An upper bound in Goldbach's problem

by

Jean–Marc Deshouillers, Andrew Granville *,
Władysław Narkiewicz and Carl Pomerance *

In memory of D.H. Lehmer.

Abstract: It is clear that the number of distinct representations of a number n as the sum of two primes is at most the number of primes in the interval $[n/2, n-2]$. We show that 210 is the largest value of n for which this upper bound is attained.

1. Introduction.

In 1742 Christian Goldbach wrote, in a letter to Euler, that on the evidence of extensive computations he was convinced that every integer exceeding 6 was the sum of three primes. Euler replied that if an even number $2n + 2$ is so represented then one of those primes must be even and thus 2, so that every even number $2n$, greater than 2, can be represented as the sum of two primes; it is easy to see that this conjecture implies Goldbach's original proposal, and it has widely become known as *Goldbach's conjecture*.

Although still unresolved, Goldbach's conjecture is widely believed to be true. It has now been verified for every even integer up to 2×10^{10} (in [3]), and there are many interesting partial results worthy of mention.

In 1930, Šnirel'man [8] proved the existence of an integer k such that every integer larger than 1 may be written as the sum of at most k primes (recently Ramaré [5] has shown that every positive even integer is the sum of at most 6 primes).

In 1937, I. M. Vinogradov [9] showed that every sufficiently large odd integer n may be written as the sum of three primes (recently Chen Jing-run and Wang [2] have shown this for all $n > 10^{43,000}$).

In 1966, Chen Jing-run [1] showed that every sufficiently large even integer n may be written as the sum of a prime and a number that has at most two prime factors.

In 1975, Montgomery and Vaughan [4] showed that there exist constants $c > 0$ and $\delta > 0$ such that there are no more than $cx^{1-\delta}$ even integers $n \leq x$ that cannot be written as the sum of two primes.

Define $g(n)$ to be the number of representations of n as $p + q$ with $p \geq q$. Goldbach's conjecture may be rephrased as $g(n) > 0$ for all even $n > 2$; in other words, the 'trivial' lower bound does not hold with equality for any even $n > 2$. A 'trivial' upper bound for $g(n)$ is given by the possibility that, for every prime p in the range $n/2 \leq p \leq n - 2$, we have

* A.G. and C.P. are supported, in part, by the National Science Foundation

$n-p$ prime, so that

$$g(n) \le \pi(n-2) - \pi(\tfrac{1}{2}n-1), \tag{1}$$

where $\pi(x)$ denotes the number of primes up to x. In analogy with Goldbach's conjecture, the fourth–named author conjectured that $n = 210$ is the largest value for which equality holds. In what follows we prove this conjecture.

Theorem. *The number 210 is the largest positive integer n that can be written as the sum of two primes in $\pi(n-2) - \pi(\frac{1}{2}n-1)$ distinct ways.*

The only other possibilities arise when $n \le 8$, or $2|n$ and $n \le 18$, or $2 \times 3|n$ and $n \le 48$, or $2 \times 3 \times 5|n$ and $n \le 90$.

It is amusing that the equality $g(210) = \pi(208) - \pi(104)$ may be verified mentally, since if p is prime and $105 \le p \le 208$, then $2 \le 210-p \le 105$ and $210-p$ is coprime to 2, 3, 5, 7 (since $210 = 2 \times 3 \times 5 \times 7$), so $210-p$ is also prime.

2. Bertrand's postulate for arithmetic progressions.

To see the relevance of Bertrand's postulate to our problem, we first deal with the case that n is odd. By a minor modification of the standard proof of Bertrand's postulate, it is easy to show that for every odd integer $n \ge 9$, there exists a prime p in the range $(n+1)/2 \le p \le n-4$; but then $n-p$ is even and $n-p > 2$, and therefore $n-p$ is not prime, so that equality fails in (1). We then check that equality holds in (1) for each odd $n < 9$. (The case of odd n is actually reproved below as part of the general case.)

This same idea may be carried over to primes in arithmetic progressions: If, for a given prime q, there exists a prime p in the range $n/2 \le p < n-q$ which belongs to the arithmetic progression n modulo q, then $n-p$ is divisible by q but $n-p > q$ and so $n-p$ cannot be prime; and therefore equality fails in (1). This can be rephrased as follows.

Lemma 1. *Suppose that equality holds in (1) for n. If q is a prime such that for each a, $1 \le a \le q-1$, there exists a prime $p \equiv a \pmod q$ with $\frac{1}{2}n \le p < n-q$, then q divides n.*

A straightforward consequence of this is the following result.

Lemma 2. *Suppose that we are given positive integers x, y, z, and sets of primes $\mathcal{P}$ and $\mathcal{Q}$ with the following properties:*

(i) the primes in $\mathcal{Q}$ are all at most z, and their product exceeds $2x$;

(ii) each prime in $\mathcal{P}$ lies in the interval $[x, x+y]$;

(iii) for each prime $q \in \mathcal{Q}$ and each integer a, $1 \le a \le q-1$, there exists a

prime $p \in \mathcal{P}$ with $p \equiv a \pmod q$.
Then equality fails in (1) for every integer n in the interval $(x+y+z, 2x]$.

