A proof of Sophie Germain primes conjecture

Marko V Jankovic

Proof of the Sophie Germain Primes Conjecture

Marko V. Jankovic

ARTORG Centre for Biomedical Engineering Research,
University of Bern, Switzerland

Abstract In this paper a proof of the existence of an infinite number of Sophie Germain primes is going to be presented. In order to do that, the basic formula for prime numbers was analyzed with the intention of finding out when this formula would produce a Sophie Germain prime and when not. The originally very difficult problem (in observational space) has been transformed into a simpler one (in generative space) that can be solved.

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.

Remark 1: *Prime numbers 2 and 3 are in a sense special primes, since they do not share some of the common features of all other prime numbers. For instance, every prime number, apart from 2 and 3, can be expressed in the form 6l + 1 or 6s- 1, where l, s ∈ N. In this paper most of the time we analyze prime numbers bigger than 3. It has to be said that both 2 and 3 are Sophie Germain primes, but that has no impact on the conclusion of this paper.*

Remark 2: *In this paper any infinite number 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). Here c_1 and c_2 are constants that belong to the set of natural numbers (c_2 can also be 0, 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

It is easy to check that any prime number (apart form 2 and 3) 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 $6s-1$. The safe prime will then be in the form $6(2s) - 1$.

We denote any composite number (that is represented as a product of prime numbers bigger than 3) with *CPN*5. A number in the form $6l + 1$ is marked with *mpl*, while a number in the form $6s - 1$ is marked with *mps* ($l, s \in N$). That means that any composite number *CPN*5 can be expressed in the form *mpl x mpl*, *mps x mps* or *mpl x mps*.

If we represent all composite numbers in *mps* form with $6k$ - 1 ($k \in N$) it must hold

$$k = \frac{CPN5 + 1}{6}. \tag{2.1}$$

Since *CPN*5 should be in the *mps* form, *CPN*5 can be generally expressed as a product *mpl* x *mps*, or

$$mpl = 6x + 1 \text{ and } mps = 6y - 1 (x, y \in N),$$

which leads to

$$CPN5 = mpl \text{ x } mps = 6(6xy - x + y) - 1, \tag{2.2}$$

or, due to symmetry

$$mpl = 6y + 1 \text{ and } mps = 6x - 1,$$

which leads to

$$CPN5 = mpl \text{ x } mps = 6(6xy + x - y) - 1, \tag{2.3}$$

If (2.2) or (2.3) is replaced in (2.1) forms of k that will not produce a Sophie Germain prime number will be obtained. Those forms are expressed by the following equations

$$k = (6x - 1)y + x \tag{2.4a}$$

$$k = (6x + 1)y - x \tag{2.4b}$$

where $x, y \in N$. These equations are equivalent (they will produce the same numbers) and can be used interchangeably.

Also, in order to have Sophie Germain pair (which consists of prime p and safe prime $2p+1$), a safe prime which is generated by $2k$ cannot be composite. If the safe number is composite the following equation must hold

$$k = \frac{CPN5 + 1}{6 \cdot 2},$$

(2.5)

where *CPN*5 is a composite number in the *mps* form. Using the same analysis as in the previous case and replacing for instance (2.2) in (2.5), additional cases in which k cannot be used to produce Sophie Germain prime pairs are obtained. They are defined by the following equation

$$k = \begin{cases} (6x+1)y - \frac{x}{2}, & x \text{ is even} \\ (6x+1)y - 3x - \frac{x+1}{2}, & x \text{ is odd} \end{cases}$$

(2.6a)

where $x, y \in N$. Alternatively, it is possible to use the equation (2.3) and replace it in (2.5) and then the following equation holds

$$k = \begin{cases} (6x-1)y + \frac{x}{2}, & x \text{ is even} \\ (6x-1)y - 3x + \frac{x+1}{2}, & x \text{ is odd} \end{cases}$$

. (2.6b)

A different equation that produces the same numbers as the equation (2.6a) is obtained.

Equations (2.4a) and (2.6a) (and other alternatives like (2.4b) and (2.6b)) give a sufficient and necessary condition for k, so that it cannot be used for the generation of prime pairs in the form $(p, 2p+1)$. In order to prove that there are infinitely many prime pairs in the form $(p, 2p+1)$ it is necessary to prove that infinitely many *natural numbers s* ($s \in N$) exist, that cannot be expressed in the form (2.4a) or (2.6a).

First, the forms of (2.4a, 2.6a) for some values of x are checked.

