A proof of Sophie Germain primes conjecture

Marko Jankovic

Proof of the Sophie Germain Primes Conjecture

Marko V. Jankovic

Institute of Electrical Engineering “Nikola Tesla”, Belgrade, Serbia,

Abstract In this paper a proof of the existence of an infinite number of Sophie Germain primes is going to be presented. Originally very difficult problem (in observational space) has been transformed into a simpler one (in generative space) that can be solved. It will be shown that Sophie Germain primes could be obtained through two stage sieve process, and that will be used to used to prove that infinitely many Sophie Germain primes exists.

1 Introduction

A prime p is a Sophie Germain prime if $2p + 1$ is a prime too [1]. In that case the prime number $2p + 1$ is called safe prime. These special primes have applications in public key cryptography, pseudorandom number generation, and primality testing; see, for example [2, 4, 6]. Originally, they have been used in the investigation of cases of Fermat's last theorem [3]. It has been conjectured that infinitely many Sophie Germain primes exist, but this was unproven (see for instance, [5]).

In this paper it is going to be proved that an infinite number of Sophie Germain primes exists. The problem is addressed in generative space, which means that prime numbers are not going to be analyzed directly, but rather their representatives, that can be used to produce them. It will be shown that Sophie Germain primes could be generated by two stage recursive type process. This process will be compared to other two stage recursion sieve process that leaves infinitely many numbers. Fact that sieve process that generate Sophie Germain primes leaves more numbers than the other sieve process, will be used to prove that infinitely many Sophie Germain primes exist.

Remark 1: *In this paper any infinite series in the form $c_1*l \pm c_2$ is going to be called a thread defined by number c_1 (in literature these forms are known as linear factors – however, it seems that the term thread is probably better choice in this context). Here c_1 and c_2 are numbers that belong to*

the set of natural numbers (c_2 can also be zero and usually is smaller than c_1) and l represents an infinite series of consecutive natural numbers in the form (1, 2, 3, ...).

2 Proof of the conjecture

It is easy to check that any prime number (apart from 2 and 3, which are Sophie Germain primes) can be expressed in the form $6l + 1$ or $6s - 1$ ($l, s \in N$). Having that in mind, it is easy to conclude that numbers in the form $6l + 1$ could never be Sophie Germain primes since their safe primes are in the form

$$2(6l + 1) + 1 = 12l + 3 = 3(4l + 1),$$

and that is a composite number divisible by 3. Hence, the prime number that can potentially be a Sophie Germain prime must be in the form $6s - 1$ (here numbers 2 and 3 are ignored, and that obviously has no impact on the analysis that follows). The safe prime will then be in the form $6(2s) - 1$.

In the text that follows, a number in the form $6l + 1$ is denoted with *mpl*, while a number in the form $6s - 1$ is denoted with *mps* ($l, s \in N$).

Here we are going to present a two stage process that can be used for generation of Sophie Germain primes. In the first stage we are going to produce prime numbers by removing all composite numbers from the set of natural numbers. In the second stage, we are going to analyze the prime numbers themselves, as a potential generators of odd primes. In the second stage all prime numbers that create composite numbers are going to be removed. Basically, we are going to implement two stage recursive process. At the end, only the prime numbers in the *mps* form, that represent the Sophie Germain primes, are going to stay. It is going to be shown that their number is infinite. It is easy to check that all numbers in *mpl* form are going to be removed from the set, based on the analysis made at the beginning of this chapter.

