A Chinese Prouhet–Tarry–Escott solution

David Broadhurst, The Open University, UK

Abstract: Jens Kruse Andersen recently set the challenge of finding complete factorizations of consecutive integers with more than 500 decimal digits. I was able to set records with up to 10 consecutive factorizations by using solutions of the ideal Prouhet–Tarry–Escott (PTE) problem, which is equivalent to finding polynomials with integer roots that differ only by an integer.

PTE solutions with degrees up to $s = 10$ were known by 1944, but the problem with $s > 10$ had received only one solution, found almost inadvertently in 1999.

It seemed to me that the ideal PTE problem might benefit from use of the Chinese remainder theorem. I shall describe how a new solution was found at degree $s = 12$ by the discipline of splitting the problem into a smart part that can be handled by Pari-GP and a brute force part that benefits from parsimonious ForTran programming.

I keep six honest serving-men
(They taught me all I knew);
Their names are What and Why and When
And How and Where and Who.

Rudyard Kipling (1865-1936)

1 Polynomials with integer roots differing by integers

Example 1 [G. Tarry, 1912]

$$A(x) \equiv (x^2-1)(x^2-9^2)(x^2-10^2) = (x^2-5^2)(x^2-6^2)(x^2-11^2) + C$$

with $C = A(\pm 5) = A(\pm 6) = A(\pm 11) = 2^6 \cdot 3^2 \cdot 5^2 \cdot 7$ is an "ideal symmetric Prouhet–Tarry–Escott" solution with even degree $s = 6$.

Example 2 [G. Tarry, 1913]

$$(x^2-5^2)(x^2-14^2)(x^2-23^2)(x^2-24^2) = (x^2-4)(x^2-16^2)(x^2-21^2)(x^2-25^2) + C$$

with $C = 2^8 \cdot 3^3 \cdot 5^2 \cdot 7^2 \cdot 11 \cdot 13$

Example 3 [A. Letac, before 1944]

$$a \equiv [12, 11881, 20231, 20885, 23738] \sim [436, 11857, 20449, 20667, 23750] \equiv b$$

$$A(x) \equiv \prod_{n=1}^{5} (x^2 - a_n^2) = B(x) + C; \quad B(x) \equiv \prod_{n=1}^{5} (x^2 - b_n^2)$$

$$C = 2^{11} \cdot 3^6 \cdot 5^2 \cdot 7^2 \cdot 11^2 \cdot 13^2 \cdot 17 \cdot 23 \cdot 37 \cdot 53 \cdot 61 \cdot 79 \cdot 83^2 \cdot 103 \cdot 107 \cdot 109^2 \cdot 113 \cdot 191$$

As explained in Sections 21.9 and 21.10 of Hardy and Wright, the problem of Prouhet and Tarry is to find a pair (a, b) of distinct sets of s integers with $\sum_{n=1}^{s}(a_n^h - b_n^h) = 0$ for all integer exponents h running from 1 to k.

There can be no solution with $k \geq s$. An **ideal** solution has $k = s - 1$ and leads to a pair of polynomials with integer roots and a difference

$$|C| = \left| \frac{\sum_{n=1}^{s}(a_n^s - b_n^s)}{s} \right| = 0 \bmod (s-1)!$$

A **symmetric** solution of even degree is one in which the roots of each polynomial occur as pairs of integers differing only in sign, as in the examples above. For odd degrees, we say that a solution is symmetric if $B(x) = -A(-x)$ as here:

Example 4 [E.B. Escott, 1913]

$A(x) = (x{+}50)(x{+}38)(x{+}13)(x{+}7)(x{-}24)(x{-}33)(x{-}51);\ A(x) = -A(-x){+}2A(0)$

By 1944 there were parametric formulas to generate any number of ideal solutions with degree s up to 8, and elliptic curve methods to generate solutions with $s = 9$ and $s = 10$, but until 1999 no ideal solution was known for $s > 10$.

Nuutti Kuosa discovered the following equality of sums of 10th powers on 3 September 1999, using an ESLP program written by Jean–Charles Meyrignac:

$$151^{10} + 140^{10} + 127^{10} + 86^{10} + 61^{10} + 22^{10} = 148^{10} + 146^{10} + 121^{10} + 94^{10} + 47^{10} + 35^{10}$$

Chen Shuwen noticed on 7 September 1999 that the same sets of integers have equal sums of 8th, 6th, 4th and 2nd powers and hence provide an ideal symmetric 12th-order PTE solution,

$$[35, 47, 94, 121, 146, 148] \sim [22, 61, 86, 127, 140, 151]$$

$$C = 2^{12} \cdot 3^9 \cdot 5^3 \cdot 7^2 \cdot 11^2 \cdot 13^2 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31$$

This "almost inadvertent" solution has a maximum root, 151, considerably smaller than one might have guessed from extrapolating previous trends.