Case x=1: $k = 7y - 4, k = 5y + 1$

Case x=2: $k = 13y - 1, k = 11y + 2$

Case x=3: $k = 19y - 11, k = 17y + 3$

Case x=4: $k = 5(5y) - 2, k = 23y + 4$ *(here we should have the equivalence of the equations (2.6a) and (2.6b) in mind)*

Case x=5: $k = 31y - 18, k = 29y + 5$

Case x=6: $k = 37y - 3, k = 5(7y+1) + 1$

Case x=7: $k = 43y - 25, k = 41y + 7$

Case x=8: $k = 7(7y) - 4, k = 47y + 8$

It can be seen that k is represented by the threads that are defined by prime numbers bigger than 3. From examples (cases x=4 and $x = 8$), it can be seen that if $(6x - 1)$ or $(6x + 1)$ represent a composite number, k that is represented by the thread defined by that number, is also represented by the thread defined by one of the prime factors of that composite number. This can be easily proved by direct calculation in the general case, where representations similar to (2.2) and (2.3) are used. Here only one case is going to be analyzed. All other cases can be analyzed analogously. In this case we assume

$$(6x - 1) = (6l + 1)(6s - 1),$$

where ($l, s \in N$). Thus, the following equation holds

$$x = 6ls - l + s.$$

Considering that and using the following representation of k that includes the form $(6x - 1)$

$$k = (6x - 1)y + x,$$

the simple calculations leads to

$$k = (6l + 1)(6s - 1)y + 6ls - l + s = (6l + 1)(6s - 1)y + s(6l + 1) - l,$$

or

$$k = (6l + 1)\big((6s - 1)y + s\big) - l,$$

which means

$$k = (6l + 1)f - l,$$

and these values of k are also represented by the thread defined by $(6l+1)$, where

$$f = (6s - 1)y + s.$$

Here the equivalency of the equations (2.4a) and (2.4b) is used. It can be seen that all patterns for k are represented by the threads defined by prime numbers bigger than 3. Now, proof that the number of natural numbers s, that cannot be represented by the models (2.4a) and (2.6a) is infinite, will be provided.

When all numbers that can be represented in the form

$$5y + 1,$$

are removed from the set of natural numbers N, it can be seen that a ratio of $r_1 = 1/5$ of all natural numbers is removed. The ratio $c_1 = 1 - 1/5 = 4/5$ of all natural numbers cannot be represented by those two patterns and they still contain some numbers that could be used for representation of Sophie Germain primes.

What does that actually mean? The proper interpretation of this result is: All natural numbers can be represented by 5 threads: $5z$, $5z$-1, $5z$-2, $5z$-3 and $5z$-4 ($z \in N$). It means that all *natural numbers* that cannot be represented by $5y+1$ (that is equivalent to $5z - 4$, for $z > 1$) can only belong to the threads that are in the form $5z$ - *3*, $5z$- 2 or $5z$ - 1 and $5z$. That means that

there are four threads that potentially contain natural numbers that can be used for the generation of Sophie Germain primes and cannot be represented by $5y+1$.

If in addition, the natural numbers in the form

$$7y - 4,$$

are removed, then the ratio of removed numbers can be calculated by the following equation (checking that every removed number is calculated only once; basically, the formula for calculation of the probability of the occurring of two events that are not mutually exclusive is applied: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$)

$$r_2 = r_1 + \frac{1}{7} - \frac{1}{7} \times r_1 = r_1 + \frac{1}{7}(1 - r_1) = \frac{1}{5} + \frac{1}{7}\left(1 - \frac{1}{5}\right) = \frac{11}{35}$$

The ratio of all natural numbers that can still be used for the generation of Sophie Germain primes, is defined by

$$c_2 = 1 - r_2 = 1 - r_1 - \frac{1}{7}(1 - r_1) = \left(1 - \frac{1}{7}\right) \times c_1 = \frac{4 \times 6}{5 \times 7}.$$

Again – the proper interpretation of this result is: All natural numbers can be represented with 35 threads in the form $35z\text{-}i$ ($z \in N$, $i \in \{0, 1, 2, \ldots, 34\}$). From previous step it is known that the 7 threads defined by $5*(7z\text{-}j) + 1$, where $z \in N$, $j \in \{0, 1, 2, \ldots, 6\}$, do not contain the numbers that can be used for the generation of Sophie Germain primes. From the current step it is known that 5 threads defined by $7*(5z\text{-}j)\text{-}4$, where $z \in N$, $j \in \{0, 1, 2, 3, 4\}$, do not contain any numbers that can be used for the generation of Sophie Germain primes. However, these formulas, from the first two steps, produce some threads that overlap:

$$35z - 4 = 5(7z - 1) + 1 = 7(5z - 0) - 4.$$

That leaves us with 24 threads that potentially contain numbers that can be used for the generation of Sophie Germain primes and cannot be represented by $5y+1$ and $7y\text{-}4$.