STAGE 1

Prime numbers can be obtained in the following way:

First, we remove all even numbers (except 2) from the set of natural numbers. Then, it is necessary to remove the composite odd numbers from the rest of the numbers. In order to do that, the formula for the composite odd numbers is going to be analyzed. It is well known that odd numbers bigger than 1, here denoted by a, can be represented by the following formula

$$a = 2n + 1,$$

where $n \in N$. It is not difficult to prove that all composite odd numbers a_c can be represented by the following formula

$$a_c = 2(2ij + i + j) + 1 = 2((2j + 1)i + j) + 1. \quad (1)$$

where $i, j \in N$. It is simple to conclude that all composite numbers could be represented by product $(i + 1)(j + 1)$, where $i, j \in N$. If it is checked how that formula looks like for the odd numbers, after simple calculation, equation (1) is obtained. This calculation is presented here. The form $2m + 1$, $m \in N$ will represent odd numbers that are composite. Then the following equation holds

$$2m + 1 = (i_1 + 1)(j_1 + 1) \quad ,$$

where $i_1, j_1 \in N$. Now, it is easy to see that the following equation holds

$$m = \frac{i_1 j_1 + i_1 + j_1}{2}.$$

In order to have $m \in N$, it is easy to check that i_1 and j_1 have to be in the forms

$$i_1 = 2i \text{ and } j_1 = 2j,$$

where $i, j \in N$. From that, it follows that m must be in the form

$$m = 2ij + i + j. \quad (2)$$

When all numbers represented by m are removed from the set of odd natural numbers bigger than 1,

only the numbers that represent odd prime numbers are going to stay. In other words, only odd numbers that cannot be represented by (1) will stay. This process is equivalent to the sieve of Sundaram [7].

The numbers that are left after this stage are prime numbers. If we denote with $\pi(n)$ number of primes smaller than n, the following equation holds [8]

$$\pi(n) \approx \frac{n}{\ln(n)}.$$

From [8] we know that following equation holds

$$\pi(n) > \frac{n}{\ln(n)}, \quad n \geqslant 17. \tag{3}$$

This inequality will be useful in analysis that follows.

STAGE 2

Now, we should analyze numbers a that are left in observational space, or prime numbers themselves. With the exception of number 2 all other prime numbers are odd numbers. Since number 2 is Sophie Germain prime it will not be removed from the set. We are interested in removal of all numbers a that will create composite number when we generate number $2a + 1$. So, once more we are interested in removal of all numbers that generate composite odd numbers. So, once more we are going to implement (2) and remove all a in the form

$$a = 2ij + i + j. \tag{4}$$

That will leave us with prime numbers in *mps* form that represent the Sophie Germain primes. As it has been already explained, prime numbers in *mpl* form produce composite odd numbers divisible by 3, when formula $2 \cdot mpl + 1$ is applied on them, so they all are going to be removed in second stage.

Let us mark the number of twin primes with π_{SGP}. Also, let us define the number of numbers that is left after two consecutive implementations of Sundaram sieve as *pd2*. The numbers obtained after

recursive implementation of two Sundaram sieves are going to be called double primes. The second stage sieve that is identical to the first stage sieve can be obtained if the prime numbers, that are the only numbers left after the first stage, are lined up next to each other and then the numbers are removed from the exactly same positions like in the first stage. It is equivalent to implementation of Sundaram sieve on ordinals of prime numbers and removal of corresponding prime. In that case it is easy to understand that the following equation would hold ($n \in N$)

$$pd2(n) \approx \frac{\pi(n)}{\ln(\pi(n))} \quad , \tag{5}$$

where $pd2(n)$ represents the number of double primes smaller than some natural number n. Since the Sophie Germain primes are obtained by implementation of the Sundaram sieve in the first stage and sieve that is similar to Sundaram sieve in the second stage, it can be intuitively concluded that numbers π_{SGP} and $pd2$ should be comparable. However, in the case of generation of Sophie Germain primes the second stage sieve defined by (4) is not equivalent to the the first stage sieve since the second stage "Sundaram" sieve (defined by (4)) is applied on an incomplete set, that is depleted by previously implemented Sundaram sieve. Here, it will be shown that number of double primes $pd2$ is smaller than the number of Sophie Germain primes, π_{SGP}. In order to understand why it is so, we are going to analyze (2) and (4) in more detail.

It is not difficult to be seen that m in (2) and (4) is represented by the threads that are defined by odd prime numbers. For details see Appendix A. Now we are going to compare stages 1 and 2 step by step, for a few initial steps (analysis can be easily extended to any number of steps).

