A Computer Study of the Goldbach Curve

By Paul D. Coddington

Abstract. The Goldbach curve $N_2(n)$ is the number of distinct ways that an even number n may be written as the sum of two primes. $N_2(n)$ is computed for n up to 2×10^7, and found to agree extremely well at large n with an asymptotic formula proposed by Littlewood and Hardy. Some other conjectures resulting from a previous calculation of Stein and Stein are also investigated, and a probable lower bound for $N_2(n)$ is proposed.

1. Introduction. The Goldbach conjecture states that every even number greater than 2 can be represented as the sum of two primes. Although it is undoubtedly true, it has remained unproven since its proposal by Goldbach in a letter to Euler in 1742. Stein and Stein [1] investigated Goldbach's conjecture by computing the Goldbach curve $N_2(n)$, which is the number of distinct ways that an even number n may be written as the sum of two primes (throughout this paper, n will denote an even positive integer). Their calculations were done for n up to 150,000. The results strongly supported the Goldbach conjecture, and also led them to make two further conjectures concerning the Goldbach curve:

CONJECTURE I. *For every integer $k > 0$ there exists at least one even number n such that $N_2(n) = k$.*

CONJECTURE II. *For every integer $k > 4$, the smallest solution n of the equation $N_2(n) = k$ is such that $n \equiv 0 \pmod 6$.*

Conjecture II is quoted by Stein and Stein as being true only for $k > 8$, but this is because in their calculations they counted 1 as being a prime number, which we do not.

Stein and Stein also devised a formula which approximates the Goldbach curve, apparently unaware that Littlewood and Hardy had previously proposed a much simpler formula for the asymptotic behavior of $N_2(n)$ [2]. The Littlewood-Hardy conjecture is that

$$N_2(n) \sim \rho(n) = C_2 \frac{n}{(\log n)^2 - 2\log n} \prod_{p|n} \frac{p-1}{p-2}, \tag{1}$$

where p is any odd prime, and

$$C_2 = \prod_p \left(1 - \frac{1}{(p-1)^2}\right) \approx 0.6601618. \tag{2}$$

The $2\log n$ term is of course redundant for an asymptotic formula, however for the modest values of n with which we are dealing it is very important.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 11P32.

In this work the calculation of the Goldbach curve has been extended to values of n up to 2×10^7. The results are compared with the formula of Littlewood and Hardy, an estimate of the error as a function of n is given, and a lower bound is proposed for $N_2(n)$. Conjectures I and II have also been tested in this extended interval.

2. The Goldbach curve and the Littlewood-Hardy conjecture. A notable feature of the Goldbach curve is that it is not "smooth" – two neighboring values of n can have greatly different values of $N_2(n)$. This feature is also present in the Littlewood-Hardy function $\rho(n)$, which can be expressed as the product of two terms, $\rho(n) = \rho_0(n)f(n)$, where $f(n)$ is the product over odd prime factors of n in equation (1). Here $\rho_0(n)$ is a smooth function, whereas $f(n)$ is highly discontinuous, being dependent on the factor structure, rather than the magnitude, of n. The minimum value taken by $f(n)$ is 1 (when n is a power of 2, and thus has no odd prime factors), so that $\rho_0(n)$ provides a lower bound on $\rho(n)$.

It is instructive to look at the upper and lower bounds (i.e. the envelope) of the Goldbach curve. Let us define the minimum extent of the curve, $N_2^{min}(n)$, as the greatest number which satisfies $N_2(m) > N_2^{min}(n)$ for all even $m > n$, and similarly the maximum extent $N_2^{max}(n)$, as being the smallest number satisfying $N_2(m) < N_2^{max}(n)$ for all even $m < n$. Table 1 shows $N_2^{min}(n)$ for various values of n. We see, for example, that all even numbers greater than 11672 can be written as the sum of two primes in more than 100 different ways, while even numbers greater than 195368 have more than 1000 such decompositions.

$N_2^{min}(n)$	n	$N_2^{min}(n)$	n
10	632	1000	195368
20	1448	1500	317972
50	4688	2000	445268
100	11672	5000	1315298
150	19246	10000	2947918
200	27908	15000	4719452
500	85616	20000	6594752

Table 1

Lower bounds $N_2^{min}(n)$ on the Goldbach curve.

$N_2^{min}(n)$ and $N_2^{max}(n)$ are plotted in Fig. 1. The Goldbach curve lies on or between these two curves, so looking at this figure it is easy to see why the Goldbach conjecture (that $N_2(n) > 0$ for $n > 2$) is believed to be true. The smooth curve in the figure is $\rho_0(n)$, which gives a lower bound on the Littlewood-Hardy function, and can be seen to closely approximate $N_2^{min}(n)$ at large n. This leads us to postulate a simple lower bound for $N_2(n)$. Taking the asymptotic form of $\rho_0(n)$, by ignoring the term linear in $\log n$ in the denominator of equation (1), gives the function $C_2 n(\log n)^{-2}$, which is a lower bound for $N_2(n)$ for all $n > 2672$ in the region we have checked ($n \leq 2 \times 10^7$). However it could easily be possible that this bound fails when n becomes so large that the difference between

$\rho_0(n)$ and its asymptotic form given above is negligible. A function which seems certain to be a true lower bound is easily constructed by merely changing the multiplicative factor from C_2 to $1/2$. This function is indeed a lower bound for all $n > 4$ in the region we have studied. This leads to the following conjecture:

$$N_2(n) > \frac{n}{2(\log n)^2} \quad \text{for all } n > 4. \tag{3}$$

Now let us look at the Goldbach curve and the Littlewood-Hardy formula in more detail. Fig. 2 shows a portion of the Goldbach curve at large n. The striking thing about the curve is the banding effect, which is explained by the Littlewood-Hardy formula as being due to the distinct values of $f(n)$, which describes the factor structure of n. Large prime factors do not have much effect on the value of $f(n)$, so it is the smaller factors which produce the different bands. At the right of the figure are arrows denoting the value of $\rho(n)$ for n close to 8.05×10^6, and having the prime factors shown to the right of the arrow. These values not only match the positions of the bands perfectly, but also explain the different densities of the bands, with the more common (i.e. smallest) factors attracting the most points.

