# ON THE LENGTH OF THE CONTINUED FRACTION OF $\sqrt{q}$

J. B. Conrey

The formula for the class number of a real quadratic field of discriminant $d$ is given by

$$h(d) = \frac{L(1,\chi_d)\sqrt{d}}{\log \epsilon_d}$$

where $\chi_d$ is the quadratic character associated with the field and $\epsilon_d$ is the fundamental unit of the field. It was conjectured by Gauss that this class number is 1 for infinitely many $d$. Numerical evidence strongly supports Gauss' conjecture. However, in this direction it is only known that $h(d) < cd^{1/2}(\log d)^{-3}$ holds for some $c > 0$ and infinitely many $d$.

The Riemann Hypothesis for $L(s,\chi_d)$, implies that

$$\frac{1}{\log\log d} \ll L(1,\chi_d) \ll \log\log d.$$

Thus, the main part of the order of magnitude of $h(d)$ is controlled by $\log \epsilon_d$.

We can give an explicit formula for $\epsilon_d$ in terms of the continued fraction of $\sqrt{d}$. Consider the continued fraction expansion of $\sqrt{d}$. It is known that the expansion is periodic after the first term, and that the last term of the period is equal to twice the first term of the expansion. Also, the periodic part, except for the last term, is symmetric. Thus, we can express the continued fraction expansion of $\sqrt{d}$ as

$$\sqrt{d} = [n; \overline{a_1, a_2, \dots, a_{k-1}, a_k, a_{k-1}, \dots, a_2, a_1, 2n}]$$

or as

$$\sqrt{d} = [n; \overline{a_1, a_2, \dots, a_{k-1}, a_k, a_k, a_{k-1}, \dots, a_2, a_1, 2n}]$$

where $n = [\sqrt{d}]$. In most of what we do it won't matter which of these two forms appear, so we will write either simply as

$$\sqrt{d} = [n; \overline{a_1, a_2, \dots, a_2, a_1, 2n}]$$

Then the fundamental unit $\epsilon_d$ is given by

$$\epsilon_d = (n, a_1, a_2, \dots, a_2, a_1) + \sqrt{d}(a_1, \dots, a_1)$$

where the (.) notation can be defined recursively by $() = 1$, $(x) = x$, and

$$(a_1, \dots, a_k) = a_k(a_1, \dots, a_{k-1}) + (a_1, \dots, a_{k-2}).$$

Supported in part by an NSF grant

It is easy to see that

$$F_{k+1} \le (a_1, \dots, a_k) \le F_{k+1}(\max\{a_1, \dots, a_k\})^k$$

where $F_k$ is the $k$th Fibonacci number, so that

$$\ell(d) \ll \log \epsilon_d \ll \ell(d) \log d$$

where $\ell(d)$ denotes the length of the period of the continued fraction expansion of $\sqrt{d}$, which, for this paper, we take to be the number of $a_i$, ignoring the $2n$ at the end of the period. Hence, a good estimate for $h(d)$ really hinges on the length of the continued fraction expansion of $\sqrt{d}$.

Approaches in the past have included attempts to construct thin sets of $d$ for which $\ell(d)$ is large. Yamamoto has done this; his work leads to the bound for $h(d)$ mentioned above. In general, it is difficult to unhinge $\log \epsilon_d$ from $h(d)$.

In this paper, we study the distribution of the numbers $\ell(d)$. That is, for a fixed value $\ell$, we calculate the number $c_\ell(N)$ of $d < (N+1)^2$ for which $\ell(d) = \ell$. We conjecture that for each $\ell$ there exists a $c_\ell > 0$ such that

$$c_\ell(N) \sim c_\ell N (\log N)^{s_\ell}$$

where $s_\ell$ is 0 in the case that the norm of the fundamental unit is -1 (i.e. $\ell$ even) and is 1 in the case that the norm of the fundamental unit is 1 (the case where $\ell$ is odd). Moreover, we conjecture that the ratios $c_\ell(N)/(N \log N)$ are uniformly bounded. We present some numerical evidence to support this conjecture. We remark that this conjecture implies that $h(d) \ll (\log d)^3$ for a positive proportion of $d$.

We are able to verify this conjecture for $0 \le \ell \le 4$. Also, we can obtain reasonable bounds for $c_5(N)$. In addition, we analyze the structure of the problem and indicate a method which could give some insight into the general case.

**Basic Properties of (.)** The first few values of (.) are:

$$() = 1 \qquad (x) = x \qquad (x, y) = xy + 1 \qquad (x, y, z) = xyz + x + z$$

$$(x, y, z, w) = xyzw + xw + zw + xy + 1$$

The connection between the [.] and the (.) notations is given by

$$[b_0; b_1 \dots, b_k] = \frac{(b_0, \dots, b_k)}{(b_1, \dots, b_k)}. \tag{1}$$

Next, there is a more general recursion formula:

$$(b_1, \dots, b_k) = (b_1, \dots, b_r)(b_{r+1}, \dots, b_k) + (b_1, \dots, b_{r-1})(b_{r+2}, \dots, b_k) \tag{2}$$

for any $r$ with $0 \le r \le k$. One can prove this easily by induction. Suppose all such rules hold for expressions (.) with fewer than $k$ arguments. Then,

$$\begin{aligned}(b_1,\dots,b_k) &= b_k(b_1,\dots,b_{k-1}) + (b_1,\dots,b_{k-2})\\ &= b_k\{(b_1,\dots,b_r)(b_{r+1},\dots,b_{k-1}) + (b_1,\dots,b_{r-1})(b_{r+2},\dots,b_{k-1})\}\\ &\quad + (b_1,\dots,b_{k-2})\\ &= (b_1,\dots,b_r)\{(b_{r+1},\dots,b_k) - (b_{r+1},\dots,b_{k-2})\}\\ &\quad + (b_1,\dots,b_{r-1})\{(b_{r+2},\dots,b_k) - (b_{r+2},\dots,b_{k-2})\}\\ &\quad + (b_1,\dots,b_{k-2})\\ &= (b_1,\dots,b_r)(b_{r+1},\dots,b_k) + (b_1,\dots,b_{r-1})(b_{r+2},\dots,b_k)\\ &\quad + \{(b_1,\dots,b_{k-2}) - (b_1,\dots,b_r)(b_{r+1},\dots,b_{k-2})\\ &\qquad - (b_1,\dots,b_{r-1})(b_{r+2},\dots,b_{k-2})\}\end{aligned}$$

as desired. A particular case of this with $r = 1$ shows that the basic recursion formula can be applied from the front of an expression (.):

$$(b_1,\dots,b_k) = b_1(b_2,\dots,b_k) + (b_3,\dots,b_k). \tag{3}$$

This leads to the fact that (.) is unchanged when the arguments are reversed:

$$(b_1,\dots,b_k) = (b_k,\dots,b_1) \tag{4}$$

as is easily seen by induction: Assuming the reversal property for lengths smaller than $k$, we have

$$\begin{aligned}(b_1,\dots,b_k) &= b_k(b_1,\dots,b_{k-1}) + (b_1,\dots,b_{k-2})\\ &= b_k(b_{k-1},\dots,b_1) + (b_{k-2},\dots,b_1)\\ &= (b_k,\dots,b_1).\end{aligned}$$