Now, we denote prime numbers bigger than 3 as $p5$, where $p5(1) = 5$, $p5(2) = 7$ and so on. After step n the ratio r_n of all numbers removed and ratio the c_n of all numbers still potentially available for the generation of Sophie Germain primes, are obtained. In the step $n + 1$ we will have

$$r_{n+1} = r_n + \frac{1}{p5(n+1)} - \frac{1}{p5(n+1)} \times r_n.$$

After a few elementary calculations, the following equation is obtained

$$r_{n+1} = r_n + \frac{1}{p5(n+1)} \times (1 - r_n).$$

Now, the following equation holds

$$c_{n+1} = 1 - r_{n+1} = 1 - r_n - \frac{1}{p5(n+1)} \times (1 - r_n) = c_n - \frac{1}{p5(n+1)} \times c_n$$

or

$$c_{n+1} = \frac{p5(n+1) - 1}{p5(n+1)} \times c_n. \tag{2.7}$$

Equation (2.7) can also be written in the following form

$$c_{n+1} = \frac{\prod_{j=1}^{n+1}(p5(j)-1)}{\prod_{j=1}^{n+1} p5(j)}. \tag{2.8}$$

Hence, after n+1-st step in which a thread defined by n+1-st $p5$ is removed, **we know that $\prod_{j=1}^{n+1}(\boldsymbol{p5(j)-1})$ threads that potentially contain numbers that can be used for the generation of Sophie Germain primes exists**.

If the process is continued until all possible patterns (defined by (2.4a) and (2.6a) related to all prime numbers bigger than 3 (and that is an infinite number)) are removed, a number C can be defined by the following equation

$$C = \lim_{n\to+\infty} \prod_{j=1}^{n} (p5(j) - 1). \tag{2.9}$$

It can easily be concluded that C is an infinite number. We know that C represents the number of threads that contain natural numbers that **cannot be represented by (2.4a), and (2.6a)**. Since the set of natural numbers is closed for multiplication, it means that every one of those threads contains at least one number and that means that the number of natural numbers that cannot be represented by equations (2.4a) and (2.6a) is infinite.

That completes the proof that the number of Sophie Germain primes is infinite.

Note: Using the sieve of Erathostenes [7] and same line of reasoning it is possible to prove that the number of prime numbers is infinite.

Note: Using the sieve presented in [8] for the generation of twin primes and same line of reasoning it is easy to prove that the number of twin prime numbers is infinite. Same can be done for cousin primes.

3 Estimation of the number of Sophie Germain primes

In chapter 2, a sieve that was presented by equations (2.4a) and (2.6a) was used for generation of Sophie Germain primes. In this chapter the other for for the generation of Sophie Germain primes is going to presented. This will help to give a reasonable estimation of the number of Sophie Germain primes that smaller that natural number n.

Prime numbers can be obtained if even numbers (except 2) are removed from the set of natural numbers, first. Then, it is necessary to remove the composite odd numbers from the set of all odd numbers. In order to do that, the formula for odd numbers and the formula for the composite odd numbers are going to be analyzed. It is well known that odd numbers a (bigger than 1) 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 presented by the following formula

$$a_c = 2(2ij + i + j) + 1 = 2\big((2j + 1)i + j\big) + 1, \tag{3.1}$$

where $i, j \in N$. It is simple to conclude that all composite numbers could be represented by the product $(i+1) * (j+1)$, where $i, j \in N$. If it is checked how that formula looks like for odd numbers, after simple calculation, equation (3.1) 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

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

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

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

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

$$m = 2ij + i + j.$$

When all numbers represented by m are removed only the numbers that represent odd prime numbers are going to stay. The odd numbers that cannot be represented by the (3.1). This process is equivalent to sieve of Sundaram [9].

Let us mark the numbers used for creation of odd prime numbers with $m2$. Those are the numbers that are left after implementation of Sundaram sieve. The number of those numbers is equivalent to the number of prime numbers smaller than natural number n (we mark it as $\pi(n)$), and can be well approximated by the following equation

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

Now, the following question can be asked:

hat will happen when Sundaram's sieve is implemented once more, now on numbers m2?

The answer is that in the second step the only numbers that are not going to be removed are prime numbers that are used for generation of some *mps* odd prime numbers. All of these numbers are in the *mps* form and actually represent Sophie Germain primes (it has been already explained that prime numbers in *mpl* form produce composite odd numbers divisible by 3, when formula 2**mpl*+1 is applied on them) . Since the same method is applied in the second step, intuitively can be concluded that the number of numbers left after the second Sundaram sieve, should be comparable to the following number *msg*:

$$msg = \frac{\pi(n)}{\left(\ln\left(\frac{\pi(n)}{2}\right) - \frac{1}{3}\right)}.$$

Let us mark a number of Sophie Germain primes smaller than natural number n with $\pi SGP(n)$. Good approximation of the $\pi SGP(n)$ is given with the following equation

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

This equation gives good approximation of the number of Sophie Germain primes smaller than natural number n, at least for the values of n till 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. Yap, S.L. Yeo, S.-H. Heng, M. Henricksen. Security analysis of GCM for communication, Security and Communication Networks 7(5) (2014), 854-864.

[7] D. Wells. *Prime Numbers: The Most Mysterious Figures in Math.*, Hoboken, NJ: John Wiley & Sons, Inc., 2005, pp. 58–59.

[8] M. Jankovic. Proof of the Twin Prime Conjecture (Together with the Proof of Polignac's Conjecture for Cousin Primes). 2020. ⟨hal-02549967⟩

[9] V. Ramaswami Aiyar. Sundaram's Sieve for Prime Numbers, The Mathematics Student. **2** (2): 73, (1934).