As a final check on the Littlewood-Hardy conjecture, the difference $D(n) = |\rho(n) - N_2(n)|$ and the error $E(n) = D(n)/N_2(n)$ have been calculated. Like the Goldbach curve, $D(n)$ and $E(n)$ are not smooth functions, so instead we investigate the maximum error $E^{max}(n)$, defined as the smallest error value satisfying $E(m) < E^{max}(n)$ for all $m > n$, as well as the local average of $E(n)$ in the neighborhood of n. Similarly we can define minimum and average values of the difference $D(n)$, and fitting a power law to these values shows that they increase roughly as $\sqrt{n}$ at large n. We therefore expect the maximum and average values of the error, which are plotted in Fig. 3, to be of order $((\log n)^2 - 2\log n)/\sqrt{n}$, and indeed this does produce a reasonable fit to the two curves at large n. In any case, the error is clearly falling with increasing n, and for $n > 10^7$ the maximum error is less than 2%, while the average error is less than 0.3%. This strongly suggests that the Littlewood-Hardy formula is indeed asymptotic to $N_2(n)$.

3. The conjectures of Stein and Stein. The two conjectures made by Stein and Stein have also been checked and found to to hold for $n \leq 2 \times 10^7$. Stein and Stein showed that over the range of n they studied, conjecture I was true for all $k \leq 1911$, with the first gap (i.e. no solution found) at 1912. Over our larger range of values, it was found that the first gap is not until 125764. We have looked at the first gap as a function of M, the maximum value of n for which the Goldbach curve is computed, and found that it increases approximately as $M^{0.85}$, which suggests that the conjecture is true.

We can also look at the number of solutions of $N_2(n) = k$ for each value of k, which we denote by $S(k)$. Just as we have done for the Goldbach curve, we can define a lower bound $S^{min}(k)$ of this function to be the greatest value such that $S(j) > S^{min}(k)$ for all $j > k$. Some of these values are given in Table 2, which shows that $S^{min}(k)$ is increasing, albeit slowly, and for $k > 24041$ there are more than 100 solutions of the equation. Hence conjecture I seems very solid.

$S^{min}(k)$	k	$S^{min}(k)$	k
10	2	60	1202
20	12	70	2486
30	88	80	7520
40	252	90	21347
50	382	100	24041

Table 2

Lower bounds $S^{min}(k)$ on the number of solutions of $N_2(n) = k$.

The truth of conjecture II is not so clear. Stein and Stein argued that this conjecture was plausible since the Goldbach curve "almost always" has a local maximum where $n \equiv 0 \pmod 6$. This is because $f(n)$ has the greatest contributions from the smallest prime factors, especially 3. In the range $36 < n <$ 150,000 studied by Stein and Stein, all multiples of 6 have more Goldbach decompositions than their immediate neighbors, with only three exceptions: $N_2(912) = N_2(910) = 31$, $N_2(1542) = N_2(1540) = 46$, and $N_2(80082) = 1005$, $N_2(80080) = 1006$. From this data it would appear that exceptions are very rare, and that when they do occur the difference between the two values of $N_2(n)$ is very small. However in the range $36 < n \leq 2 \times 10^7$ there are 282 exceptions, which gives a rate of about 28 exceptions per million even integers. This is small enough to indicate why a counterexample to conjecture II has not been found, but large enough to arouse suspicion that a counterexample may exist.

To see how these exceptions can arise, let us look at the smallest one beyond the range of the Stein and Stein calculation, which is $N_2(190188) = 2069$, $N(190190) = 2095$. The prime factors of 190188 are 2,3, and 587, giving $f(190188) \approx 2.0034$, whereas 190190 has prime factors 2,5,7,11,13 and 19, giving an f of about 2.0535. Thus it is possible with enough small prime factors to overcome the lack of a factor of 3 and get an $f(n)$ above 2. This can produce quite substantial differences in the values of $N_2(n)$ for exceptional pairs, for example $N_2(5895888) = 37220$, while $N_2(5895890) = 39199$, a difference of nearly 2000.

We have seen, however, that the equation $N_2(n) = k$ has many solutions for large k, so it may be that this acts in favor of average, rather than exceptional, behavior, and conjecture II will prove to be true after all.

Acknowledgements. The calculations took a total of about 600 hours using two parallel computers, a 20 processor Sequent Symmetry and a 20 processor Encore Multimax, with some checks done using a DECstation 3100 workstation. I would like to thank the Caltech Concurrent Supercomputer Facility and the Advanced Computing Research Facility at Argonne National Laboratory for the use of these machines.

Caltech Concurrent Computation Program
California Institute of Technology
Pasadena, CA 91125, USA
E-mail: paulc@wega.caltech.edu

Current address:
Physics Department
Syracuse University
Syracuse, NY 13244, USA

1. M.L. STEIN & P.R. STEIN, "New experimental results on the Goldbach conjecture", *Math. Mag.*, v. 29, 1965, pp. 72-80.

2. G.H. HARDY & J.E. LITTLEWOOD, "Some problems of 'partitio numerorum'; III: On the expression of a number as a sum of primes", *Acta Math.*, v. 44, 1923, pp. 1-70.