THE 392 PROBLEM

Aaron Meyerowitz[1]
Department of Mathematics, Florida Atlantic University, Boca Raton, Florida
meyerowi@fau.edu

John Selfridge

Received: 7/1/11, Revised: 11/2/11, Accepted: 1/19/12, Published: 10/12/12

Abstract

Suppose that a is a positive non-square integer and we wish to multiply a subset of $\{a, a+1, a+2, \dots\}$ to form a square so that a is used and we minimize the largest number used. Now add the requirement that we must use at most f factors in forming the square. Erdős, Malouf, Selfridge, and Szekeres conjectured that if $a > 392$ then using $f = 3$ factors results in a smaller maximum than using $f = 2$ factors. Their paper on this subject reduces the conjecture to the case that a is a term in a certain sequence $\{s_N^2\}$, where $2s_N$ grows almost as fast as 6^N. We give theoretical and computational results which establish the result for $N < 20{,}000$ and for several infinite classes that comprise a positive proportion of the subscripts N.

1. Introduction

In many factoring algorithms, for example, see [2], one is interested in finding a non-empty subset of a given set of integers with product a square. The given set of numbers may be an image of an interval under a polynomial or something more complicated. There are two ways that this problem can be idealized: (1) assume that the numbers are randomly chosen from a long interval, (2) assume that the given set consists of all the integers in an interval $[a, b]$ with a given and b minimal. The first idealization has been considered in several papers, such as [1]. This paper, which is a sequel to [3], considers the second idealization.

Specifically, for a positive integer a, let $b = b(a)$ denote the least integer with $b \geq a$ such that some subset of $[a, b]$ that contains a, has a square product. For example, if $a = 2$ then $b = 6$ where the desired subset is $\{2, 3, 6\}$. We will only consider the case that a itself is not square because otherwise $b(a) = a$.

Let $f \geq 2$ be an integer and let $b = b_f(a)$ denote the least $b > a$ such that a subset of $[a, b]$ of size at most f and containing a has a square product. Then

[1]Supported by a grant from the Number Theory Foundation

$b_2(a) \geq b_3(a) \geq \cdots$, and eventually, this sequence stabilizes on $b(a)$.

For example, we have

$$\begin{aligned} 15 \times 60 &= 30^2 \\ 15 \times 20 \times 27 &= 90^2 \\ 15 \times 18 \times 20 \times 24 &= 360^2, \end{aligned}$$

and so we see that $b_2(15) = 60$, $b_3(15) = 27$ and $b(15) = 24$ as $b_f(15) = 24$ for all $f \geq 4$.

Let $a = rs^2$ with $r > 1$ and square-free. It is immediate that $b_2(a) = r\left(s+1\right)^2$. Do we ever have $b_2(a) = b_3(a)$? Yes, this occurs for $a = 8$ and $a = 392$. Erdős et al. [3] have the

Conjecture 1. (The 392 Problem) When $a > 392$ is not a square, $b_3(a) < b_2(a)$.

They reduce the problem dramatically. They show that $b_3(a) < b_2(a)$ for all non-square $a > 8$ except $a = 392$ and perhaps some (other) numbers of the form $a = 2s_N^2$ where s_N is the Nth term of the sequence $0, 2, 14, 84, 492, 2870, 16730 \ldots$ with recurrence

$$s_{N+2} = 6s_{N+1} - s_N + 2$$

Thus the conjecture reduces to answering the question "Is it the case that $b_3(a) < b_2(a)$ for $a = 2s_N^2$ when $3 \leq N$?"

In this paper we investigate these remaining open cases. The conjecture is still open but we will show among other things

- $b_3(a) < b_2(a)$ if $a = 2s_N^2$ for $3 \leq N \leq 20{,}000$.
- $b_3(a) < b_2(a)$ if $a = 2s_N^2$ and $N \pmod{21} \in \{0, 7, 8, 10, 11, 15, 16, 20\}$ or $N \pmod{15} \in \{0, 6, 7, 8, 14\}$.

There is no difficulty in principle in extending the first result well past $N = 20{,}000$. The second result can be expanded to include many more moduli. However, the next smallest moduli are 68, 205 and 253

The main focus of this paper is the theoretical results behind some extensive calculations (many of which are of independent interest) and the most interesting findings.

Here are the key steps of [3]: Recall that $a = rs^2$ with $r > 1$ and square-free.

- If r is not prime then $b_3(a) < b_2(a)$. Thus we assume $a = ps^2$ with p prime.
- For all q, if $p \neq q^2 + 1$ then $b_3(a) < b_2(a)$.
- If $p = q^2 + 1$ for some q and $p > 2$ then $b_3(a) < b_2(a)$.

This reduces the conjecture to showing that for every $s > 2$ except $s = 14$ there are integers x, y such that $2s^2 < x, y < 2(s+1)^2$ and $2s^2xy$ is a square. Evidently we must have $x = Mu^2$ and $y = 2Mv^2$ for some $M > 1$. This is most likely to work when $\left|u^2 - 2v^2\right|$ is small. Indeed we can almost always find a solution with $\left|u^2 - 2v^2\right| = 1$, i.e. one such that $\frac{u}{v}$ is a convergent to $\sqrt{2}$. For example $\frac{7}{5}$ and $\frac{17}{12}$ are two successive convergents and $\left(u^2, 2v^2\right) = (49, 50)$ can be used for $15 \leq s \leq 83$ while $\left(u^2, 2v^2\right) = (289, 288)$ works for $85 \leq s \leq 491$. However, neither works for $s = 84$ The only solutions for $\left[2 \cdot 84^2, 2 \cdot 85^2\right]$ are $(M, u, v) = (3, 69, 49)$ and $(7, 45, 32)$. For $\left[2 \cdot 492^2, 2 \cdot 493^2\right]$ there is just one solution: $(M, u, v) = (7, 263, 186)$.