From the table, that follows, it can be noticed that threads defined by the same number in first and second stage will not remove the same percentage of numbers. The reason is obvious – consider for instance the thread defined by 3: in the first stage it will remove 1/3 of the numbers left, but in the second stage it will remove ½ of the numbers left, since the thread defined by 3 in stage 1 has already removed one third of the numbers (odd numbers divisible by 3 in observation space). So,

only odd numbers (in observational space) that give residual 1 and -1 when they are divided by 3 are left, and there is approximately same number of numbers that give residual -1 and numbers that give residual 1, when the number is divided by 3 (see Appendix A). Same way of reasoning can be applied for all other threads defined by the same prime in different stages.

Table 1 Comparison of the stages 1 and 2 for threads defined by a few smallest primes

Step	Stage 1	Step	Stage 2
1	Remove even numbers (except 2) amount of numbers left 1/2	1	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left 1/2
2	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left 2/3 of the numbers left in previous step	2	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left 3/4 of the numbers left in previous step
3	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left 4/5 of the numbers left in previous step	3	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left 5/6 of the numbers left in previous step
4	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left 6/7 of the numbers left in previous step	4	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left 9/10 of the numbers left in previous step
5	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left 10/11 of the numbers left in previous step	5	Remove numbers defined by thread defined by 13 (obtained for $i = 6$) amount of numbers left 11/12 of the numbers left in previous step
6	Remove numbers defined by thread defined by 13 (obtained for $i = 6$) amount of numbers left 12/13 of the numbers left in previous step	6	Remove numbers defined by thread defined by 17 (obtained for $i = 8$) amount of numbers left 15/16 of the numbers left in previous step

From Table 1 can be seen that in every step, except step 1, threads in the second stage will leave bigger percentage of numbers than the corresponding threads in the first stage. This could be easily understood from the analysis that follows:

- suppose that we have two natural numbers j, k such that $j - 1 \geq k$ ($j, k \in N$), then the following set of equations is trivially true

$$j + k - 1 \geqslant 2k$$

$$-j-k+1\leqslant -2k$$

$$jk-j-k+1\leqslant jk-2k$$

$$(j-1)(k-1)\leqslant (j-2)k$$

$$\frac{k-1}{k}\leqslant \frac{j-2}{j-1}$$

The equality sign holds only in the case $j = k + 1$. In the set of prime numbers there is only one case when $j = k + 1$ and that is in the case of primes of 2 and 3. In all other cases $p(i) – p(i - 1) > 1$, ($i > 1$, $i \in N$, $p(i)$ is i-th prime number). So, in all cases $i > 2$

$$\frac{p(i-1)-1}{p(i-1)}<\frac{p(i)-2}{p(i)-1} \quad .$$

From Table 1 (or last equation) we can see that bigger number of numbers is left in every step of stage 2 then in the stage 1 (except 1st step). From that, we can conclude that after every step bigger than 1, part of the numbers that is left in stage 2 is bigger than number of numbers left in the stage 1 (that is also noticeable if we consider amount of numbers left after removal of all numbers generated by threads that are defined by all prime numbers smaller than some natural number). From previous analysis we can safely conclude that the following equation holds

$$\pi_{SGP} > pd2 = \lim_{n \to \infty} pd2(n) \quad .$$

Having in mind (3) and (5), we can say that for some n big enough the following inequality holds

$$pd2(n) > \frac{\pi(n)}{\ln(\pi(n))}. \tag{6}$$

From (3) we know that previous inequality holds for $\pi(n) \geq 17$. It can be realized that n that is big enough is $n \geq 59$, since 59 is the 17th prime number. By checking its first derivative, it can be shown that the function $x/\ln(x)$ is monotonous rising function for $\ln(x) > 1$, or $x > e$ (e is Euler's number). Having that in mind, and equations (3) and (6), we can conclude that the following inequality holds ($n \geq 59$)

$$pd2(n) > \frac{\frac{n}{\ln(n)}}{\ln\left(\frac{n}{\ln(n)}\right)} = \frac{n}{\ln(n)(\ln(n) - \ln(\ln(n)))}.$$

Since it it easy to show that following holds

$$\lim_{n \to \infty} \frac{n}{\ln(n)(\ln(n) - \ln(\ln(n)))} = \infty,$$

then it is easy to understand that the following equation holds

$$pd2 = \lim_{n \to \infty} pd2(n) = \infty.$$

Now, we can safely conclude that the number of Sophie Germain primes is infinite. That concludes the proof.