Finally, we show that

$$(b_1,\dots,b_k)(b_2,\dots,b_{k+1}) - (b_1,\dots,b_{k+1})(b_2,\dots,b_k) = (-1)^k. \tag{5}$$

For the left hand side of the above is

$$\begin{aligned}&= (b_1,\dots,b_k)\{(b_2,\dots,b_k)b_{k+1} + (b_2,\dots,b_{k-1})\}\\ &\quad - (b_1,\dots,b_{k+1})(b_2,\dots,b_k)\\ &= (b_2,\dots,b_k)\{(b_1,\dots,b_{k+1}) - (b_1,\dots,b_{k-1})\}\\ &\quad + (b_1,\dots,b_k)(b_2,\dots,b_{k-1}) - (b_1,\dots,b_{k+1})(b_2,\dots,b_k)\\ &= -\{(b_1,\dots,b_{k-1})(b_2,\dots,b_k) - (b_1,\dots,b_k)(b_2,\dots,b_{k-1})\}\\ &= \cdots = (-1)^{k-1}\{(b_1)(b_2) - (b_1 b_2)()\} = (-1)^k.\end{aligned}$$

Using these properties we can prove

**Lemma 1.** *Suppose that*

$$\sqrt{q} = [n; \overline{a_1, a_2, \ldots, a_2, a_1, 2n}].$$

*Then*

$$q = \frac{(n, a_1, \ldots, a_1, n)}{(a_1, \ldots, a_1)}.$$

$$q = n^2 + \frac{(2n, a_1, a_2, \ldots, a_2)}{(a_1, \ldots, a_1)}.$$

*Proof.* It is easy to see that

$$\sqrt{q} = [n; a_1, a_2, \ldots, a_2, a_1, n + \sqrt{q}].$$

By (1), we have

$$\sqrt{q} = \frac{(n, a_1, \ldots, a_1, n + \sqrt{q})}{(a_1, \ldots, a_1, n + \sqrt{q})}.$$

By the recursive property, we then find

$$\sqrt{q} = \frac{(n + \sqrt{q})(n, a_1, \ldots, a_1) + (n, a_1, \ldots, a_2)}{(n + \sqrt{q})(a_1, \ldots, a_1) + (a_1, \ldots, a_2)}$$

which leads to

$$\sqrt{q} = \frac{\sqrt{q}(n, a_1, \ldots, a_1) + (n, a_1, \ldots, a_1, n)}{\sqrt{q}(a_1, \ldots, a_1) + (a_1, \ldots, a_1, n)}.$$

Now we have an equation

$$\sqrt{q} = \frac{A\sqrt{q} + B}{C\sqrt{q} + A}.$$

This expression simplifies to $q = B/C$. Thus, we have the formula

$$q = \frac{(n, a_1, \ldots, a_1, n)}{(a_1, \ldots, a_1)}.$$

We remark that the point of this paper really is to count how often the expression

$$\frac{(n, a_1, \ldots, a_1, n)}{(a_1, \ldots, a_1)}$$

is an integer. We write this in a slightly different form.

**Corollary.** *With the above notation,*

$$q = n^2 + \frac{(2n, a_1, a_2, \ldots, a_2)}{(a_1, \ldots, a_1)}.$$

*Proof.* We note that by several applications of the recursion formula, at least once from each end, we have

$$\begin{aligned}(n, a_1, \dots, a_1, n) &= n(a_1, \dots, a_1, n) + (a_2, \dots, a_1, n) \\ &= n^2(a_1, \dots, a_1) + n(a_1, \dots, a_2) + n(a_2, \dots, a_1) + (a_2, \dots, a_2) \\ &= n^2(a_1, \dots, a_1) + 2n(a_1, \dots, a_2) + (a_2, \dots, a_2) \\ &= n^2(a_1, \dots, a_1) + (2n, a_1, \dots, a_2).\end{aligned}$$

The statement of the Corollary now follows.

In order to count the $q$ with a given length of continued fraction expansion, we will instead count the number of $n$ and $a_i$ for which

$$(a_1, \dots, a_1) \mid (2n, a_1, a_2, \dots, a_2).$$

In order to be sure that this gives an accurate count for the $q$, we need to know that each $q$ can be expressed in at most one way as

$$q = n^2 + \frac{(2n, a_1, a_2, \dots, a_2)}{(a_1, \dots, a_1)}.$$

Strictly speaking, this is not completely true. If we accidentally took two periods instead of 1 from the continued fraction expansion of $\sqrt{q}$ we would arrive at a different expression. However, in that expression, at least one of the $a_i$ would be equal to $2n$. However, it is known that the $a_i$ that form the smallest period satisfy $a_i \le n$. With this extra proviso, the expression for $q$ above is unique, since all of the steps in the proofs of Lemma 1 and its Corollary are reversible.

**Lemma 2.** *Let $q$ be a positive integer, $q$ not equal to a perfect square. Then there is a unique positive integer $n$ and a unique symmetric tuple $(a_1, \dots, a_1)$ of positive integers with $a_i \le n$ such that*

$$q = \frac{(n, a_1, \dots, a_1, n)}{(a_1, \dots, a_1)} = n^2 + \frac{(2n, a_1, a_2, \dots, a_2)}{(a_1, \dots, a_1)}.$$

We can now give an expression for $c_\ell(N)$, namely

$$c_\ell(N) = \sum_{n \le N} \sum_{\substack{a_1, \dots, a_k \le n \\ (a_1, \dots, a_1) \mid (2n, a_1, \dots, a_2)}} 1$$

where $k = [\ell/2]$. This expression can be rewritten as

$$c_\ell(N) = \sum_{a_1, \dots, a_k \le N} \sum_{\substack{\max\{a_i\} \le n \le N \\ (2n, a_1, \dots, a_2) \equiv 0 \mod (a_1, \dots, a_1)}} 1.$$

The congruence condition can be expressed as

$$2n(a_1, \dots, a_2) \equiv -(a_2, \dots, a_2) \mod (a_1, \dots, a_1).$$

Now applying (5) we find that

$$(a_1,\dots,a_2)(a_2,\dots,a_1)-(a_1,\dots,a_1)(a_2,\dots,a_2)=(-1)^{\ell-1}.$$

In particular,

$$(a_1,\dots,a_2)^2\equiv(-1)^{\ell-1}\mod (a_1,\dots,a_1).$$

Thus, our problem is reduced to counting the sets of $a_i\le N$ and $n\le N$ with $n\ge a_i$ for each $i$ such that

$$2n\equiv(-1)^{\ell}(a_1,\dots,a_2)(a_2,\dots,a_2)\mod (a_1,\dots,a_1).$$

We give some examples for small $\ell$.