After expending roughly 10^{17} floating point operations on Newton's method, Peter Borwein, Petr Lisonek and Colin Percival [Math. Comp. 72 (2002) 2063-2070] found no further solution with degree $s > 10$.

It seemed to me that this problem is more suited to modular arithmetic than to a numerical method. At degree $s = 12$, I found all non-trivial ideal symmetric solutions modulo 41 and all such solutions modulo 53. Then I used the Chinese remainder theorem (CRT) to generate solutions with maximum root less than 41×53 and C coprime to both 41 and 53. Thus I found a new solution, with

$$[472, 639, 1294, 1514, 1947, 2037] \sim [257, 891, 1109, 1618, 1896, 2058]$$

$$C = 2^{14} \cdot 3^8 \cdot 5^3 \cdot 7^2 \cdot 11^2 \cdot 13^2 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 37 \cdot 43 \cdot 47 \cdot 61 \cdot 89 \cdot 191 \cdot 419$$

found in 72 GHz-hours of running gfortran with this CRT strategy.

Perhaps a CRT strategy might lead to a solution with degree $s > 12$.

2 Amateur number theory d'après Gaston Tarry

From MacTutor at St Andrews:

Gaston Tarry:
Born: 27 Sept 1843 in Villefranche de Rouergue, Aveyron, France
Died: 21 June 1913 in Le Havre, France

Gaston Tarry was born in Villefranche de Rouergue which is located in south central France. He attended the Lycée Saint-Louis in Paris where he became interested in mathematics. He never looked for an academic career as a mathematician, however, and joined the Contributions Diverses de l'Administration des Finances (French Financial Administration). He spent the whole of his working life in Algeria. France had invaded Algeria in 1830 and French rule was established by 1847. Tarry was part of the French administration in Algeria, retiring in 1902.

Although an amateur mathematician, Tarry had an amazing ability to analyse combinatorial problems. One has simply to feel amazement at some of the problems he solved using purely combinatorial and calculating skills... Even more surprising [David: sic] is the fact that his mathematical achievements came after the age of fifty... [David: Hmmm...]

A notable (but not ideal) solution by Tarry:

$$2^k + 3^k + 14^k + 18^k + 39^k + 43^k + 45^k + 49^k + 55^k + 61^k + 76^k + 86^k + 92^k + 96^k$$

and

$$1^k + 5^k + 11^k + 21^k + 36^k + 42^k + 48^k + 52^k + 54^k + 58^k + 79^k + 83^k + 94^k + 95^k$$

are equal for each integer exponent k from 0 to 10.

The amateur path that led me to be interested in the Tarry problem began when Jens Kruse Andersen recently created the URL
`http://hjem.get2net.dk/jka/math/consecutive_factorizations.htm`
for records of complete factorizations of consecutive integers above 500 decimal digits.

Thanks to polynomial forms adapted from suggestions by Jaroslaw Wroblewski, I was able to add new records for 6, 8 and 10 consecutive factorizations, within a few days.

Clearly, those records may be improved upon, merely by throwing more cycles at similar polynomials. However, the more interesting question, for number theorists, is whether there are polynomials significantly better than those that were used for the recent records.

3 Progress from 2000 CE to 400 CE

Fermat, Euler, Lagrange, Gauss: To express a number as the sum of two squares in 4 different ways we may use 3 Gaussian primes.

Example: The product of $5 = |2 + \mathrm{i}|^2$, $13 = |3 + 2\mathrm{i}|^2$ and $17 = |4 + \mathrm{i}|^2$, is equal to:
$|(2 + \mathrm{i})(3 + 2\mathrm{i})(4 + \mathrm{i})|^2 = 9^2 + 32^2$
$|(2 - \mathrm{i})(3 + 2\mathrm{i})(4 + \mathrm{i})|^2 = 31^2 + 12^2$
$|(2 + \mathrm{i})(3 - 2\mathrm{i})(4 + \mathrm{i})|^2 = 33^2 + 4^2$
$|(2 - \mathrm{i})(3 - 2\mathrm{i})(4 + \mathrm{i})|^2 = 23^2 + 24^2$

