)$ and $F_p(f)$ have less "infrastructure" than G_n and behave more like truly random sets. For example, let $w_f(n)$ denote the number of vertices of the convex hull of $F_n(f)$. We now let

$$h_f(n) = \frac{8}{3}\left(\log\left(\#F_n(f)\right) + \gamma - \log 2\right).$$

The histograms in Figures 12–14 show the relative difference $(w_f - h_f)/h_f$ for random quadratic and cubic polynomials. For the histogram of Figure 12, we chose a random value of n in the interval $[10000, 300000]$. Then based on the value of n, we randomly chose the coefficients a, b, c and took $f(x,y)$ to be the polynomial

$$f(x,y) = y - ax^2 - bx - c.$$

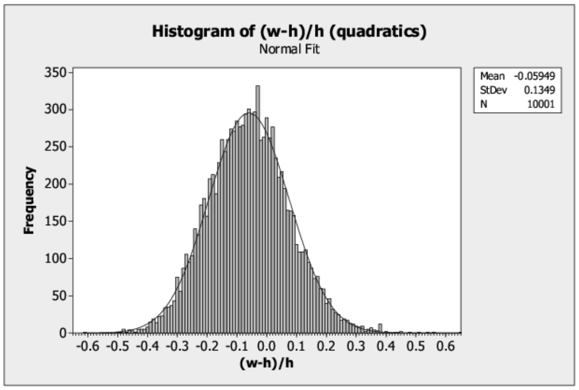

FIGURE 12. Frequency histogram of $(w_f - h_f)/h_f$ for random quadratics f over random n, min $= -0.607$, max $= 0.65$.

We did this for 10,000 values of n. For the histogram of Figure 13 we repeated this same experiment with random quadratic polynomials for 1000 random primes in the interval $[7919, 611953]$. For the histogram of Figure 14 we repeated our first numerical experiment (again for 10,000 values of n), but this time with random cubics

$$f(x, y) = y - ax^3 - bx^2 - cx - d.$$

The histograms of Figures 12–14 suggest that the quantities

$$\frac{w_f(n) - h_f(n)}{h_f(n)} \quad \text{and} \quad \frac{w_f(p) - h_f(p)}{h_f(p)}$$

are both normally distributed with mean 0, and so we make the following "Erdős–Kac"-type conjectures.

Let

$$\Phi_\sigma(z) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt$$

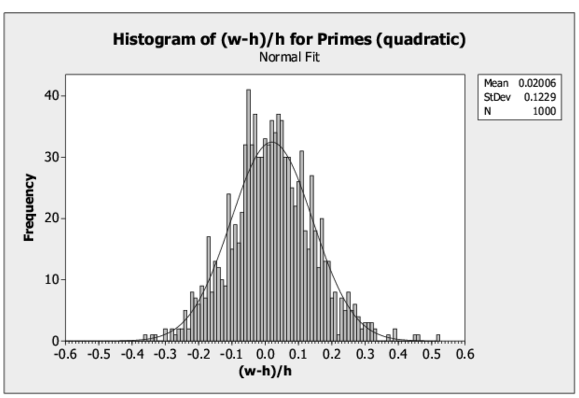

FIGURE 13. Frequency histogram of $(w_f - h_f)/h_f$ for random quadratics f over random p, min $= -0.355$, max $= 0.518$.

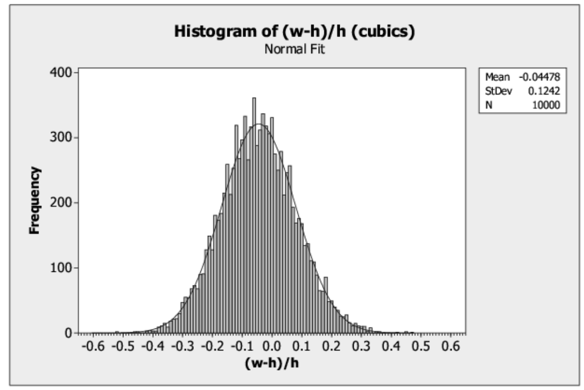

FIGURE 14. Frequency histogram of $(w_f - h_f)/h_f$ for random cubics f over random n, min $= -0.525$, max $= 0.473$.

denote the cumulative distribution function of a normal distribution with mean 0 and variance σ^2.