Proof: Suppose that equality holds in (1) for some n in $(x+y+z, 2x]$. By (iii), and the ranges for $p \in \mathcal{P}$ and $q \in \mathcal{Q}$ given in (ii) and (i) respectively, we can invoke Lemma 1 to show that q divides n for each $q \in \mathcal{Q}$. But then n is divisible by the product of all of the primes in $\mathcal{Q}$, so that $n \le 2x$ contradicts (i).

Using tools of analytic number theory we can then deduce the following weak version of our Theorem.

Proposition. *There exists an effectively computable constant n_0 such that equality fails in (1) for $n \ge n_0$.*

Proof: It is well known that one can give an effective uniform estimate for the number of primes $p \le x$ with $p \equiv a \pmod q$, provided $q \le (\log x)^{2-\varepsilon}$, for some fixed $\varepsilon > 0$. Taking $\varepsilon = 1/2$, these estimates are strong enough to imply that there exists an effectively computable x_0 such that, if $x \ge x_0$ then the hypothesis of Lemma 2 holds for $y = x/2$, $\mathcal{Q}$ the set of primes up to $z = 2\log x$, and $\mathcal{P}$ the set of primes in the interval $[x, 3x/2]$.

In the next section we will give a different proof of the Proposition. This new proof will have the advantage that we will obtain $n_0 = 2 \times 10^{24}$, whereas we were only able to get $n_0 = 10^{520}$ by the methods of this section (one could do much better here if one assumed the Generalized Riemann Hypothesis). Lemma 2 is, however, useful for computation, and we will use it to close the gap between 210 and 2×10^{24}.

3. Sieve methods.

Either Brun's sieve or Selberg's sieve may be used to prove the estimate $g(n) = O(n \log\log n / \log^2 n)$. Then, as $\pi(n-2) - \pi(\frac{1}{2}n - 1) \gg n/\log n$ by the Prime Number Theorem, this implies that equality fails in (1) for sufficiently large n, so providing another proof of the Proposition. We will now make this proof explicit so as to obtain $n_0 = 2 \times 10^{24}$ in the Proposition.

Combining Lemma 5 and (2.4) of [6] we have for $N \ge e^{48}$ that

$$g(2N) \;\le\; 10.57\, \kappa_N \,\frac{N}{(8.2 + \log N)\log N}, \qquad \text{where } \kappa_N = \prod_{p \mid N,\; p>2} \left(\frac{p-1}{p-2}\right). \tag{2}$$

From this we shall deduce the following result.

Lemma 3. *If $N \ge 10^{24}$ then*

$$g(2N) \;<\; 0.961\, \frac{N}{\log N}. \tag{3}$$

Theorem 2 of [7] implies that $\pi(2N-2) - \pi(N-1) \ge 0.961 N/\log N$ for $N > e^{50}$; combining this with Lemma 3 we can deduce a value for n_0 in the Proposition above:

Explicit Proposition. *Equality fails in (1) for* $n \geq 2 \times 10^{24}$.

It remains only to give a
Proof of Lemma 3: Let $p_1 = 3, p_2 = 5, p_3 = 7, \ldots$ be the sequence of odd primes, and let N_j be the product of the first j of them. Suppose that $10^{24} \leq N < N_{18}$ so that N has at most 17 distinct odd prime divisors. Thus, as $\log N > 55$,

$$\frac{10.57\ \kappa_N}{8.2 + \log N} \leq \frac{10.57\ \kappa_{N_{17}}}{8.2 + 55} < 0.96,$$

and so (3) follows from (2) for such values of N.

Now, if $N \geq N_{18}$ then there exists a value of $j \geq 18$ for which $N_j \leq N < N_{j+1}$, so that $\kappa_N/(8.2 + \log N) \leq \kappa_{N_j}/(8.2 + \log N_j)$. Lemma 3 then follows from

(4) $$\frac{10.57\ \kappa_{N_j}}{8.2 + \log N_j} < 0.961, \text{ for } j \geq 18.$$

To see (4) first note that it holds for $j = 18$ by direct computation. Further, the sequence on the left of (4) is decreasing for $j \geq 4$. To see this first note that the ratio of consecutive terms is

(5) $$\left(\frac{p_j - 1}{p_j - 2}\right) \frac{(8.2 + \log N_{j-1})}{(8.2 + \log N_j)}.$$

But, $N_{j-1} \leq 3 \times 5 \times 7 \times 9 \times \ldots \times p_{j-1} \leq p_{j-1}^{(p_{j-1}-1)/2} < p_j^{p_j - 3}/e^{8.2}$ for $j \geq 4$, so the expression in (5) is less than 1. This completes the proof of Lemma 3.