So $A = (y - 23^2)(y - 24^2)$, $B = (y - 12^2)(y - 31^2)$ and $C = (y - 9^2)(y - 32^2)$ differ only by integers. Moreover $A - B = 3(B - C)$ which is why I was able to be sure of 3 of the 8 consecutive factorizations in the 600-digit record starting with:

$$n = (y - 23^2)(y - 24^2)/55440 - 6$$

where $y = (z(5z + 9)/2 - 31)^2$ and $z = (1320(10^{22} + 1932187))^3$

Thanks to Gauss, n was constructed such that $n + 2$, $n + 3$ and $n + 6$ are each the product of known divisors with at most 152 digits, where the large divisors are suited for SNFS (special number field sieve).

Such divisors may then be factorized one at a time, albeit with some patience, after one has used GMP-ECM to find 5 prime factors with more than 500 digits in the other 5 members of the sequence.

No method has so far been devised to produce 3 nearby free lunches that are significantly better than this three-fold PTE solution by quartics.

Thus to achieve 7 consecutive factorizations at 1404 decimal digits, I had to find 5 much larger primes, with only two 12-th order free lunches, for the record starting with

$$n = (y - 11^4)(y - 35^2)(y - 47^2)(y - 94^2)(y - 146^2)(y - 148^2)/m - 4$$

where $m = 674402945596760540160000$ and $y = (m(10^{96} + 10624986) + 22)^2$.

That record used the best-so-far PTE solution of Chen et al. But if one were to find an ideal solution with degree $s > 12$, improvement of such consecutive factorization records might be possible.

Having learned from Peter Borwein's papers that no PTE solution was found at degree $s = 11$ with maximum root less than 2000, I began to realize just how difficult it might be to find a solution with $s > 12$.

But then I thought that maybe some old maths might help.

Sun Zi wrote a relevant mathematical manual (suanjing) some time around 400 CE.

"The Sunzi suanjing mentions the mein as an item of taxation, and the hu tiao system. These two were first established in 280 AD. So the book could not have been written before this date. ... A new scale between chih and tuan was established in 474 AD; the Sunzi, still using the old scale by Wu Ch'en-Shih's emendation, cannot be older than 473 AD." L Wang, The date of the Sunzi suanjing and the Chinese remainder theorem, in Proc. Tenth Internat. Conf. History of Science, 1962 (Paris, 1964), 489-492.

Q: Suppose we have an unknown number of objects. When counted in threes, 2 are left over, when counted in fives, 3 are left over, and when counted in sevens, 2 are left over. How many objects are there?

A: Multiply the number of units left over when counting in threes by 70, add the product of the number of units left over when counting in fives by 21, and then add the product of the number of units left over when counting in sevens by 15. If the answer is 106 or more then subtract multiples of 105.

```
forprime(p=3,7,print1((p/105)%p*105/p" "))
70 21 15
```

4 Smart Pari-GP and parsimonious ForTran

I began by trying to regenerate the 12th-order Chen solution

$$[35, 47, 94, 121, 146, 148] \sim [22, 61, 86, 127, 140, 151]$$

First I used the `polrootsmod` command of Pari-GP to find all the independent non-trivial 12th-order ideal solutions modulo each prime $p \leq 61$. By non-trivial, I mean that C is not divisible by p, so the polynomials differ modulo p. By independent, I mean that I count only 1 of the $p-1$ solutions that are related by multiplying both sets of integers by an integer coprime to p.

Let $N_s(p)$ denote the number of independent non-trivial solutions modulo p at an even order s.

For the primes

$$p = 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61$$

the corresponding numbers of independent 12-th order solutions are

$$N_{12}(p) = 1, 0, 0, 1, 0, 2, 4, 6, 6, 9, 11, 24, 36$$

To identify a solution takes only 7 integers: the 6 roots of one polynomial in the interval $[0, (p-1)/2]$ and a single root of the other polynomial, to determine C.

Gluing together solutions modulo p_1 with solutions modulo p_2 clearly leads to a combinatoric explosion. At even degree s there appear to be

$$G_s(p_1, p_2) = 2^s(s/2)!s(p_1-1)(p_2-1)N_s(p_1)N_s(p_2)$$

non-trivial solutions modulo p_1p_2, with $G_{12}(59, 61) > 10^{14}$, for example.

In fact, the frequent presence of zero roots and non-zero multiple roots in the solutions modulo p_1 and modulo p_2 means that this is a considerable overestimate, by a factor of more than 8 in the case of $G_{12}(41, 53)$, for example.