**Proposition.** *We have the following estimates:*

$$\begin{aligned}
c_0(N)&=N\\
c_1(N)&\sim\frac{3}{2}N\log N\\
c_2(N)&\sim\left(\frac{\pi}{4}\coth\frac{\pi}{2}-\frac{1}{2}\right)N\\
c_3(N)&<2\zeta(2)N\log N\\
c_4(N)&\ll N\\
c_5(N)&\ll N\log^3 N.
\end{aligned}$$

*Proof.*
For $\ell=0$ we have $\sqrt{q}=[n;\overline{2n}]$ so that $q=n^2+1$ and so $c_0(N)=N$.
For $\ell=1$ we have $\sqrt{q}=[n;\overline{x,2n}]$ so that

$$q=n^2+\frac{2n}{x}$$

and

$$\begin{aligned}
c_1(N)&=\sum_{n\le N}\sum_{\substack{x\le n\\ x\mid 2n}}1\\
&=\sum_{n\le N}(d(2n)-1).
\end{aligned}$$

This sum can be evaluated elementarily, or we can observe that

$$\begin{aligned}
\sum_{n=1}^{\infty}\frac{d(2n)}{n^s}&=2^s\sum_{n\ \text{even}}\frac{d(n)}{n^s}\\
&=2^s\left(\zeta(s)^2-(1-2^{-s})^2\zeta(s)^2\right)\\
&=2(1-2^{-s-1})\zeta(s)^2
\end{aligned}$$

so that

$$c_1(N) = \sum_{n\le N} d(2n) + O(N) \sim \frac{3}{2} N \log N.$$

For $\ell = 2$ we have $\sqrt{q} = [n; \overline{x, x, 2n}]$ so that

$$q = n^2 + \frac{(2n, x)}{(x, x)} = n^2 + \frac{2nx+1}{x^2+1}$$

and using $x^2 \equiv -1 \mod x^2+1$,

$$\begin{aligned} c_2(N) &= \sum_{x\le N} \sum_{\substack{x\le n\le N \\ 2n\equiv x \mod x^2+1}} 1 \\ &= \sum_{n\le N} \sum_{\substack{x\le n, M\ge 1 \\ 2n=x+M(x^2+1)}} 1. \end{aligned}$$

The condition $M \ge 1$ is necessary to ensure that $x \le n$. Note that $x$ and $M$ must both be even. We substitute $x = 2y$, $M = 2L$ and find that

$$\begin{aligned} c_2(N) &= \sum_{n\le N} \sum_{\substack{y\le n/2 \\ n=y+L(4y^2+1)}} 1 \\ &= \sum_{y+L(1+4y^2\le N} 1 = \sum_{1+y+4y^2\le N} \sum_{L\le \frac{N-y}{1+4y^2}} 1 \\ &= \sum_{1+y+4y^2\le N} \left( \frac{N-y}{1+4y^2} + O(1) \right) \\ &= N \sum_{y=1}^{\infty} \frac{1}{1+4y^2} + O(N^{1/2} \log N). \end{aligned}$$

Hence,

$$c_2(N) = N\left( \frac{\pi}{4} \coth \frac{\pi}{2} - \frac{1}{2} \right) + O(n^{1/2})$$

in a fairly straightforward way through Cauchy's Theorem applied to a contour integral of $(\cot \pi z)/(1+4z^2)$. Note that the constant here is $= 0.356\ldots$.

For $\ell = 3$ we have $\sqrt{q} = [n; \overline{x, y, x, 2n}]$ so that

$$q = n^2 + \frac{(2n, x, y)}{(x, y, x)} = n^2 + \frac{2n(1+xy)+y}{2x+x^2y}.$$

This leads to

$$c_3(N) = \sum_{x,y\le N} \sum_{\substack{x,y\le n\le N \\ 2n\equiv -y(1+xy) \mod 2x+x^2y}} 1.$$

The condition on $n$ requires that

$$2n = Lx(2+xy) - y(1+xy) = y + (Lx-y)(2+xy)$$

for some $L$. Let $M = Lx - y$. Clearly, $M > 0$, or else $2n = y + M(xy+2) < 0$. Also, $M > 0$ ensures that $x, y \le n$. Our sum is now

$$c_3(N) = \sum_{\substack{0<y+M(xy+2)<2N \\ y+M(xy+2) \text{ even} \\ M+y\equiv 0 \mod x}} 1.$$

We handle this sum with the following lemma.

**Lemma 3.** *We have*

$$\sum_{\substack{kmn\le x \\ m+n\equiv 0 \mod k}} 1 = \zeta(2)x\log x + O(x) \tag{6}$$

*where the sum is over positive integers $k, m, n$.*

*Proof.* We consider the generating function

$$\sum_{m+n\equiv 0 \mod k} \frac{1}{(kmn)^s} = \sum_k \frac{1}{k^s} \sum_{a=1}^{k} \sum_{n\equiv -a \mod k} \frac{1}{n^s} \sum_{m\equiv a \mod k} \frac{1}{m^s}.$$

In the innermost sum we replace $m$ by $mk+a$, factor out $k^{-s}$, and add and subtract a term $(m+1)^{-s}$ to see that the above is

$$= \sum_k \frac{1}{k^{2s}} \sum_{a=1}^{k-1} \sum_{n\equiv -a \mod k} \frac{1}{n^s} \sum_{m=0}^{\infty} \left(\frac{1}{(m+a/k)^s} - \frac{1}{(m+1)^s}\right) + \zeta(s)^2\zeta(2s).$$

Note that the term corresponding to $a = k$ is 0. The sum over $m$ is

$$= \sum_{m=0}^{\infty} \int_{m+a/k}^{m+1} su^{-s-1}\, du$$

and is majorized by

$$\int_{a/k}^{\infty} su^{-s-1}\, du = (a/k)^{-s}.$$

Thus, the generating function is $\zeta(s)^2\zeta(2s)$ plus a term which is majorized by

$$\sum_k \frac{1}{k^s} \sum_{a=1}^{k-1} \frac{1}{a^s} \sum_{n\equiv -a \mod k} \frac{1}{n^s}.$$

This expression is, in turn, majorized by

$$\sum_k \frac{1}{k^s} \sum_{a=1}^{k-1} \frac{1}{a^s} \left( \frac{1}{(k-a)^s} + k^{-s}\zeta(s) \right).$$

It is not difficult to see that this function has a simple pole at $s = 1$ with a bounded residue. Thus, we conclude that

$$\sum_{\substack{k,m,n \\ k|m+n}} \frac{1}{(kmn)^s} = \zeta(s)^2\zeta(2s) + Z(s)$$

for a function $Z(s)$ which has at worst a simple pole at $s = 1$. From this analysis the claim follows.