2. Notation and Some Estimates

Let $\alpha = 1 + \sqrt{2} \approx 2.41421$ and $\overline{\alpha} = -1/\alpha = 1 - \sqrt{2} \approx -.41421$. Also, $\alpha^n = x_n + y_n\sqrt{2}$.

The *Nth critical interval* is $CI_N = \left[2s_N^2, 2(s_N+1)^2\right]$ where

$$s_N\sqrt{2}, (s_N+1)\sqrt{2} = \frac{y_{2N+1} \pm 1}{2}\sqrt{2} = \frac{\alpha^{2N+1} - \alpha^{-2N-1} \pm 2\sqrt{2}}{4}.$$

Thus the integers in CI_N have about $1.531\,N$ decimal digits.

We will sometimes write $[a, b]$ to mean $[\min(a, b), \max(a, b)]$ This should cause no confusion.

We now turn to the remaining cases of the 392 conjecture:

Conjecture 2. For $N > 2$, there are integers M, u, v such that the interval $\left[Mu^2, 2Mv^2\right]$ is properly contained in the Nth critical interval $\left[2s_N^2, 2(s_N+1)^2\right]$.

A *core* is an interval $\left[u^2, 2v^2\right]$ (order not important) with relatively prime ends. The *gap* of the interval is $g = \left|u^2 - 2v^2\right|$. We associate an element of $\mathbb{Z}[\sqrt{2}]$ with this core:

$$\left[u^2, 2v^2\right] \Leftrightarrow u + v\sqrt{2}$$

We will call $u + v\sqrt{2}$ a *starter* if $\gcd(u, 2v) = 1$ and either $u > 2v > 0$ or $v > u$. i.e., if $\left(u + v\sqrt{2}\right)\left(\sqrt{2} - 1\right) = (2v - u) + (u - v)\sqrt{2}$ has a negative coefficient. Then we call $|2v - u| + |u - v|\sqrt{2} = u' + v'\sqrt{2}$ the *paired starter*.

Claim 3. *Let $\sigma = u + v\sqrt{2}$ be a starter with gap $g = \left|u^2 - 2v^2\right|$ and define $\sigma' = |2v - u| + |u - v|\sqrt{2} = u' + v'\sqrt{2}$. Then*

(i) $\left|u - v\sqrt{2}\right| = \frac{g}{\sigma}$;

(ii) $\sigma' = \frac{g\alpha}{\sigma}$;

(iii) *σ' is a starter;*

(iv) *Both σ and σ' have the same gap $g = \left|u^2 - 2v^2\right| = \left|u'^2 - 2v'^2\right|$.*

These definitions given do not quite fit the case of gap $g = 1$. The strict inequalities do not allow either $1+0\sqrt{2}$ or $1+\sqrt{2}$ to be a starter. This will not be a problem since we are considering precisely the few cases which have no solution with gap $g = 1$.

Consider an interval $[2s^2, 2(s+1)^2]$ such as the interval $CI_N = \left[2s_N^2, 2\left(s_N+1\right)^2\right]$. A *solution* of this interval is a smaller interval $\left[Mu^2, 2Mv^2\right]$ properly contained in it. We assume that $\gcd(u, 2v) = 1$. Then the interval $\left[u^2, 2v^2\right]$ is called the *core* of the solution and the numbers , $g = \left|u^2 - 2v^2\right|$ and M are called the *gap* and the *broad multiplier* of the solution. We reserve the term *multiplier* (or *narrow multiplier*) for m, the largest odd square-free divisor of M. Note that $\left[Mu^2, 2Mv^2\right] = \left[mU^2, 2mV^2\right]$ where

$$(U, V) = \begin{cases} \left(\sqrt{\frac{M}{m}}u, \sqrt{\frac{M}{m}}v\right) & \frac{M}{m} \text{ is square} \\ \left(\sqrt{\frac{2M}{m}}v, \sqrt{\frac{M}{2m}}u\right) & \frac{M}{2m} \text{ is square} \end{cases}.$$

We have verified the conjecture up to critical interval 20,000. The integers in this interval have magnitude about 10^{30620}. We have two approaches to find solutions. The *multiplier method* seeks solutions with relatively small (narrow) multiplier and the *starter method* seeks solutions with a relatively small gap (more precisely with $u + v\sqrt{2}$ relatively small). The starter method can be carried out with large integer arithmetic or with calculations modulo $2g^2$ combined with some floating point computations.

It would be nice to be able to fix m and find many or all U, V, N such that

$$2s_N^2 < mU^2, 2mV^2 < 2\left(s_N+1\right)^2.$$

It is highly plausible that there are solutions for about $\frac{\sqrt{2}}{m}$ of the intervals but we can neither prove this nor find cases short of doing the obvious large integer calculations.