What was required was a smart way of outputting from Pari-GP a manageably small set of combinatoric instructions to the brute-force low-overhead ForTran test loops that quickly enabled ForTran to avoid testing the same gluing twice.

When that was achieved, the next issue was to find a fast filter, inside the innermost ForTran loop, to reject solutions modulo p_1p_2 that cannot be solutions modulo other primes.

Here I used the fact that at degree $s = 12$ the constant

$$C = \prod_{n=1}^{6} (a_n^2 - b_1^2)$$

must be divisible by the 31-bit integer $m = 1388423575$. The multiplications in the product-form for C were performed modulo m, with only one being needed in the innermost loop.

That fast filter had a high rejection rate, allowing time for a further 31-bit modular test of C, and two further tests of the identity

$$\prod_{n=1}^{6} (a_n^2 - b_1^2) = \prod_{n=1}^{6} (a_n^2 - b_2^2)$$

modulo a pair of pseudo-random 31-bit primes, inside a rarely used further loop that generated an alternative CRT candidate, b_2.

The principle was to design each of the 4 ForTran filters according to how often it was invoked. The first was massively aided by Pari-GP pre-processing of the combinatorics; later filters could be more sophisticated according to how rarely they were needed. At the end of the ForTran run with $s = 12$, $p_1 = 41$ and $p_2 = 53$, only 22 candidates emerged from the 4 modular filters, for close scrutiny by Pari-GP.

These included all 14 of the multiples of the Chen solution with maximum root less than 41×53, the new solution reported to the NMBRTHRY list, and merely 7 frauds, which had correctly passed all 4 modular filters without being integer solutions in Z.

At present I am running this strategy with $s = 14$, $p_1 = 61$ and $p_2 = 67$.

5 The Open University in Canada

Distance learning has been thriving in the UK for more than 35 years. Now the OU takes a significant proportion of its students from further afield. I thought that I should advertise a few of 100 or so OU courses available in Canada:

Advanced database technology (M877)
Communicating science (S804)
Contemporary issues in science learning (SEH806)
Current issues in public management and social enterprise (B857)
Distributed applications and e-commerce (M879)
Educational enquiry (E891)

Environmental ethics (TXX861)
Fundamentals of senior management (BZX713)
Imaging in medicine (S809)
Information security management (M886)
Innovations in e-learning (H807)
Issues in brain and behaviour (SD805)
Language and literacy in a changing world (E844)
Molecules in medicine (S807)
Relational database systems (M876)
Science and the public (S802)
Software requirements for business systems (M883)
User interface design and evaluation (M873)
War, intervention and development (TUXX875)
Web systems integration (M887)

6 Acknowledgements

I thank

- Jens Kruse Andersen, for his computational challenge
- Jaroslaw Wroblewski, for alerting me to the Chen–Kuosa–Meyrignac solution
- Peter Borwein, Colin Ingalls, Petr Lisonek and Colin Percival, for the quality of their exposition of the Prouhet–Tarry–Escott problem
- Kevin Buzzard for advice and encouragement
- Jonathan and Judith Borwein for their wonderful hospitality here in Halifax
- Gaston Tarry, for remaining an amateur of mathematics well into his 60s
- Sun Zi, for Problem 26 in Chapter 3 of his suanjing

David Broadhurst
Coast to Coast Seminar
October 12, 2007, 3:30pm Atlantic Time
Presented from Dalhousie

Appendix

I accessed these 6 URLs, during my talk:

1: What: A new Prouhet-Tarry-Escott solution in the NMBRTHRY list:
`http://listserv.nodak.edu/cgi-bin/wa.exe?A2=ind0709&L=nmbrthry&P=R994`

2: Why: Largest consecutive factorizations recorded by Jens Kruse Andersen:
`http://hjem.get2net.dk/jka/math/consecutive_factorizations.htm`

3: When: 1600-year-old math by Sun Zi:
`http://www-groups.dcs.st-and.ac.uk/~history/Biographies/Sun_Zi.html`

4: How: Brute force by ForTran:
`http://physics.open.ac.uk/~dbroadhu/fortend.txt`

5: Where: The Open University spreads its wings:
`http://www3.open.ac.uk/courses/countries/Canada.shtm`

6: Who: A very helpful review by Peter Borwein and Colin Ingalls:
`http://www.cecm.sfu.ca/~pborwein/PAPERS/P98.ps`