Conjecture 6.1. *Let $\mathcal{F} = (f_n)$ be a sequence of polynomials $f_n(x, y) \in \mathbb{Z}_n[x, y]$ of a fixed degree $d \geq 2$, chosen uniformly at random over the residue ring $\mathbb{Z}_n$, and let*

$$\sigma_{\mathcal{F}}(N) = \sqrt{\frac{1}{N} \sum_{n \leq N} \left(w_{f_n}(n)/h_{f_n}(n) - 1\right)^2},$$

$$\rho_{\mathcal{F}}(N) = \sqrt{\frac{1}{\pi(N)} \sum_{p \leq N} \left(w_{f_p}(p)/h_{f_p}(p) - 1\right)^2}.$$

Then for any real z,

$$\frac{\#\left\{n \leq N : \left(w_{f_n}(n) - h_{f_n}(n)\right)/h_{f_n}(n) \leq z\right\}}{N \Phi_{\sigma_{\mathcal{F}}(N)}(z)} \to 1,$$

$$\frac{\#\left\{p \leq N : \left(w_{f_p}(p) - h_{f_p}(p)\right)/h_{f_p}(p) \leq z\right\}}{\pi(N) \Phi_{\rho_{\mathcal{F}}(N)}(z)} \to 1,$$

with probability 1 (over the choice of $\mathcal{F} = (f_n)$) as $N \to \infty$.

ACKNOWLEDGMENTS

We thank the referees for their careful reading of the article. The manuscript substantially benefited from their comments. In particular, we are indebted to the referee who suggested using the result of [Saias 97] to show that

$$\#\left\{n \leq x : v(n) = 2\left(\tau(n-1) - 1\right)\right\} \gg \frac{x}{\log x}.$$

We are also grateful to Kevin Ford for suggesting the set $\mathcal{A}(x)$ that arises in the proof of Theorem 3.4, to Daniel Sutantyo for computing $\psi(3/4)$, and to Anthony Aidoo and Marsha Davis for assistance with the frequency histograms.

During the preparation of this paper, the second author was supported in part by ARC grant DP0556431.

REFERENCES

[Bombieri 66] E. Bombieri. "On Exponential Sums in Finite Fields." *Amer. J. Math.* 88 (1966), 71–105.

[Buchta and Reitzner 97] C. Buchta and M. Reitzner. "Equiaffine Inner Parallel Curves of a Plane Convex Body and the Convex Hulls of Randomly Chosen Points." *Prob. Theory Relat. Fields* 108 (1997), 385–415.

[Cobeli and Zaharescu 01] C. Cobeli and A. Zaharescu. "On the Distribution of the $\mathbb{F}_p$-points on an Affine Curve in r Dimensions." *Acta Arithmetica* 99 (2001), 321–329.

[Crandall and Pomerance 05] R. Crandall and C. Pomerance. *Prime Numbers: A Computational Perspective*. Berlin: Springer-Verlag, 2005.

[Dickman 30] K. Dickman. "On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude." *Ark. Math. Astr. Fys.* 22 (1930), 1–14.

[Finch 03] S. R. Finch. *Mathematical Constants*, Encyclopedia of Mathematics and Its Applications, 94, Cambridge: Cambridge Univ. Press, 2003.

[Ford 08] K. Ford. "The Distribution of Integers with a Divisor in a Given Interval." To appear in *Ann. Math.*, 2008.

[Ford et al. 05] K. Ford, M. R. Khan, I. E. Shparlinski, and C. L. Yankov. "On the Maximal Difference between an Element and Its Inverse in Residue Rings." *Proc. Amer. Math. Soc.* 133 (2005), 3463–3468.