We apply the claim to the sum in question, and obtain an upper bound

$$c_3(N) \le \sum_{\substack{Mxy \le 2N \\ M+x \equiv 0 \mod y}} 1 \sim 2\zeta(2) N \log N + O(N).$$

It should be possible to get an asymptotic formula here with more work.

Next we take up the case of $\ell = 4$. Here we have $\sqrt{q} = [n; \overline{x, y, y, x, 2n}]$ and

$$q = n^2 + \frac{(2n, x, y, y)}{(x, y, y, x)} = n^2 + \frac{2n(x + y + xy^2) + (1 + y^2)}{(1 + xy)^2 + x^2}.$$

Then,

$$c_4(N) = \sum_{x,y \le N} \sum_{\substack{x,y \le n \le N \\ 2n \equiv (1+y^2)(x+y+xy^2) \mod ((1+xy)^2+x^2)}} 1$$

Here we are only aiming for an upper bound of the order of magnitude $N$, so we immediately make some simplifying assumptions. First of all,

$$c_4(N) \le \sum_{0 < (1+y^2)(x+y+xy^2) - M((1+xy)^2+x^2) < 2N} 1$$

where $x$ and $y$ are positive integers $\le N$ and $M$ is any integer. (Clearly, the complicated expression in the condition of summation is supposed to be $2n$. We have relaxed the condition that this expression is even and that $x$ and $y$ be smaller than $n$ which is why we have an inequality rather than equality.) The terms with $M < 0$ make a contribution which is

$$\ll \sum_{|M|x^2y^2 \le 2N} 1 \ll N.$$

The terms with $M = 0$ contribute

$$\ll \sum_{xy^4 \le 2N} 1 \ll N.$$

Now we consider the terms with $M > 0$. We let

$$g = y^2 + 1 - Mx \tag{7}$$

and

$$h = gy - M. \tag{8}$$

We then have to estimate

$$\sum_{0<gx+h(1+xy)<2N} 1$$

where $x$ and $y$ are positive integers $\le N$ and $g$ and $h$ are constrained by the equations (7) and (8). We claim that $g > 0$ and $h \ge 0$. First of all, $2n = g(x + y + xy^2) - M(1 + xy)$, so that $M > 0$ and $2n > 0$ imply that $g > 0$. Next, suppose that $h < 0$. Then, $gy < M \le Mx \le y^2$ so that $g \le y$. But, $2n = h(1 + xy) + gx > 0$ implies that $gx > (-h)(1 + xy) > (-h)xy \ge xy$ so that $g > y$. This contradiction implies that $h \ge 0$.

We now estimate the contribution from the terms with $h = 0$. This condition implies that $M = gy$ so that $g(1 + xy) = y^2 + 1$ or $y^2 - gxy + (1 - g) = 0$. Thus, each pair $(g, x)$ leads to at most 2 values of $y$. In fact, most pairs $(g, x)$ will not give a solution $y$ since that requires the discriminant $g^2x^2 + 4g - 4$ will not usually be a square. In fact,

$$g^2x^2 + 4g - 4 = u^2$$

implies that

$$(gx - u)(gx + u) = 4 - 4g.$$

We may as well assume that $u \ge 0$. Then $(gx + u) \mid (4 - 4g)$ implies that $x \le 4$. Thus, we see that the contribution from these $h = 0$ terms is

$$\ll \sum_{\substack{gx\le 2N \\ g^2x^2+4g-4= \text{ square}}} 1 \ll \sum_{\substack{gx\le 2N \\ x\le 4}} 1 \ll N.$$

Now we turn our attention to the case $g > 0$, $h > 0$. Eliminating $M$ from (7) and (8) gives

$$g(1 + xy) = y^2 + 1 + xh.$$

Thus, given $h$, $x$, and $y$, the value of $g$ is determined. Also, we see that

$$hx \equiv -1 - y^2 \mod (1 + xy)$$

whence

$$h \equiv y^3 + y \mod (1 + xy).$$

Thus, the sum in question is

$$\ll \sum_{\substack{hxy\ll N \\ h\equiv y^3+y \mod (1+xy)}} 1.$$

We remark that this sum is trivially

$$\ll \sum_{xy \ll N} \left( \frac{N}{(xy)^2} + 1 \right) \ll N \log N.$$

So, we are trying to save a $\log N$.

We can do this, but the argument is rather lengthy, and we omit it for the time being.

Now let us consider the case $\ell = 5$. We have

$$\sqrt{q} = [n; \overline{x, y, z, y, x, 2n}]$$

and

$$q = \frac{(n, x, y, z, y, x, n)}{(x, y, z, y, x)} = n^2 + \frac{(2n, x, y, z, y)}{(x, y, z, y, x)}.$$

The condition that $q$ is an integer translates to the congruence condition:

$$2n \equiv -(x, y, z, y)(y, z, y) \mod (x, y, z, y, x).$$

So, we need to count how many $n \leq N$ satisfy the above for some set of $x, y, z \leq n$. We will only be looking for an upper bound of the order of magnitude $N \log^3 N$.