Here we will state the following conjecture: for n big enough, number of Sophie Germain primes is given by the following equation

$$\pi_{SGP}(n) \sim 2C_2 \frac{\pi(n)}{\ln(\pi(n))},$$

where C_2 is twin prime constant [9]. Why it is reasonable to make such conjecture is explained in Appendix B. If we mark the number of primes smaller than some natural number n with $\pi(n) = f(n)$, where function $f(n)$ gives good estimation of the number of primes smaller than n, than $\pi_{SGP}(n)$, for n big enough, is given by the following equation

$$\pi_{SGP}(n) \sim 2C_2 \cdot f(f(n)).$$

If particular case $f(n) = Li(n)$, the following equation holds

$$\pi_{SGP}(n) \sim 2C_2 \cdot \int_2^n \left(\frac{d\pi(x)}{\ln(\pi(x))}\right) = 2C_2 \cdot \int_2^n \left(\frac{dx}{\ln(x)\ln\left(\int_2^x \left(\frac{dt}{\ln(t)}\right)\right)}\right).$$

For small number n, starting from the following formula for the prime numbers smaller than some

natural number n

$$\pi(n) \approx \frac{n}{\ln\left(\frac{n}{2}\right) - \frac{1}{3}},$$

a good estimation of Sophie Germain primes smaller than natural number n is given by the following formula

$$\pi_{SGP}(n) \approx (1 + 0.0129 \cdot \log(n)) \frac{n}{\left(\ln\left(\frac{n}{2}\right) - \frac{1}{3}\right)\left(\ln\left(\frac{n}{2}\right) - \ln\left(\ln\left(\frac{n}{2} - \frac{1}{3}\right)\right) + \frac{1}{3}\right)}.$$

This equation gives good approximation of the number of Sophie Germain primes smaller than natural number n, at least for the values of $n \leq 10^{14}$ (estimation for values $n = 10^3$ and n = 10^4 is actually correct).

References

[1] T. Agoh. On Sophie Germain primes, Tatra Mt. Math. Publ. 20(65) (2000), 65-73.

[2] M. Agrawal, N. Kayal, N. Saxena. PRIMES is in P, Ann. Of Math. 160(2)(2004), 781-793

[3] H.M. Edwards. Fermat's Last Theorem: A Genetic Introduction to Algebraic Number Theory, Springer, 2000.

[4] R.A.J. Matthews. Maximally periodic reciprocals, Bull. Inst. Math. Appl. 28 (1992), 147-148.

[5] V. Shoup. A Computational Introduction to Number Theory and Algebra, Cambridge University Press, 2009.

[6] W.-S. Tap, S.I. Yeo, S.-H. Heng, M. Henricksen. Security analysis of GCM for communication, Security and Communication Networks 7(5) (2014), 854-864.

[7] V. Ramaswami Aiyar. Sundaram's Sieve for prime numbers, The Mathematics Student, 2(2) (1934), 73.

[8] J.B. Rosser, L. Schoenfeld. (1962) Approximate formulas for some functions of prime numbers. Illinois Journal of Mathematics, 6(1), pp. 64-94.

[9] 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. 44, (1923), pp.1-70.

APPENDIX A.

Here it is going to be shown that m in (2) is represented by threads defined by odd prime numbers. Now, the form of (2) for some values of i will be checked.