For a given starter $u + v\sqrt{2}$ we define a sequence of cores $\left[u_n^2, 2v_n^2\right]$ by

$$\begin{aligned}(u_0, v_0) &= (u, v) \\ (u_n, v_n) &= (u_{n-1} + 2v_{n-1}, u_{n-1} + v_{n-1}).\end{aligned}$$

So

$$\begin{aligned}u_n + v_n\sqrt{2} = \left(u_0 + v_0\sqrt{2}\right)\left(1+\sqrt{2}\right)^n &= \sigma\alpha^n \\ u_n &= u_0x_n + 2v_0y_n \\ v_n &= u_0y_n + v_0x_n.\end{aligned}$$

Also,

$$\left|u_n - v_n\sqrt{2}\right| = \left|u_0 - v_0\sqrt{2}\right| \left|\left(1-\sqrt{2}\right)^n\right| = g\sigma^{-1}\alpha^{-n}$$

$$\left\{|u_n|, \left|v_n\sqrt{2}\right|\right\} = \left\{\left|\left(u_n + v_n\sqrt{2}\right) \pm \left(u_n - v_n\sqrt{2}\right)\right|\right\} = \left|\frac{\sigma\alpha^n \pm g\sigma^{-1}\alpha^{-n}}{2}\right|$$

$$u_n^2 + 2v_n^2 = \frac{\sigma^2\alpha^{2n} + g^2\sigma^{-2}\alpha^{-2n}}{2}.$$

We call this sequence of cores a *family*. All the cores in a family will have the same gap $\left|u_n^2 - 2v_n^2\right|$. Each core belongs to a unique family whose starter can be found by iterating $(u_{n-1}, v_{n-1}) = (2v_n - u_n, v_n - u_n)$ until a starter results. Here are (the) two families with gap 89:

n	u_n	v_n	u'_n	v'_n
-2	19	-15	17	-10
-1	-11	4	-3	7
0	3	7	11	4
1	17	10	19	15
2	37	27	49	34
3	91	64	117	83

One starter is $3 + 7\sqrt{2} \Leftrightarrow [9, 98]$ and the paired starter is $11 + 4\sqrt{2} \Leftrightarrow [32, 121]$. Note that

$$\left(\left|u'_{n-1}\right|, \left|v'_{n-1}\right|\right) = \left(|u_{-n}|, |v_{-n}|\right).$$

This can be explained by recalling that

$$\left\{|u_n|, \left|v_n\sqrt{2}\right|\right\} = \left|\frac{\sigma\alpha^n \pm g\sigma^{-1}\alpha^{-n}}{2}\right|$$

$$\left\{|u'_n|, \left|v'_n\sqrt{2}\right|\right\} = \left|\frac{\sigma'\alpha^n \pm g\sigma'^{-1}\alpha^{-n}}{2}\right|$$

Where

$$\sigma' = \frac{g\alpha}{\sigma}.$$

An integer $g > 1$ occurs as a gap exactly when all of its prime divisors are of the form $8k \pm 1$. The number of families corresponding to this gap is 2^d where d is the number of distinct prime divisors. If desired, it is not too hard to use the factorization routine over $\mathbb{Z}\left[\sqrt{2}\right]$ of MAPLE to find the starters corresponding to a given gap.

Claim 4. *Let $\sigma = u + v\sqrt{2}$ be a starter with gap $g = \left|u^2 - 2v^2\right| > 1$. Then the ratio $\rho = \frac{g}{\sigma^2}$ satisfies $1 > \rho > 3 - 2\sqrt{2} = \alpha^{-2} \approx .171\,57$*

Proof. Let $\rho' = \frac{g}{\sigma'^2}$ where $\sigma' = u' + v'\sqrt{2} = \frac{g\alpha}{\sigma}$ is the paired starter. Surely $\rho = \frac{g/\sigma}{\sigma} = \frac{|u-v\sqrt{2}|}{u+v\sqrt{2}} < 1$ and similarly $\rho' < 1$. We finish by noting that $\rho\rho' = \frac{g}{\sigma^2}\frac{g}{\sigma'^2} = \frac{1}{\alpha^2}$ since $\sigma'\sigma = g\alpha$. □

We mention, but do not prove or use, these sharper estimates

$$1 - \frac{2\sqrt{2}}{\sigma} \geq \frac{g}{\sigma^2} \geq \left(3 - 2\sqrt{2}\right)\left(1 + \frac{2\sqrt{2}\alpha}{\sigma}\right).$$

Any potential solution of the critical interval CI_N has the form

$$2s_N^2 < MU^2, 2MV^2 < 2\left(s_N + 1\right)^2.$$

Here

$$U + V\sqrt{2} = \left(u_k + v_k\sqrt{2}\right) = \left(u + v\sqrt{2}\right)\left(1 + \sqrt{2}\right)^k$$

for some starter $\left(u + v\sqrt{2}\right)$ and $k > 0$. Thus,

$$2s_N^2 < M\left(u_k\right)^2, 2M\left(v_k\right)^2 < 2\left(s_N + 1\right)^2.$$

We call this a *type j solution* where $j = k - N$. We will describe it with the code

$$N, u + v\sqrt{2}, j, m$$

where m is the narrow multiplier. It turns out that $j = 0$ is by far the most common case.