We resort to a factorization that is available for odd $\ell$. Namely, if $\ell = 2k - 1$, then

$$\begin{aligned}(a_1, \ldots, a_k, \ldots, a_1) &= (a_1, \ldots, a_k)(a_{k-1}, \ldots, a_1) + (a_1, \ldots, a_{k-1})(a_{k-2}, \ldots, a_1) \\ &= (a_1, \ldots, a_{k-1})\{(a_1, \ldots, a_k) + (a_1, \ldots, a_{k-2})\}.\end{aligned}$$

Then, the congruence

$$2n \equiv -(a_1, \ldots, a_2)(a_2, \ldots, a_2) \mod (a_1, \ldots, a_1)$$

reduces to a pair of simultaneous congruences:

$$\begin{cases} 2n & \equiv -R \mod (a_1, \ldots, a_{k-1}) \\ 2n & \equiv R \mod \{(a_1, \ldots, a_k) + (a_1, \ldots, a_{k-2})\} \end{cases}$$

where

$$R = (a_2, \ldots, a_{k-1})\{(a_2, \ldots a_k) + (a_2, \ldots, a_{k-2})\}.$$

In the case $\ell = 5$ this leads to

$$\begin{cases} 2n & \equiv 2y + y^2 z \mod 1 + xy \\ 2n & \equiv -(2y + y^2 z) \mod 2x + z + xyz \end{cases}$$

We can derive most of our bound from the second of these congruences, simply ignoring the first one except in one case. The second congruence implies that $2n = M(2x + z + xyz) - (2y + y^2 z)$ for some $M$. Clearly, $M > 0$. Put $g = Mx - y$. Then,

$$2n = Mz + g(yz + 2).$$

We claim that $g \ge 0$. Suppose that $g < 0$. Then $Mx - y = g \le -1$ so that $y \ge Mx + 1$. Also, $Mz + g(yz + 2) \ge 1$ so that

$$Mz \ge (-g)(yz+2) + 1 \ge yz + 3 \ge z(Mx+1) + 3 \ge Mxz + z + 3$$

which is a contradiction. Therefore, $g \ge 0$. The terms with $g > 0$ contribute an amount which is

$$\ll \sum_{\substack{0<Mz+g(yz+2)<2N \\ g=Mx-y}} 1 \ll \sum_{gxy \ll N} \sum_{Mx=g+y} 1.$$

This sum can be handled similarly to the situation that arose in the $\ell = 3$ case, leading to a bound $O(N \log^3 N)$ by using

**Lemma 4.** *We have*

$$\sum_{mn \le x} \frac{d(m+n)}{mn} \sim \frac{1}{2} x \log^2 x.$$

*Proof.* The proof is almost the same as that of Lemma 3. The relevant generating function is

$$\begin{aligned}
\sum_{m,n} \frac{d(m+n)}{(mn)^s} &= \sum_{m,n} \frac{1}{(mn)^s} \sum_{k \mid m+n} 1 \\
&= \sum_k \sum_{a=1}^{k} \sum_{n \equiv -a \mod k} \frac{1}{n^s} \sum_{m \equiv a \mod k} \frac{1}{m^s} \\
&= \sum_k \frac{1}{k^s} \sum_{a=1}^{k-1} \sum_{n \equiv -a \mod k} \frac{1}{n^s} \sum_{m=0}^{\infty} \left( \frac{1}{(m+a/k)^s} - \frac{1}{(m+1)^s} \right) \\
&\quad + \zeta(s)^3.
\end{aligned}$$

We continue as before and deduce the Lemma.

For the terms with $g = 0$, we need to use the first congruence as well. In this case, we have $Mx = y$ and our system may be written as

$$\begin{cases} 2n & \equiv (Mx, z, Mx) \mod (x, Mx) \\ 2n & \equiv -(Mx, z, Mx) \mod (x, Mx, z) + x \end{cases}$$

This system simplifies to

$$\begin{cases} 2n & \equiv -zM + 2Mx \mod Mx^2 + 1 \\ 2n & \equiv zM \mod Mx^2 z + 2x + z \end{cases}$$

Then, $2n = zM + r(Mx^2 z + 2x + z)$ for some $r \ge 0$. The contribution of terms with $r > 0$ is

$$\ll \sum_{rzMx^2 \ll N} 1 \ll N \log^2 N.$$

If $r = 0$, then $2n = Mz$ so that the first congruence implies that

$$2Mx \equiv 2Mz \mod Mx^2 + 1.$$

The contribution of these terms is

$$\ll \sum_{\substack{2n=Mz\ll N \\ y=Mx\ll N \\ 2x\equiv 2z \mod Mx^2+1}} 1$$
$$\ll \sum_{Mx\ll N} 1 + \sum_{\substack{Mx\ll N, Mz\ll N \\ z=x+s(Mx^2+1)/2}} 1 \ll N \log N.$$

**Larger $\ell$.**

We can begin the $\ell = 6$ case, but can't quite estimate the relevant sums. Perhaps it is instructive to include our partial work here. We have

$$\sqrt{q} = [n; \overline{x, y, z, z, y, x, 2n}].$$

We have to count the number of $x, y, z, M$ such that

$$0 < 2n = (x, y, z, z, y)(y, z, z, y) - M(x, y, z, z, y, x) << N.$$

Note, the terms with $M \leq 0$ contribute $\ll N$. We let

$$g = (y, z, z, y) - Mx.$$

Clearly, if $m \leq 0$ then $g > 0$. Now,

$$2n = g(x, y, z, z, y) - M(x, y, z, z).$$

Therefore, if $M > 0$, then $g > 0$. Hence, it is always the case that $g > 0$. Next, we let

$$h = M - gy.$$

Then,

$$2n = g(x, y, z) - h(x, y, z, z).$$

Note that $g(1 + xy) = (y, z, z, y) - hx$; in particular, $g \mid (y, z, z, y) - hx$, and that $M$ is determined once $g, h, y$ are given. Thus, the terms here with $h < 0$ contribute

$$\ll \sum_{hxyz^2\ll N} ((y, z, z, y) - hx) \ll N^{1+\epsilon}.$$

If $h = 0$, then $M = gy$ and these terms contribute

$$\ll \sum_{gxyz\ll N} 1 \ll N \log^3 N.$$

Let

$$i = g - hz.$$

Note that if $h \le 0$, then $i > 0$. Also,

$$2n = i(x, y, z) - h(x, y).$$

So, if $h > 0$, then $i > 0$. Thus, in all cases $i > 0$. Now, let

$$j = h - iz.$$

Then

$$2n = ix - j(x, y).$$

Finally, let $k = i - jy$. Then,

$$2n = kx - j.$$

We claim that $j < 0$. For, we calculate that

$$\begin{aligned} M &= h + gy = (h + (i + hz)y = h(1 + yz) + iy = h(y, z) + iy \\ &= (iz + j)(y, z) + iy = i(z(y, z) + y) + j(y, z) = i(y, z, z) + j(y, z) \\ &= (k + jy)(y, z, z) + j(y, z) = j(y, z, z, y) + k(y, z, z). \end{aligned}$$

Thus,

$$g = (y, z, z, y) - Mx = (y, z, z, y)(1 - xj) - kx(y, z, z).$$

Suppose that $j > 0$. Then it must be the case that $k < 0$, since $g > 0$. But then $2n = kx - j < 0$ which is a contradiction. Hence, $j < 0$ as claimed.