Case i = 1: $m = 3j + 1$,

Case i = 2: $m = 5j + 2$,

Case i = 3: $m = 7j + 3$,

Case i = 4: $m = 9j + 4 = 3(3j + 1) + 1$,

Case i = 5: $m = 11j + 5$,

Case i = 6: $m = 13j + 6$,

Case i = 7: $m = 15y + 7 = 5(3j + 1) + 2$,

Case i = 8: $m = 17j + 8$,

It can be seen that m is represented by the threads that are defined by odd prime numbers. From examples (cases $i = 4$, $i = 7$), it can be seen that if $(2i + 1)$ represent a composite number, m that is represented by thread defined by that number also has a representation by the thread defined by one of the prime factors of that composite number. That can be proved easily in the general case, by direct calculation, using representations similar to (2). Here, that is going to be analyzed. Assume that $2i + 1$ is a composite number, the following holds

$$2i + 1 = (2l + 1)(2s + 1)$$

where (l, $s \in N$). That leads to

$$i = 2ls + l + s.$$

The simple calculation leads to

$$m = (2l + 1)(2s + 1)j + 2ls + l + s = (2l + 1)(2s+1)j + s(2l + 1) + l$$

or

$$m = (2l+1)((2s+1)j + s) + l$$

which means

$$m = (2l + 1)f + l,$$

and that represents the already exiting form of the representation of m for the factor $(2l + 1)$, where

$$f = (2s + 1)j + s.$$

In the same way this can be proved for (4) .

Note: *It is not difficult to understand that after implementation of stage 1, the number of numbers in residual classes of some specific prime number are equal. In other words, after implementation of stage 1, for example, all numbers divisible by 3 (except 3, but it does not affect the analysis) are removed. However, the number of numbers in the forms 3k + 1 and 3k + 2 (alternatively, 3k – 1) are equal. The reason is that the thread defined by any other prime number (bigger than 2) will remove the same number of numbers from the numbers in the form 3k + 1 and from the numbers in the form 3k + 2. It is simple to understand that, for instance, thread defined by number 5, is going to remove 1/5 of the numbers in form 3k + 1 and 1/5 of the numbers in form 3k + 2. This can be proved by elementary calculation. That will hold for all other primes and for all other residual classes.*

APPENDIX B.

Here, asymptotic density of numbers left, after implementation of the first and the second "Sundaram" sieve is going to be calculated. After first k steps of the first Sundaram sieve, after removal of all composite even numbers, density of numbers left is given by the following equation

$$c_k = \frac{1}{2} \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j)}\right),$$

where $p(j)$ is j-th prime number.

In the case of second "Sundaram" sieve the density of numbers left after the first k-steps is given by the following equation

$$c2_k = \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j) - 1}\right) = \prod_{j=2}^{k+1} \left(\frac{p(j) - 2}{p(j) - 1}\right).$$

So, if implementation of first sieve will result in the number of prime numbers smaller than n which we denote as $\pi(n)$, than implementation of the second sieve on some set of size $\pi(n)$ should result in the number of numbers $gp(n)$ that are defined by the following equation (for some big enough n)

$$gp(n) = r_{S2SI}(n) \cdot \frac{\pi(n)}{\ln(\pi(n))} \quad ,$$

where $r_{S2SI}(n)$ is defined by the following equation (k is the number of primes smaller or equal to sqrt(n))

$$r_{S2SI}(n) = \frac{c2_k}{c_k} = \frac{\prod_{p>2,\, p \le \sqrt{n}} \left(\frac{p-2}{p-1}\right)}{\prod_{p \le \sqrt{n}} \left(\frac{p-1}{p}\right)} = 2 \prod_{p>2,\, p \le \sqrt{n}} \left(\frac{p-2}{p-1}\right)\left(\frac{p}{p-1}\right) \approx 2C_2.$$

For n that is not big, $gp(n)$ should be defined as

$$gp(n) = f_{COR}(n) \cdot 2C_2 \cdot \frac{\pi(n)}{\ln(\pi(n))},$$

where $f_{COR}(n)$ represents correction factor that asymptotically tends toward 1 when n tends to infinity.