Given a solution $N, u + v\sqrt{2}, j, m$ we set $n = N - j$ so $k = N + j = n + 2j$. Hence for a type j solution we are considering possible integer values of M such that:

$$2s_{n+j}^2 < M\left(u_{n+2j}\right)^2, 2M\left(v_{n+2j}\right)^2 < 2\left(s_{n+j} + 1\right)^2$$

Since we have exact expressions for all these quantities we can given exact expressions for the (perhaps empty) interval of *rational* values M^* such that

$$2s_{n+j}^2 < M^*\left(u_{n+2j}\right)^2, 2M^*\left(v_{n+2j}\right)^2 < 2\left(s_{n+j} + 1\right)^2.$$

Putting these expressions into useful form and getting useful information is somewhat more work.

Consider a fixed starter $\sigma = u + v\sqrt{2}$. For brevity we also define

$$\begin{aligned}
\tau &= \overline{\sigma} = u - v\sqrt{2} \\
\rho &= \frac{|\tau|}{\sigma} \\
g &= \left|u^2 - 2v^2\right| = |\sigma\tau| = \sigma^2\rho \\
f_1 &= u^2 + 2v^2 = \frac{\sigma^2 + \tau^2}{2} = \sigma^2\frac{1 + \rho^2}{2} \\
f_2 &= -4uv = \frac{-\left(\sigma^2 - \tau^2\right)}{\sqrt{2}} = -\sigma^2\frac{1 - \rho^2}{\sqrt{2}}.
\end{aligned}$$

Claim 5. *For the fixed starter* $\sigma = u + v\sqrt{2}$, *the* exact interval *of rational values* M^* *such that*

$$2s_{n+j}^2 < M^* \left(u_{n+2j}\right)^2, 2M^* \left(v_{n+2j}\right)^2 < 2\left(s_{n+j}+1\right)^2$$

(i) *has center at* $C_n + O\left(\alpha^{-2n}\right)$ *where*

$$C_n = \frac{x_{2n+2}f_1 + y_{2n+2}f_2}{2g^2} = \frac{\alpha^{2n+2} + \alpha^{-2n-2}\rho^{-2}}{4\sigma^2}$$

(ii) *and radius* $r_j + O\left(\alpha^{-4n}\right)$ *where*

$$r_j = \frac{1}{\sigma^2\alpha^{2j-1}}\left(\sqrt{2} - \frac{g}{2\alpha^{2j-1}\sigma^2}\right) = \frac{1}{\sigma^2\alpha^{2j-1}}\left(\sqrt{2} - \frac{\rho}{2\alpha^{2j-1}}\right).$$

Also, throughout this interval

$$\sqrt{M^*}\sigma = \frac{\alpha^{n+1}}{2} + O\left(\alpha^{-2n}\right) = x_{n+1} + O\left(\alpha^{-n}\right).$$

In contrast to the exact interval, we call $(C_n - r_j, C_n + r_j)$ the *estimated interval*. We take as a

Guiding Principle. The error terms in the claim are so small that we can discard them and still detect all valid solutions while rejecting all non-solutions.

This is well supported by experience but not rigorously established. There are strong results bounding the errors. We present only some of them here.

We wish to determine if there is an integer in the exact interval. Using the Guiding Principle we instead examine the estimated interval. This gives a nice advantage. Note that C_n is a rational number with odd numerator. To determine if there is some integer in this estimated interval it is enough to find the remainder of the numerator modulo $2g^2$.

In discarding the error terms there are two potential types of error. We might incorrectly include an integer which is not in the interval or exclude one which is. However any presumed solution can be checked with exact large integer arithmetic and near misses can be noted and checked similarly. The Guiding Principle is supported by the fact that no errors of either type showed up in extensive calculations including an independent exact check for the first 25 critical intervals. In practice the errors are exceedingly small even for small n and decrease rapidly. A typical example is:

Example 6. The starter $25 + 12\sqrt{2}$ has gap $g = 25^2 - 2 \cdot 12^2 = 337$ and provides solutions for intervals $6, 7$ and 8 all with broad multiplier $M = 189$ (so the narrow

multiplier is $m = 21$).

critical interval	type	estimated center $\pm$ radius for broad multiplier	center error	radius error
6	-1	$\left(189 - \frac{41}{2g^2}\right) \pm \frac{123.808}{2g^2}$	$\frac{5.49\times10^{-3}}{2g^2}$	$\frac{1.63\times10^{-7}}{2g^2}$
7	0	$\left(189 - \frac{41}{2g^2}\right) \pm \frac{368.353}{2g^2}$	$\frac{1.025\times10^{-4}}{2g^2}$	$\frac{9.07\times10^{-11}}{2g^2}$
8	1	$\left(189 - \frac{41}{2g^2}\right) \pm \frac{73.417}{2g^2}$	$\frac{6.23\times10^{-4}}{2g^2}$	$\frac{5.36\times10^{-12}}{2g^2}$

Note that $x_8 = 577$ and $\sqrt{189}\left(25 + 12\sqrt{2}\right) = 576.999\,842$. In all these solutions the error in the estimated center and radius for the interval of rational multipliers is insignificant for the conclusion that the interval contains an integer. In each case we estimate the center to be at $189 - \frac{41}{227\,138} = 189 - \frac{41}{2g^2}$. The range of broad rational multipliers for a solution $9, 25 + 12\sqrt{2}, 2, m$ is about $\left(189 - \frac{41}{2g^2}\right) \pm \frac{12.89}{2g^2}$ and any small error terms do not alter the conclusion that there are no integer multipliers in this range.

