Searching for large Twin Smooth Integers

using solutions to the Prouhet-Tarry-Escott problem

Knud Ahrens

Faculty of Computer Science and Mathematics
University of Passau, Germany

ALCOCRYPT 2023
23^{rd} of February

Outline

Twin Smooth Primes

Definition

An integer $n \in \mathbb{N}$ is called *twin smooth* if $n \pm 1$ both are smooth.

- SQISign needs twin smooth primes.
- Larger primes improve security.
- "Smoother" primes improve efficiency.

The Prouhet-Tarry-Escott problem

Prouhet-Tarry-Escott (PTE) problem of size n

$$a_1^j + \cdots + a_n^j = b_1^j + \cdots + b_n^j \text{ with } \{a_1, \ldots, a_n\} \cap \{b_1, \ldots, b_n\} = \emptyset$$

for all $1 \leq j \leq k$ and $a_i, b_i \in \mathbb{Z}$.

A solution with $k = n - 1$ is called ideal and the corresponding polynomials

$$a(x) = \prod_{i=1}^{n}(x - a_i), \quad b(x) = \prod_{i=1}^{n}(x - b_i)$$

only differ by a constant $C \in \mathbb{Z}$ (Borwein & Ingalls, 1994).

Using PTE solutions

Since $|2a(\ell)/C - 2b(\ell)/C| = 2$ we can check:

smoothness	$a(\ell)$ and $b(\ell)$ are smooth
C-integrality	$a(\ell) \equiv b(\ell) \equiv 0 \mod C$ or $a(\ell)/C, b(\ell)/C \in \mathbb{Z}$
primality	$a(\ell)/C + b(\ell)/C$ is prime

Costello, Meyer & Naehrig (2021) tested the smoothness of many different PTE solutions at the same time using shared factors.

Motivation

Parameters of the PTE solutions:

- Increasing size n allows to decrease the smoothness bound.
- Smaller difference C gives larger number of multiples of C.
- There are only few known ideal PTE solutions with $n > 6$ and $\log_2 C < 60$. (Costello *et al.*, 2021)

Check C-integrality first and have fewer smoothness tests.

For 512 bit primes PTE solutions with $n > 6$ give better results.

C-integrality

Let $R = \{r \in \mathbb{Z} \mid a(r) \equiv b(r) \equiv 0 \mod C, 0 \leq r < C\}$ then

- only about $|R|/C$ of the elements in $a(\mathbb{Z})$ and $b(\mathbb{Z})$ are C-integral.
- the C-integral elements are $a(r + kC)$ and $b(r + kC)$ with $r \in R$ and $k \in \mathbb{Z}$.

Finding R is a one-time pre-computation.

We will see that $|R|/C \to 0$ for $C \to \infty$.

Smoothness

To check smoothness of $a(\ell)$ and $b(\ell)$ we do the following:

- Create a look-up table using a sieve of Eratosthenes.
- Look up the factors $(\ell - a_i)$ and $(\ell - b_i)$.

For our purposes using a polynomial sieve or the smoothness test of Bernstein (2004) is slower than using this look-up table.

Implementation

For $\log_2 C > 32$ we need 128 bit variables.

Standard floating point arithmetic is not precise enough for our needs.

Implementations in Sage and C can be found at:
`https://git.fim.uni-passau.de/ahrens/twin-smooth-integers`

Results

For $n = 7$:

- Only 10 ideal PTE solutions with $\log_2 C \leq 60$ are known.
- Our tests took less than $1/20$ of the time of the naive approach.

For $n \geq 9$:

- Per n at most two ideal PTE solutions with small C are known.
- Heuristically, our algorithm is at least twice as fast as the naive approach.

Counting roots modulo C

Let $R = \{r \in \mathbb{Z}/C\mathbb{Z} \mid f(r) = 0\}$ for $C \in \mathbb{N}_{>0}$ and f of the form

$$f(x) = \prod_{i=1}^{n}(x - z_i)$$

where $z_i \in \mathbb{Z}$ are known.

How many solutions $|R|$ are there and how large is the density $|R|/C$?

Notation

All solutions of $f(x) \equiv 0 \bmod p^e$ are also solutions modulo p.

Let $m_e(z) = |\{z_i \mid z \equiv z_i \bmod p^e\}|$ then

- $1 \leq m_e(z) \leq n$
- $m_i(z) \leq m_j(z)$ for $i > j$
- $m_0(z) = n$

Example

Let $g(x) = (x-0)(x-2)(x-4)$ with $z = 2$ and $p = 2$.
Then $m_0(2) = 3, m_1(2) = 3, m_2(2) = 1$.

$$g\big(z + \sum_{j\geq 1} d_j p^j\big) = p^3\big(0 + \sum_{j\geq 1} d_j p^{j-1}\big)\big(1 + \sum_{j\geq 1} d_j p^{j-1}\big)\big(-1 + \sum_{j\geq 1} d_j p^{j-1}\big)$$
$$\equiv p^3\big(d_1\big)\big(1 + d_1\big)\big(-1 + d_1\big) \mod p^4$$

$z + d_1 p = 2 \pm 2$ for $d_1 \neq 0$.

Set $d_1 = 0$ for solutions above 2 modulo p^e with $e > m_1(2) = 3$.

Induction

Assume $d_1, \ldots, d_{k-1} = 0$ for all solutions above z modulo p^e with $e > \sum_{i=1}^{k-1} m_i(z)$.
Look at $g(z + \sum_{j \geq k} d_j p^j)$ and set $d_k = 0$ for all solutions modulo p^e with $e > \sum_{\ell=1}^{k} m_\ell(z)$ with the same argument as in the example.

By induction we can set $d_k = 0$ for all solutions modulo p^e with $e > \sum_{\ell=1}^{k} m_\ell(z)$ for all $k \in \mathbb{N}$.