4. Computations.

In order to complete the proof of the Theorem we need to fill the gap left by the Explicit Proposition of the previous section. We performed some computations to achieve this:

Computational Proposition. *Equality fails in (1) for each* n *in the range* $210 < n \leq 2 \times 10^{24}$.

At first sight it might seem possible to rule out each n in this range by simply finding a prime p, $n/2 \leq p \leq n - 2$, such that $n - p$ is not prime; however doing 2×10^{24} such searches is prohibitively expensive.

Lemma 2 can evidently be used to quickly rule out wide ranges of values of n. One way to use Lemma 2 is to choose $z = 2 \log x$. Then we search through $x, x+1, x+2, \ldots$ for primes for the set $\mathcal{P}$. Each time we find such a p we note, for each $q \leq z$, the residue class in which p belongs, modulo q. A weak heuristic argument suggests that by the time we have searched (for elements of $\mathcal{P}$) as far as $x+(\log x)^{2+\epsilon}$, we will certainly have found a set $\mathcal{Q}$ satisfying both (i) and (iii) of Lemma 2.

In practice this works fine for small values of x, but for x around, say, 10^{20}, it becomes very slow to test each of the integers $x, x+1, x+2, \dots$ for primality. To be sure, it is still easy to pick out the 'industrial grade primes' using a Fermat test, but proving these numbers prime begins to become time consuming. However there are some extremely fast primality tests for integers of certain special types; one such being the following.

Proth primality test (1878). *If* $p \equiv 1 \pmod{2^k}$ *where* $2^k > \sqrt{p}$*, and if there exists an integer* a *for which* $a^{(p-1)/2} \equiv -1 \pmod{p}$ *then* p *is prime.*

So, in order to use Lemma 2, we choose k so that $2^k > \sqrt{2x}$ and then search the interval $[x, x+y]$ for primes $p \equiv 1 \pmod{2^k}$ using the Proth primality test, checking whether any of the primes $a \leq 47$ satisfy the criterion $a^{(p-1)/2} \equiv -1 \pmod{p}$. (If some such a satisfies $a^{(p-1)/2} \not\equiv \pm 1 \pmod{p}$, we of course discard p, since it is composite. If each prime $a \leq 47$ satisfies $a^{(p-1)/2} \equiv 1 \pmod{p}$, we also discard p.)

Enrico Bombieri kindly programmed the above in C for us (for which we'd like to thank him), using a multiprecision routine. For simplicity's sake he applied Lemma 2 using the first construction above, nineteen times, to rule out all n in the range $210 < n \leq 9,330,712$. Then he used the second construction above (that is, using Proth's test) for all remaining n, which took a further sixty-three applications of Lemma 2. In all cases the set $\mathcal{P}$ contained at most 465 elements (which corresponds to searching for primes through about $5\log^2 x$ numbers around x), and we had $z \leq 67$. The total run time on a Sparc 2 was just under 75 minutes. A copy of the computer code is available upon request from the second–named author.

Acknowledgement. Three of us (Deshouillers, Granville and Pomerance) would like to acknowledge the hospitality of the Institute for Advanced Study in Princeton where some of the work for this paper was done.

References

[1] J. R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Acta. Math. Sci. Sinica, I, **16** (1973), 157–176; II, **21** (1978), 421–430.

[2] J. R. Chen and T. Wang, *On the odd Golbach problem*, Acta. Math. Sci. Sinica, **32** (1989), 702–718.

[3] A. Granville, J. van de Lune and H. J. J. te Riele, *Checking the Goldbach conjecture on a vector computer*, in Number Theory and Applications (pp. 423–433), R.A. Mollin ed., Kluwer Acad., 1989.

[4] H. L. Montgomery and R. C. Vaughan, *The exceptional set in Goldbach's problem*, Acta Arith. **27** (1975), 353–370.

[5] O. Ramaré, *On Šnirel'man's constant*, preprint.

[6] H. Riesel and R. C. Vaughan, *On sums of primes*, Ark. Math. **21** (1983), 45–74.

[7] J. B. Rosser and L. Schoenfeld, *Approximate formulae for some functions of prime numbers*, Ill. J. Math. **6** (1962), 64–94.

[8] L. Šnirel'man, *Über additive Eigenschaften von Zahlen*, Ann. Inst. Polytechn. Novocerkask **14** (1930), 3–28; and Math. Ann. **107** (1933), 649–690.

[9] I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, C.R. Acad. Sci. URSS **15** (1937), 6-7.

Addresses: (Deshouillers) Mathématiques Stochastiques, BP26, Université Bordeaux 2, 33076 BORDEAUX Cedex, France.

(Granville and Pomerance) Department of Mathematics, University of Georgia, Athens, Georgia 30602, U.S.A.

(Narkiewicz) Instytut Matematyczny, Uniwersytet Wrocławski, Plac Grunwaldzki 2/4, PL-50-384 Wrocław, Poland.