Consider using the same starter but increasing N. The estimated radii r_j are the same and the estimated centers C_N are always rational numbers with odd numerator and denominator $2g^2$. But each time we increase from N to $N + 1$ the already minuscule error terms decrease by an order of magnitude. This is further illustrated by the tables of center and radius errors in section 5. In some cases we will give a short justification of some result using the Guiding Principle and elsewhere give a more delicate and unconditional proof. An example is the remark following this:

Lemma 7. *For a fixed starter* $\sigma = u + v\sqrt{2}$

(i) $r_j < 0$ *for* $j < -1$;

(ii) $r_{-1} < 0$ *unless* $\rho < 20 - 14\sqrt{2} \approx .201\,010\,126\,8$;

(iii) $r_{-1} < r_0$;

(iv) $r_0 > r_1 > \cdots > 0$;

(v) $\frac{1}{2g^2} > \frac{1}{\alpha g^2} > r_j$ *for* $j > 1 + \frac{\log_\alpha g}{2}$.

Proof. Recall that

$$\begin{aligned} r_j &= \frac{1}{\sigma^2\alpha^{2j-1}}\left(\sqrt{2} - \frac{g}{2\alpha^{2j-1}\sigma^2}\right) \\ &= \frac{1}{2\sigma^2\alpha^{2j-1}}\left(2\sqrt{2} - \frac{\rho}{\alpha^{2j-1}}\right) \end{aligned}$$

and that the ratio $\rho = \frac{g}{\sigma^2}$ satisfies $\alpha^{-2} < \rho < 1$. If $j < -1$ then

$$2\sqrt{2} - \frac{\rho}{\alpha^{2j-1}} \leq 2\sqrt{2} - \alpha^5\rho < 2\sqrt{2} - \alpha^5\left(\alpha^{-2}\right) = -\left(7 + 3\sqrt{2}\right).$$

For $j = -1$

$$2\sqrt{2} - \frac{\rho}{\alpha^{2j-1}} = 2\sqrt{2} - \alpha^3\rho = \alpha^3\left(20 - 14\sqrt{2} - \rho\right).$$

For any j,

$$\begin{aligned} r_{j+1} - r_j &= \frac{1}{2\sigma^2\alpha^{2j+1}}\left(2\sqrt{2} - \frac{\rho}{\alpha^{2j+1}}\right) - \frac{1}{\sigma^2\alpha^{2j-1}}\left(2\sqrt{2} - \frac{\rho}{\alpha^{2j-1}}\right) \\ &= \frac{1}{2\sigma^2\alpha^{2j+1}}\left(\left(1-\alpha^2\right)2\sqrt{2} - \frac{\rho}{\alpha^{2j+1}}\left(1-\alpha^4\right)\right) \\ &= \frac{\left(1-\alpha^2\right)}{2\sigma^2\alpha^{2j+1}}\left(2\sqrt{2} - \frac{\rho\left(1+\alpha^2\right)}{\alpha^{2j+1}}\right). \end{aligned}$$

Thus when $j = -1$

$$r_0 - r_{-1} = \frac{\alpha\left(1-\alpha^2\right)}{2\sigma^2}\left(2\sqrt{2} - \rho\alpha\left(1+\alpha^2\right)\right)$$

and hence $r_0 > r_{-1}$ since

$$2\sqrt{2} - \rho\alpha\left(1+\alpha^2\right) < 2\sqrt{2} - \alpha^{-1}\left(1+\alpha^2\right) = 0.$$

For $j \geq 0$ we have $r_{j+1} < r_j$ since

$$2\sqrt{2} - \frac{\rho\left(1+\alpha^2\right)}{\alpha^{2j+1}} > 2\sqrt{2} - \frac{\left(1+\alpha^2\right)}{\alpha^{2j+1}} \geq 2\sqrt{2} - \frac{\left(1+\alpha^2\right)}{\alpha} = 0.$$

Finally, if $j > 1 + \frac{\log_\alpha g}{2}$ then $\alpha^{2j-2} > g$ so

$$\begin{aligned} r_j &= \frac{1}{2\sigma^2\alpha^{2j-1}}\left(2\sqrt{2} - \frac{\rho}{\alpha^{2j-1}}\right) < \frac{\sqrt{2}}{\sigma^2\alpha^{2j-1}} \\ &= \frac{\rho}{\alpha}\frac{\sqrt{2}}{g\alpha^{2j-2}} < \frac{1}{\alpha}\frac{\sqrt{2}}{g\alpha^{2j-2}} < \frac{1}{\alpha g^2}. \end{aligned}$$

□

Remark 8. This exact result about the estimated radii r_j can be combined with the Guiding Principle to conclude that, as in the example above, every solution $N, u + v\sqrt{2}, j, m$ of type $j > 0$ is connected in a string of related solutions, all with the same multiplier, to a solution of type 0.

$$N - i, u + v\sqrt{2}, j - i, m \text{ for } 0 \leq i \leq j.$$

Similarly, a solution $N, u + v\sqrt{2}, -1, m$ is always accompanied by a solution $N + 1, u + v\sqrt{2}, 0, m$.

3. Algorithms

All computations were done with MAPLE.

3.1. Multiplier Searches

The algorithm used is the obvious one. It uses the relatively fast integer quotient and integer square root operations. Naive but plausible arguments predict that a narrow multiplier m will provide solutions for about $\frac{\sqrt{2}}{m}$ of the intervals ($\frac{1}{\sqrt{m}}$ for y and $\frac{\sqrt{2}}{\sqrt{m}}$ for x). This agrees well with our computations although the intervals solved are not well distributed. The phenomenon of a string of consecutive solutions all with the same multiplier, as in the example above with $m = 21$ and $M = 189$, can cause clustering followed by relatively long empty gaps.