Number of Solutions

Let $\sum_{\ell=1}^{j-1} m_\ell(z) < e \leq \sum_{\ell=1}^{j} m_\ell(z)$ then there are p^{e-j} solutions for $f(x) \equiv 0 \bmod p^e$ and they are of the form $x = z + \sum_{i=j}^{e-1} d_i p^i$.

The most roots above z exist if all z_i are equal. Then $\sum_{\ell=1}^{j} m_\ell(z) = nj$ and the number of solutions above z modulo p^e is $p^{e-\lceil \frac{e}{n} \rceil} \leq p^{e\frac{n-1}{n}}$.

The total number of roots is therefore strictly less than $np^{e\frac{n-1}{n}}$.

Bounding $|R|$ and $|R|/C$

Let $C = \prod_{j=1}^{t} p_j^{e_j}$ be the prime factorisation of C.

Asymptotically $t \approx \log\log C$ and the number of solutions is

$$|R| < \prod_{j=1}^{t}\left(n p_j^{e_j \frac{n-1}{n}}\right) = n^t \left(\prod_{j=1}^{t} p_j^{e_j}\right)^{\frac{n-1}{n}} = n^t C^{\frac{n-1}{n}} \approx n^{\log\log C} C^{\frac{n-1}{n}}.$$

The density of solutions is therefore bounded by

$$\frac{|R|}{C} < \frac{n^t C^{\frac{n-1}{n}}}{C} = n^t C^{-\frac{1}{n}} \approx \frac{n^{\log\log C}}{\sqrt[n]{C}} \underset{C\to\infty}{\longrightarrow} 0.$$

Conclusion

A polynomial $f \in \mathbb{Z}[X]$ of the form $f(x) = \prod_{i=1}^{n}(x - z_i)$ with $z_i \in \mathbb{Z}$ has less than $np^{e\frac{n-1}{n}}$ solutions to $f(x) \equiv 0$ modulo p^e.

For $C = \prod_{j=1}^{t} p_j^{e_j}$ the number of solutions modulo C is bounded by $|R| < n^t C^{\frac{n-1}{n}}$ and the density $|R|/C$ asymptotically goes to zero for C tending towards infinity.

Our approach to optimise for one ideal PTE solution at a time is therefore (heuristically) faster than the naive or the multi-solution approach for $n = 7$ and $n \geq 9$. Hence, it is better suited to find large twin smooth primes.

Example

Let $g(x) = (x-0)(x-2)(x-4)$ and $p = 2$. Then the solutions to $g(x) \equiv 0 \bmod p^e$ are

$$
\begin{aligned}
e = 1 &: && 0 \\
e = 2 &: && 0, 2 \\
e = 3 &: && 0, 2, 4, 6 \\
e = 4 &: && 0, 4, 8, 12, && 2, 6, 10, 14 \\
e = 5 &: && 0, 4, 8, 12, 16, 20, 24, 28, && 2, 10, 18, 26 \\
e = 6 &: && 0, 8, 16, 24, 32, 40, 48, 56, && 4, 12, 20, 28, 36, 44, 52, 60, && 2, 18, 34, 50
\end{aligned}
$$

and their number is

$$
\begin{aligned}
2^{e-1} &\text{ for } 0 < e \leq 3, \\
2^{e-2} + 4 &\text{ for } 3 < e \leq 5, \qquad < 3 \cdot 2^{e\frac{3-1}{3}} \\
8 + 8 + 4 &\text{ for } e > 5.
\end{aligned}
$$

Example

Let $f(x) = (x-3)(x-3)(x-2)$ and $p = 3$. Then the solutions to $f(x) \equiv 0 \bmod p^e$ are

$$\begin{aligned}
e = 1 &: \quad 2, \quad 0 \\
e = 2 &: \quad 2, \quad 0, 3, 6 \\
e = 3 &: \quad 2, \quad 3, 12, 21 \\
e = 4 &: \quad 2, \quad 3, 12, 21, 30, 39, 48, 57, 66, 75 \\
e = 5 &: \quad 2, \quad 3, 30, 57, 84, 111, 138, 165, 192, 219
\end{aligned}$$

and their number is

$$1 + 3^{e - \lceil \frac{e}{2} \rceil} < 3 \cdot 3^{e \frac{3-1}{3}}.$$

p-adic Smoothness

The p-adic norm $|\cdot|_p$ is defined as $|x|_p = p^{-\nu_p(x)}$ for $x \neq 0$ and $|0|_p = 0$.

$$f(z+s) \equiv 0 \bmod p^e \Leftrightarrow \nu_p(f(z+s)) \geq e \quad \Leftrightarrow \quad |f(z+s)|_p \leq p^{-e}$$
$$d_1, \ldots, d_k = 0 \Leftrightarrow \nu_p(s) > k \quad \Leftrightarrow \quad |z-(z+s)|_p < p^{-k}$$

"How many solution above z are there modulo powers of p?"

"How smooth is f at z with respect to $|\cdot|_p$?"

References

Daniel J. Bernstein (2004). How to find Smooth Parts of Integers. `https://cr.yp.to/factorization/smoothparts-20040510.pdf`.

Peter Borwein & Colin Ingalls (1994). The Prouhet-Tarry-Escott problem revisited. *Enseign. Math. (2)* **40**(1-2), 3–27. ISSN 0013-8584.

Craig Costello, Michael Meyer & Michael Naehrig (2021). Sieving for Twin Smooth Integers with Solutions to the Prouhet-Tarry-Escott Problem. In *Advances in Cryptology – EUROCRYPT 2021*, Anne Canteaut & François-Xavier Standaert, editors, 272–301. Springer International Publishing, Cham. ISBN 978-3-030-77870-5.