We also note that we can show that

$$k = q - n^2 = \frac{(2n, x, y, z, z, y)}{(x, y, z, z, y, x)}.$$

This argument works in the general case, too. Consider

$$\sqrt{q} = [n; \overline{a_1, \dots, a_1, 2n}].$$

Let us consider the case that $\ell$ is even. We have to count the $n \le N$ such that

$$2n = (a_1, \dots, a_2)(a_2, \dots, a_2) - g_0(a_1, \dots, a_1).$$

Here, $g_0$ is arbitrary. However, the terms with $g_0 \le 0$ cause no difficulty. We make substitutions:

$$\begin{aligned} g_1 &= (a_2, \dots, a_2) - g_0 a_1 \\ g_2 &= g_0 - g_1 a_2 \\ g_3 &= g_1 - g_2 a_3 \\ &\dots \\ g_k &= g_{k-2} - g_{k-1} a_k \\ g_{k+1} &= g_{k-1} - g_k a_k \\ &\dots \\ g_{2k-1} &= g_{2k-3} - g_{2k-2} a_2 \\ g_{2k} &= g_{2k-2} - g_{2k-1} a_1. \end{aligned}$$

These substitutions lead to

$$\begin{aligned}
2n &= g_1(a_1,\ldots,a_2) - g_0(a_1,\ldots,a_3)\\
&= g_1(a_1,\ldots,a_4) - g_2(a_1,\ldots,a_3)\\
&= g_3(a_1,\ldots,a_4) - g_2(a_1,\ldots,a_5)\\
&= g_3(a_1,\ldots,a_6) - g_4(a_1,\ldots,a_5)\\
&\ldots\\
&= g_k(a_1,\ldots,a_k) - g_{k-1}(a_1,\ldots,a_{k-1})\\
&= g_k(a_1,\ldots,a_{k-2}) - g_{k+1}(a_1,\ldots,a_{k-1})\\
&\ldots\\
&= g_{2k-3}(a_1,a_2,a_3) - g_{2k-4}(a_1,a_2)\\
&= g_{2k-3}(a_1) - g_{2k-2}(a_1,a_2)\\
&= g_{2k-1}(a_1) - g_{2k-2} = -g_{2k}.
\end{aligned}$$

When we put all of the substitutions together, we can solve for $g_0$ as follows:

$$\begin{aligned}
g_0 &= g_2 + g_1 a_2\\
&= g_3(a_2) + g_2(a_2,a_3)\\
&= g_4(a_2,a_3) + g_3(a_2,a_3,a_4)\\
&= \ldots\\
&= g_{2k-1}(a_2,\ldots,a_2) + g_{2k-2}(a_2,\ldots,a_3)\\
&= g_{2k}(a_2,\ldots,a_1) + g_{2k-1}(a_2,\ldots,a_2).
\end{aligned}$$

We claim that $g_1 > 0$. For if $g_0 \le 0$ then the assertion follows from the substitution equation for $g_1$; if $g_0 > 0$, then the assertion follows from the equation for $2n$ in terms of $g_1$ and $g_0$. Similarly, $g_i > 0$ for all odd $i$.

Now we claim that $g_{2k-2} < 0$. By combining

$$g_1 = (a_2,\ldots,a_2) - g_0 a_1$$

and

$$g_0 = g_{2k-1}(a_2,\ldots,a_2) + g_{2k-2}(a_2,\ldots,a_3)$$

to eliminate $g_0$, we obtain

$$g_1 = (a_2,\ldots,a_2)(1 - g_{2k-2}a_1) - g_{2k-1}a_1(a_2,\ldots,a_3).$$

Thus, if $g_{2k-2} > 0$, then $g_{2k-1} < 0$. But then $2n = g_{2k-1}(a_1) - g_{2k-2} < 0$ which is a contradiction. Therefore, $g_{2k-2} < 0$ as claimed.

Now let $t = -g_{2k-2} > 0$. The significance of this inequality is that we have an equation

$$2n = g_{2k-3}a_1 - g_{2k-2}(a_1,a_2) = g_{2k-3}a_1 + t(a_1,a_2),$$

in which all the terms are positive. To rephrase the original question, we have to count how many solutions $(a_1,\ldots,a_k,g_0,\ldots,g_{2k-1}$ there are to

$$0 < g_{2k-3}a_1 + t(a_1,a_2) \ll N$$

subject to all of the substitution equations, and subject to $0 < a_i \le N$ for each $i$.

It might be of interest to note also that

$$g_{2k-1} = q - n^2 = \frac{(2n, a_1, \dots, a_2)}{(a_1, \dots, a_1)}.$$

To see this, recall, from the last expression for $2n$ that $2n = -g_{2k}$. By the last formula for $g_0$,

$$\begin{aligned} g_0 &= g_{2k}(a_2, \dots, a_1) + g_{2k-1}(a_2, \dots, a_2) \\ &= -2n(a_2, \dots, a_1) + g_{2k-1}(a_2, \dots, a_2). \end{aligned}$$

But,

$$2n = (a_1, \dots, a_2)(a_2, \dots, a_2) - g_0(a_1, \dots, a_1).$$

Eliminating $g_0$ from these two equations gives

$$\begin{aligned} 2n = (a_1, \dots, a_2)(a_2, \dots, a_2) - g_{2k-1}(a_1, \dots, a_1)(a_2, \dots, a_2) \\ + 2n(a_1, \dots, a_1)(a_2, \dots, a_2). \end{aligned}$$

This simplifies to

$$g_{2k-1}(a_1, \dots, a_1)(a_1, \dots, a_2) = 2n(a_1, \dots, a_2)^2 + (a_1, \dots, a_2)(a_2, \dots, a_2)$$

or

$$g_{2k-1} = \frac{(2n, a_1, \dots, a_2)}{(a_1, \dots, a_1)}.$$

**Concluding Remarks.** Why should we expect the conjectures to be true? The simple answer is that when we count how many $n \le N$ satisfy the congruence

$$2n \equiv (-1)^\ell (a_1, \dots, a_2)(a_2, \dots, a_2) \mod (a_1, \dots, a_1)$$

we expect there to be about

$$\frac{N}{(a_1, \dots, a_1)}$$

solutions. When we sum this over all $\ell$-tuples $(a_1, \dots, a_1)$ with $a_i \le N$ we get $\approx N$ for even $\ell$ and $\approx N \log N$ for odd $\ell$. The $\log N$ factor occurs in the odd case because the $a_k$ in the middle of $(a_1, \dots, a_1)$ appears only once and $\sum_{a_k \le N} \frac{1}{a_k} \sim \log N$.