[Graham 72] R. L. Graham. "An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set." *Inform. Process. Lett.* 1 (1972), 132–133.

[Granville et al. 05] A. Granville, I. E. Shparlinski, and A. Zaharescu. "On the Distribution of Rational Functions along a Curve over $\mathbb{F}_p$ and Residue Races." *J. Number Theory* 112 (2005), 216–237.

[Hall and Tenenbaum 88] R. Hall and G. Tenenbaum. *Divisors*, Cambridge Tracts in Mathematics, 90. Cambridge: Cambridge Univ. Press, 1988.

[Hardy and Wright 79] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*, 5th ed. Oxford: Oxford University Press, 1979.

[Heath-Brown 92] D. R. Heath-Brown. "Zero-Free Regions for Dirichlet L-functions, and the Least Prime in an Arithmetic Progression." *Proc. London Math. Soc. (3)* 64 (1992), 265–338.

[Hooley 57] C. Hooley. "An Asymptotic Formula in the Theory of Numbers." *Proc. London Math. Soc.* 7 (1957), 396–413.

[Khan 01] M. R. Khan. "Problem 10736: An Optimization with a Modular Constraint." *Amer. Math. Monthly* 108 (2001), 374–375.

[Khan and Shparlinski 03] M. R. Khan and I. E. Shparlinski. "On the Maximal Difference between an Element and Its Inverse Modulo n." *Period. Math. Hung.* 47 (2003), 111–117.

[Rényi and Sulanke 63] A. Rényi and R. Sulanke. "Über die konvexe Hülle von n zufällig gewählten Punkten." *Z. Wahrscheinlichkeitstheorie* 2 (1963), 75–84.

[Saias 97] É. Saias, "Entiers á Diviseurs Denses 1," *J. Number Theory* 62 (1997), 163–191.

[Shparlinski 07] I. E. Shparlinski. "Distribution of Points on Modular Hyperbolas." In *Sailing on the Sea of Number Theory: Proc. 4th China–Japan Seminar on Number Theory, Weihai, 2006*, pp. 155–189. River Edge, NJ: World Scientific, 2007.

[Stepanov and Shparlinski 89] S. A. Stepanov and I. E. Shparlinski. "Estimation of Trigonometric Sums with Rational and Algebraic Functions" (in Russian). *Automorphic Functions and Number Theory* 1 (1989), 5–18.

[Tenenbaum 76] G. Tenenbaum. "Sur Deux Fonctions de Diviseurs." *J. London Math. Soc.* 14 (1976), 521–526.

[Tenenbaum 79] G. Tenenbaum. "Lois de Réparitions des Diviseurs." *J. London Math. Soc.* 20 (1979), 165–176.

[Tenenbaum 95] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*. Cambridge: Cambridge Univ. Press, 1995.

[Vajaitu and Zaharescu 02] M. Vajaitu and A. Zaharescu. "Distribution of Values of Rational Maps on the $\mathbb{F}_p$-Points on an Affine Curve." *Monathsh. Math.* 136 (2002), 81–86.

[Zhang 96] W. Zhang. "On the Distribution of Inverses Modulo n." *J. Number Theory* 61 (1996), 301–310.

[Zheng 96] Z. Zheng. "The Distribution of Zeros of an Irreducible Curve over a Finite Field." *J. Number Theory* 59 (1996), 106–118.

Mizan R. Khan, Department of Mathematics and Computer Science, Eastern Connecticut State University, Willimantic, CT 06226 (khanm@easternct.edu)

Igor E. Shparlinski, Department of Computing, Macquarie University, Sydney, NSW 2109, Australia (igor@ics.mq.edu.au)

Christian L. Yankov, Department of Mathematics and Computer Science, Eastern Connecticut State University, Willimantic, CT 06226 (yankovc@easternct.edu)

Received April 19, 2006; accepted in revised form January 4, 2008.