3.2. Starter Searches

A straightforward algorithm using large integer arithmetic is easy to give. Here is a version of the modular starter algorithm. Given a starter $u_0 + v_0\sqrt{2}$ it will find all solutions involving cores $u_k + v_k\sqrt{2}$ from the associated family. The version given here runs from 3 to some top n. It is not hard to alter it to simply check a specified interval. The running time and expected frequency of solutions are somewhat more complicated to describe than for the multiplier method.

- Compute the *gap* $g = \left|u_0^2 - 2v_0^2\right|$, the *magnitude* $\sigma = u_0 + v_0\sqrt{2}$ and the constants $f_1 \equiv \left(u_0^2 + 2v_0^2\right) \pmod{2g^2}$ and $f_2 \equiv -4u_0v_0\sqrt{2} \pmod{2g^2}$.

- Define the positive integers

$$H_j = \left\lfloor 2g^2 r_j \right\rfloor_{\text{odd}} = \left\lfloor 2g^2 \left(\frac{\alpha\sqrt{2}}{(\alpha^j\sigma)^2} - \frac{\alpha^2 g}{2\left(\alpha^j\sigma\right)^4} \right) \right\rfloor_{\text{odd}} \quad \text{for } j = -1, 0, 1, 2, \ldots, j_{\text{top}}$$

  Here $\lfloor\cdot\rfloor_{\text{odd}}$ means truncate to an odd integer. We only consider H_{-1} when it might be positive (Namely if $\rho > 20 - 14\sqrt{2}$). $H_0 \geq H_1 > H_2 > \cdots > H_{j_{top}}$. Here j_{top} is the largest j with $H_j > 0$.

- For $n \geq 3$ consider the estimated center $C_n = \frac{x_{2n+2}f_1 + y_{2n+2}f_2}{2g^2}$ and compute $c_n = x_{2n+2}f_1 + y_{2n+2}f_2$ modulo $2g^2$ (note that $2g^2$ and f_2 are even while x_n, f_1 and c_n are odd).

  1. If $\min\left(c_n, 2g^2 - c_n\right) \leq H_{-1}$ then $u_{n-2} + v_{n-2}\sqrt{2}$ gives a solution for CI_{n-1}
  2. If $\min\left(c_n, 2g^2 - c_n\right) \leq H_0$ then $u_n + v_n\sqrt{2}$ gives a solution for CI_n
  3. If $\min\left(c_n, 2g^2 - c_n\right) \leq H_1$ then $u_{n+2} + v_{n+2}\sqrt{2}$ gives a solution for CI_{n+1}
  4. etc.

If $\min\left(c_n, 2g^2 - c_n\right) < H_j$ then $u_{n+2j} + v_{n+2j}\sqrt{2}$ gives a solution for I_{n+j}. we called this a *type j solution*. After step 1) we stop the first time we get failure, although we may save near misses for future examination. All the solutions which occur before failure will have the same multiplier.

- Check all solutions and near misses for possible errors of inclusion or omission.

A naive exact integer routine is faster for checking, say, the first 25 intervals. Also, there are no round off concerns. The given algorithm uses integer arithmetic modulo $2g^2$ and is effective for, say, the 10^6th critical interval (modulo some attention to the error terms). One might make two passes, first looking only for type 0 solutions and then returning to check each one found for related solutions of type $j \neq 0$. This avoids computing H_1 etc. for starters which give no solutions in the range under consideration. The numerators satisfy the recurrence

$$c_{n+1} \equiv 6c_n - c_{n-1} \pmod{2g^2}.$$

Thus each can be computed from the previous two without a computation of x_{2n+2} and y_{2n+2}.

The proportion of intervals having a solution arising from a fixed starter $\sigma = u_0 + v_0\sqrt{2}$, is, as we have said, a fraction c_σ with denominator P_g. It is plausible (in a sense which can be made precise) that (on the average, when σ is not too small) c_σ is about $\frac{1}{\sigma}$ and that there about $\frac{2\sqrt{2}}{\pi^2}s$ starters with $\left|u_0 + v_0\sqrt{2} - s\right| < \frac{1}{2}$.

4. The First 25 and the First 20,000 Critical Intervals

To find all the solutions for CI_3 through CI_{25} we first find all the type 0 solutions for each interval then go back and check which ones give solutions of type $j = -1$ or $j > 0$ for adjacent intervals. Every type 0 solution for CI_n will have a starter $\sigma = u + v\sqrt{2}$ and broad multiplier M with $\sigma\sqrt{M}$ almost exactly x_{n+1}. Hence it will suffice to do a multiplier search of odd squarefree numbers $m \leq m_{\text{top}}$ and a starter search for starters $\sigma = u + v\sqrt{2}$ (with $\gcd(u, 2v) = 1$) such that $|\sigma| \leq \sigma_{\text{top}}$ where the upper bounds are chosen so that $m_{\text{top}}\sigma_{\text{top}} > x_{n+1} + 1$. In this range it turns out to be faster (in MAPLE) to do the starter search using exact large integer arithmetic. (This avoids any concerns about roundoff error as well.)