The uniform boundedness of the implicit constants is suggested by the graphs below.

Two questions motivated by this work are as follows.

(1) Show that

$$\sum_{\max\{a_i\} \le N} \frac{1}{(a_1, \dots, a_1)} \ll N \log^{s_\ell}$$

holds uniformly in $\ell$, where $s_\ell = 0$ for $\ell$ even, and $s_\ell = 1$ for $\ell$ odd.

(2) Let $f_3(n)$ stand for the number of triples $(x, y, z)$ of positive integers such that

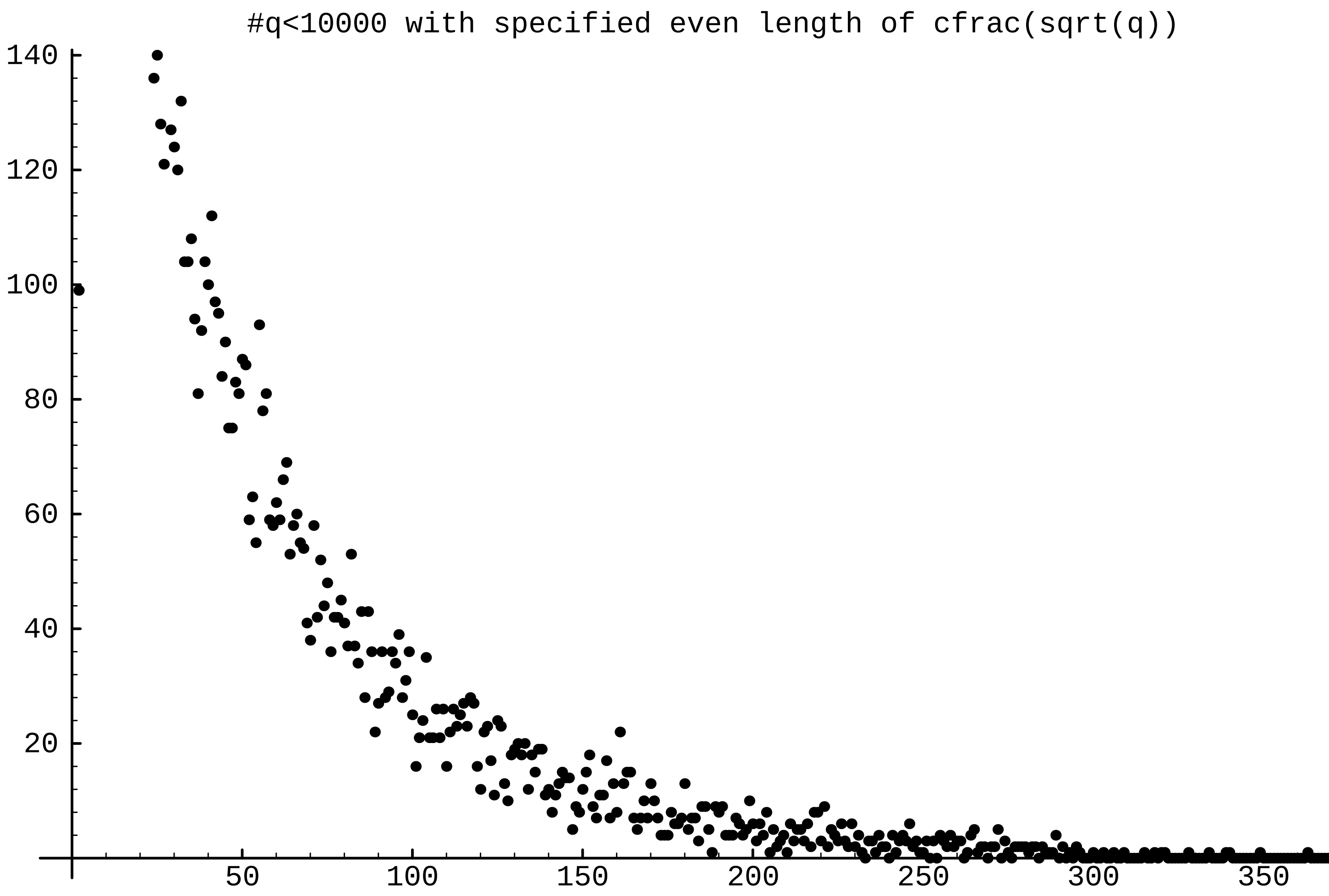
$$n = (x, y, z) = xyz + x + z.$$

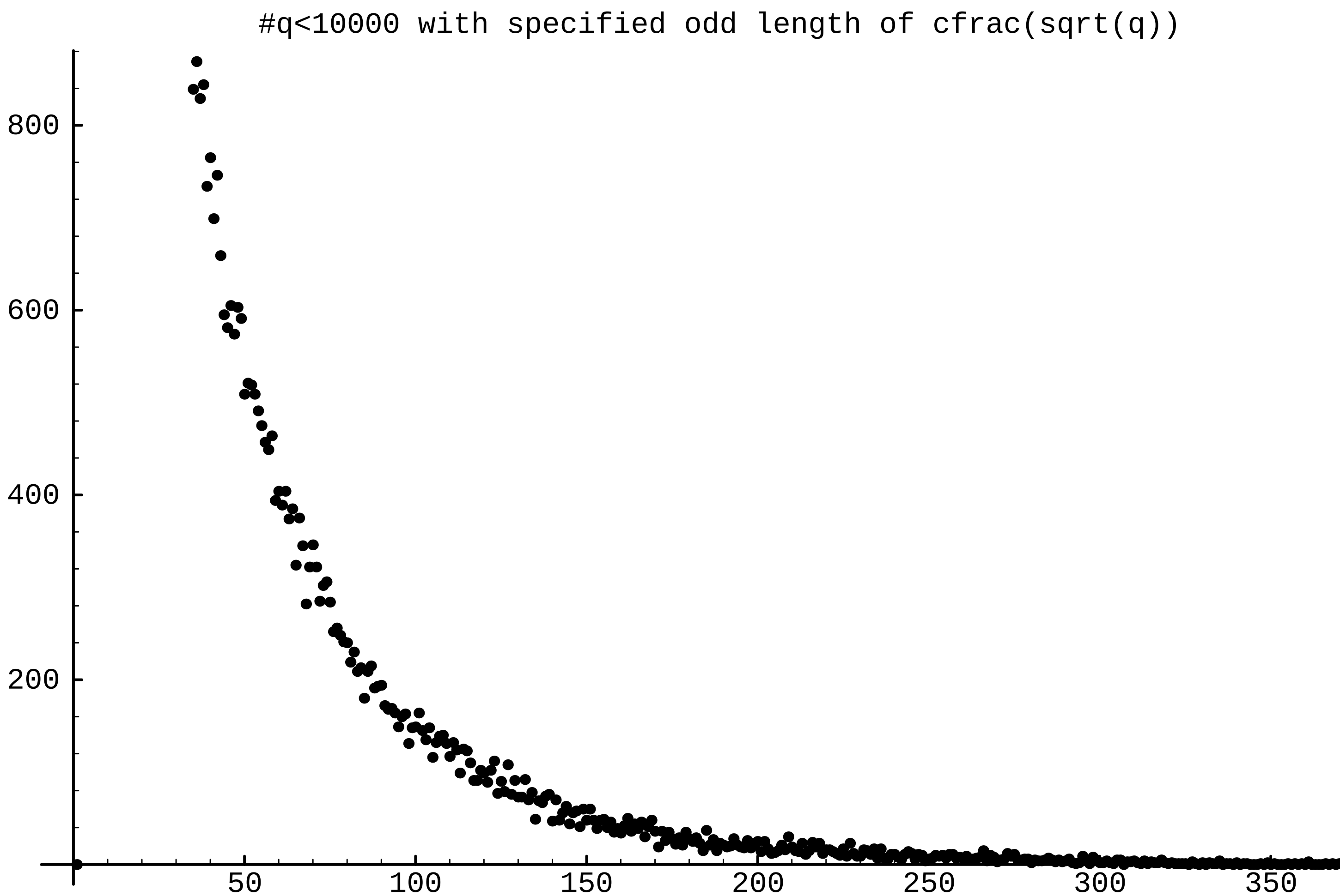
Show that

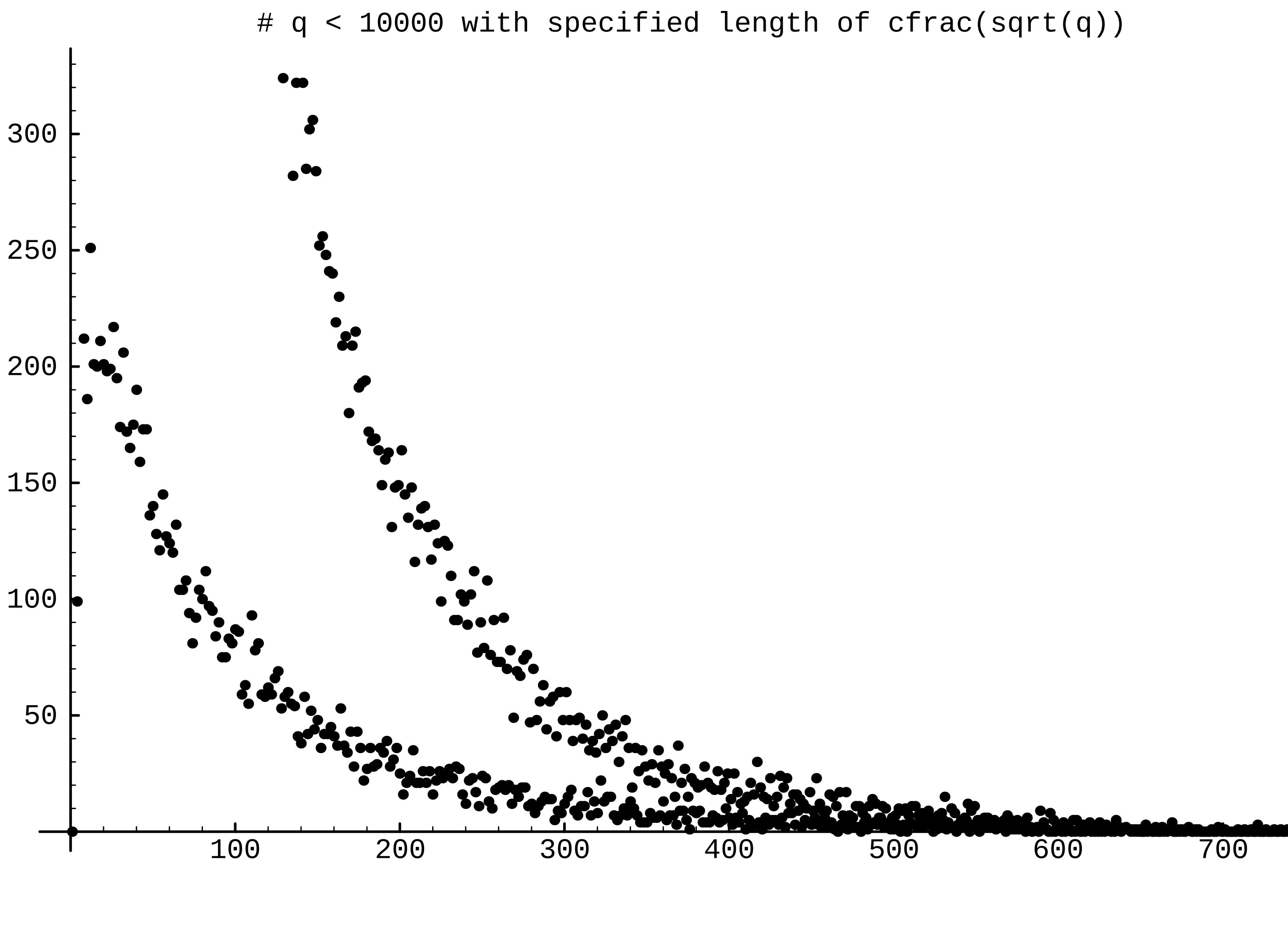
$$f(n) \ll_\epsilon n^\epsilon$$

for any $\epsilon > 0$.

Finally, we present some data in graphical form. The horizontal axis has lengths $\ell$ of continued fractions. Above each integer on the horizontal axis, there is a dot at the height which is equal to the number of $q \leq 10^5$ for which $\ell(q) = \ell$. The first graph has even $\ell$, the second graph has odd $\ell$, and the third graph has all $\ell$. odd $\ell$, and both combined.


#q<10000 with specified even length of cfrac(sqrt(q))
140
120
100
80
60
40
20
50
100
150
200
250
300
350


#q<10000 with specified odd length of cfrac(sqrt(q))
800
600
400
200
50
100
150
200
250
300
350


# q < 10000 with specified length of cfrac(sqrt(q))
300
250
200
150
100
50
100
200
300
400
500
600
700