This approach does not account for the type -1 solutions of the top interval. We deal with this by considering the starters which are larger than $\sigma_{\max}(25)$ and have ρ small enough to give such a solution and are small enough that the multiplier search could miss them (it took 20 hours to show that there were none).

We used the following procedure to find some solutions for each critical interval from 3 up to 20,000

- Use the simple (but carefully written) exact integer MAPLE algorithm to find all solutions with narrow multiplier $m < m_{\text{top}}(n)$. This gets harder as n increases so $m_{\text{top}}(n)$ was smaller for large n than for small.
- Attempt to carry the multiplier search further if needed in order to find the smallest multipliers.
- Find all solutions with starter less than 200.
- Attempt to find the smallest starter among the solutions (if it is larger).
- Continue both searches far enough to find at least one solution.

The results of the modular computations were verified with exact integer arithmetic. No near misses were found to be solutions and no incorrect solutions were offered.

For critical intervals 3 to 20,000 all have either a minimal multiplier below 1000 or a solution with starter $u_0 + v_0\sqrt{2}$ such that $\max(u_0, v_0\sqrt{2}) < 70$ EXCEPT:

interval	starters	mmin
639	$[511, 69]$	1963
9624	$[45, 127], [279, 37], [695, 69], [525, 557]$	?
16677	$[71, 35], [161, 73], [211, 94], [455, 227]$	?
16835	$[25, 74], [229, 38], [1, 331], [83, 483]$	?

(1)

Considering starters with $906 \geq v > u$ or $1283 \geq u > 2v$, solutions were found for all intervals except:

Interval	2386	2537	7120	7430	10047	14297	18063	19993
Multiplier	7	5	3	5	91	413	3	215

(2)

We also know *all* multipliers up to 30000 for the first 137 critical intervals. The smallest known multipliers for critical interval 590 are $1294449, 4977589$ and 16073971.

5. Periodic Solutions

This section is concerned with results like the following which provides solutions for 10 out of every 21 critical intervals.

Claim 9. *Let σ be one of the two starters $1 + 2\sqrt{2}$ and $3 + \sqrt{2}$ for gap 7. Then σ gives a type j solution of CI_N exactly if $N \geq 1$ and*

(i) $\sigma = 1 + 2\sqrt{2}$, $j = 0$ *and* $N \pmod{21} \in \{0, 7, 8, 10, 15, 20\}$

(ii) $\sigma = 1 + 2\sqrt{2}$, $j = 1$ *and* $N \pmod{21} \in \{11, 16\}$

(iii) $\sigma = 3 + \sqrt{2}$, $j = 0$ *and* $N \pmod{21} \in \{0, 5, 10, 12, 20\}$

(iv) $\sigma = 3 + \sqrt{2}$, $j = 1$ *and* $N \pmod{21} = 11$

More generally,

Theorem 10. *For a fixed starter σ with gap g there is an integer $N_0 = N_0(\sigma)$ and a period $P = P_g$ such that for all $N \geq N_0$ the truth of the statement "σ gives a type j solution of CI_N " depends only on j and the congruence class of N modulo P. The period P, initial value N_0, and appropriate congruence classes can all be determined.*

It follows from the **Guiding Principle** that $N_0(\sigma) = 3$ for all σ. We first discuss the general situation. Then we give a careful treatment of the two starters with gap 7 followed by results for other starters.

Given a fixed starter $\sigma = u_0 + v_0\sqrt{2}$ the modular starter algorithm first computes several radii $\frac{H_j}{2g^2}$ with H_j an odd integer and the constants $f_1 = u_0^2 + 2v_0^2$ and $f_2 = -4u_0v_0$. It then decides if $u_{n+2j} + v_{n+2j}\sqrt{2}$ gives a type j solution for critical interval CI_{n+j} by considering the odd integer $x_{2n+2}f_1 + y_{2n+2}f_2$ and checking if the distance from $\frac{x_{2n+2}f_1 + y_{2n+2}f_2}{2g^2}$ to the nearest integer is $\leq \frac{H_j}{2g^2}$. This distance depends only on the congruence class of $x_{2n+2}f_1 + y_{2n+2}f_2$ modulo $2g^2$. This in turn depends only on the congruence classes of x_{2n+2} and y_{2n+2}. There are only $\left(2g^2\right)^2$ possible values for (x_k, y_k) modulo $2g^2$ since each pair determines the next:

$$(x_n, y_n) = \begin{cases} (1,0) & n = 0 \\ (x_{n-1} + 2y_{n-1}, x_{n-1} + y_{n-1}) & n > 0 \end{cases}$$

there is some period $P = P_g \leq 4g^4$ such that for all n

$$(x_{2n+2P+2}, y_{2n+2P+2}) \equiv (x_n, y_n) \pmod{2g^2}.$$

We may take P_g to be the smallest positive integer with

$$(x_{2P}, y_{2P}) \equiv (1, 0) \pmod{2g^2}.$$

Then, with one restriction, the truth of "$u_{n+2j} + v_{n+2j}\sqrt{2}$ gives a type j solution for critical interval CI_{n+j}" depends only on j and the congruence class of n modulo P. The restriction is that the $O\left(\alpha^{-2n}\right)$ errors in approximating radii and centers do not cause an error of either type. This is why the restriction $N = n + 2j > N_0$ is included. The Guiding Principle is the claim that the errors are insignificant for $N \geq 3$.

Actually, P_g is much less than $4g^4$. In fact P_g is either equal to or a divisor of $g\prod \frac{p-1}{2}$ where the product is over the prime divisors of g. Equality is fairly common.

We will summarize the results of claim 7 for the two starters with $g = 7$ as follows:

	7	21
$1+2\sqrt{2}$		$0, 7, 8, 10, 11^1, 15, 16^1, 20$
$3+\sqrt{2}$		$0, 5, 10, 11^1, 12, 20$

With this notation we have

	17	136
$5+2\sqrt{2}$		$0, 9, 13, 15, 16^1, 42, 60,$
		$68, 77, 81, 83, 84^1, 110, 128$
$1+3\sqrt{2}$		$0, 7, 25, 26^1, 52, 53^1, \mathbf{54}^2, \mathbf{54}, 57, 58, 67,$
		$68, 75, 93, 94^1, 120, 121^1, \mathbf{122}^2, \mathbf{122}, 125, 126, 135$

	23	253
$3+4\sqrt{2}$		$0, 8, 55, 69, 94, 95^1, 96^2, 107, 111, 126,$
		$127, 147, 148^1, 174, 179, 200, 208, 209^1, 227, 240, 243, 249$
$5+\sqrt{2}$		$0, 3, 4^1, 9, 12, 25, 44, 45^1, 52, , 53^1, 73, 78, 79^1, 105, 106^1, 116, 126,$
		$127^1, 140, 141, 145, 158, 159^1, 160^2, 183, 191, 197, 203, 244, 252$

	31	15
$7+3\sqrt{2}$		$0, 7$
$1+4\sqrt{2}$		$6, 7, 8^1, 14$

	41	205
$7+2\sqrt{2}$		$0, 3, 53, 61, 81, 102, 103^1, 104^2, 114, 125, 204$
$1+4\sqrt{2}$		$0, 79, 90, 102, 103^1, 104^2, 123, 143, 151, 201, 204$

	217	105
$17+6\sqrt{2}$		0
$11+13\sqrt{2}$		$0, 52$
$15+2\sqrt{2}$		$34, 51, 52, 53^1, 104$
$5+11\sqrt{2}$		$0, 104$

A few remarks are in order.

1. In the details above for $g = 7$ we see that $\left(1+\sqrt{2}\right)^6 \equiv 1 \pmod{2\times 7}$ while $\left(1+\sqrt{2}\right)^{42} \equiv 1 \pmod{2\times 7^2}$. This is typical of many primes g where $P_g = \frac{g(g-1)}{2}$.
2. Note that P_{41} is only 205. This is related to the fact that $x_5 = 41$.
3. The period $P_{31} = 15$ is surprisingly small. This is because of a very special property of 31. As above $\left(1+\sqrt{2}\right)^{30} \equiv 1 \pmod{2\times 31}$ and this is the smallest

power which works. However, it is also true that $\left(1+\sqrt{2}\right)^{30} \equiv 1 \pmod{2\times 31^2}$.

$$\begin{aligned}(x_{30}, y_{30}) &= (152\,139\,002\,499, 107\,578\,520\,350)\\ &= \left(3^2 11 \times 19 \times 59 \times 601 \times 2281, 2 \times 5^2 7 \times 29 \times 31^2 41 \times 269\right)\end{aligned}$$

At least up to 100000, no other prime has this property that the first $k > 0$ such that $(x_k, y_k) \equiv (1,0) \pmod{p}$ is also the first such that $(x_k, y_k) \equiv (1,0) \pmod{p^2}$.

4. The starter $1+3\sqrt{2}$ for $g = 17$ gives two solutions for CI_{54}. One of type 0 and one of type 2.

5. From the table above it seems as though P_{17} should be 68. There is actually a central symmetry which occurs when P_g is even (as for $g = 73, 89, 97$).

6. This chart shows how many out of each 105 consecutive intervals have a solution with the indicated gap. The third column shows how many of these do not have a solution from a previous row. Thus the four starters for $g = 217$ each have period 105 and between them solve 6 of every 105 intervals. Of these 6 all but one has a solution with gap 7 or 31. This accounts for 69 of every 105 critical intervals.

gap	solved	new
7	50	50
31	35	18
217	6	1

7. Here is the same information for $7, 31, 17$ and their products (in that order). The numbers are now out of every 14280 consecutive intervals

gap	solved	new
7	68000	6800
31	4760	2448
17	3360	1184
119	1045	278
217	816	95
527	294	72
3689	55	12

So gaps 7, 31 and 17 account for just over 73% of the intervals and this increases to just over 76.25% if we include the other four gaps with period dividing 14280.

8. It does not seem promising to cover all the integers by such arithmetic progressions. There are critical intervals which have no solutions with small starters. For example critical interval 639 has no starters smaller than $511 + 69\sqrt{2}$

with gap $g = 511^2 - 2 \cdot 69^2 = 251\,599 = 311 \times 809$ and period $P_{251599} = 3938782345 = 15\,655 \times 251\,599 = 5 \times 31 \times 101 \times 311 \times 809$.

References

[1] E. Croot, A. Granville, R. Pemantle, and P. Tetali, On sharp transitions in making squares, Ann. of Math., to appear.

[2] R. Crandall and C. Pomerance, Prime Numbers: A Computational Perspective, 2nd ed., Springer, New York, 2005.

[3] P. Erdős, J.L. Malouf, J.L. Selfridge and E. Szekeres, Subsets of an interval whose product is a power, Disc. Math. **200** (1999), 137-147