

Bounds in Twin Prime Distribution

Serge Dolgikh

National Aviation University

Abstract: An efficient numerical method of traversal of twin prime pairs was applied to verify bounds on the distribution of twin primes at an arbitrary distance. It was shown that if proven formally, the bounds will be sufficient for the proof of the twin prime conjecture. Heuristic arguments in favor of the bounds obtained with numerical modeling are given and connections with other problems in distributions of prime sequences briefly discussed.

1 Introduction

This work in the domain of modeling and analysis of twin prime distribution [1] follows the approach of prime modulo cycles (prime harmonics) described in [2] of which a brief outline will be given here. We consider the positions of odd multiple of 3, k as the keys of possible twin prime pair that follows: $p = (k + 2; k + 4)$. Such positions are called twin prime candidates (tp-candidates). Each candidate position potentially identifies a twin prime pair, unless it is "spoiled" by a collision with a prime sequence $h_p(x) = m \times p$, or prime harmonic of prime order p , in which case one of the elements in the pair defined by the candidate position is divisible by p and the candidate is eliminated. Candidate positions that have not been eliminated by any prime harmonics become true twin primes. This framework leads to an iterative method of calculation of tp-candidate positions remaining after arbitrary number of prime harmonics in an arbitrary range, by traversing the positions of possible collisions with the candidates of the previous iteration. The method proved to be fast and efficient, with computations complexity of $\log \log p$.

Based on the results in [2] for a given order p , the candidates within certain range $d(p) = p^2$ cannot be eliminated at orders higher than p and therefore define true twin primes. A simple explanation is that any composite number $x < p^2$ must have a factor $f < p$ and therefore, would have been "eliminated" by the lower prime harmonic of that order. On this basis, the method can be used to calculate the number of true twin primes at an arbitrary range with computations complexity of $a \log \log p$.

Application of the method to prime orders up to the first 1,000 allowed to obtain seemingly stable bounds, both upper and lower, on the number of twin primes in the verified ranges that can be used as a basis for theoretical analysis. It is shown that if proven formally, the lower bound would be sufficient for the proof of the twin prime conjecture.

2 Twin Prime Traversal

In this section the method of twin prime traversal is described and results for a range of prime orders presented.

Preliminary notes:

We consider the set of positive odd integers as defined in [2]. A relative position of an integer x to the center of prime set cycle is therefore $(x + 1)/2$.

2.1 Twin Prime Traversal Method

The method is based on iterative traversing the sequence of tp-candidate positions in a given range that are known with collision positions of prime harmonics at iteration orders.

Step 1

In the first step, an array K with an arbitrary length d is filled with ones, indicating potential candidate positions from 3 and upwards. The maximum candidate position contained in K is therefore, $3 + 6 \times (d - 1)$. In the initial (zero) iteration, before any prime harmonics are applied none of candidate positions were eliminated.

Step 2

Beginning with the first iteration $p = 5$, for a given prime order p possible collision positions C_p with tp-candidates defined in Step 1 are calculated as $(p + 1)/2 + k \times p - 1$, 2 i.e., $((2k - 3)/2, (2k - 1)/2)$, up to the maximum $k_m(p)$, dictated by the range d .

The positions of possible collisions $C_p(k)$ are traversed sequentially, and in the case of an actual collision, i.e. $3 \mid C_p(k)$, the value of the candidate sequence K at the corresponding position is switched to 0, indicating that the candidate position has been eliminated by a collision with prime harmonic h_p .

Step 3

The process described in Step 2 is repeated iteratively for prime orders from 5 to the given order p . Computational complexity of the method can be estimated straightforwardly as:

$$N = \sum_{l=5}^p 2 \times k_m(l) \sim \sum_{l=5}^p \frac{6d}{l} \sim A \log \log p \quad (1)$$

After the process is complete, the number of tp-candidates in the range d remaining after prime harmonic up to p can be found immediately by summing K up to the candidate position c_m corresponding to d , $2 + 3c_m = d$. Importantly, as shown in [2], Lemma 2 and discussed earlier, for a given order p the candidates in the range p^2 cannot have

collisions with prime harmonics of orders higher than p and therefore, must identify true twin primes. This observation allows to find the actual number of twin primes at arbitrary distance d by applying the traversal process up to order $p = \sqrt{d}$.

$$N_{lp}(d) = \sum_0^{c_m(d)} K_{p=\sqrt{d}} \quad (2)$$

The resulting actual number of twin primes can be used to verify bounds produced from heuristic analysis [2] that will be discussed in the following sections.

The method proved to be efficient and fast, with the results up to several hundred of prime orders obtained on regular hardware in seconds of runtime.

2.2 Numerical Evaluation of TP-candidate Bounds

The analysis of interaction of prime modulo cycles produced several heuristic bounds on the number of tp-candidates at an arbitrary range d , (15)-(20) [2] that can be evaluated with the modeling method described in the previous section. Specifically we will consider the χ_2 bound, defined as:

$$N_{\chi}(p) = d \times \chi_2(p) = d \prod_{l=5}^p \left(1 - \frac{2}{l}\right) \quad (3)$$

and its logarithmic approximation, N_{lg} :

$$N_{lg}(p) \propto \frac{d}{\log^2 p} = \frac{A_2 d}{\log^2 p} \quad (4)$$

The results of numerical modeling in the first 500 prime orders (up to $p = 3617$) are illustrated in Figure 1.

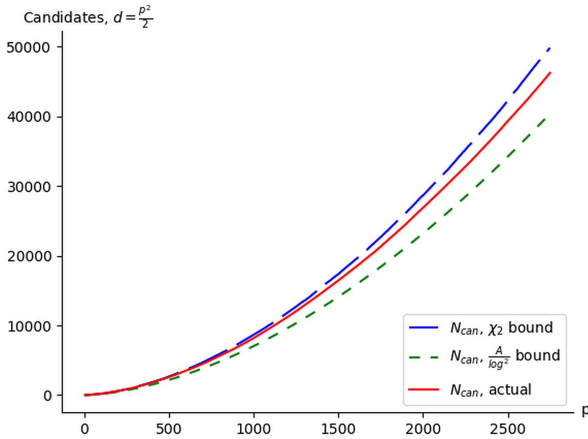


Figure 1: Twin prime bounds vs actual

As can be observed from the results of the modeling in Figure 1 and Table 1, the χ_2 bound, that is the exact mean density of the number of tp-candidates at order p in the prime set cycle of the same order, produces an estimate somewhat higher than the actual number of candidates in

the range and can be hypothesized to represent an upper bound. On the other hand, the logarithmic bound N_{lg} (with the constant $A_2 = 2/3$) was consistently below the actual number and can be conjectured to represent the lower bound. It can be noted also that the gap between the logarithmic bound and the actual count displays monotonous growth pointing at the possibility of a positive term of the next order, possibly, $\sim \frac{1}{\log^3 p}$.

In Table 1 the results of numerical modeling with the traversal method defined in Section 2.1 are presented for the number of twin primes given by the presumably, upper and lower bounds, χ_2 and \log^2 and the actual number obtained with the numerical method in the range p^2 . An

Table 1: Twin prime bounds at ranges $d = p^2$

Prime order (ordinal)	χ_2 bound	\log^2 bound	Actual
19 (8)	14	14	20
101 (26)	191	160	209
151 (36)	364	303	385
199 (46)	565	473	588
251 (54)	838	691	834
307 (63)	1175	963	1162
547 (101)	3101	2521	2986
863 (150)	6700	5459	6376
1229 (201)	12317	9998	11673
1597 (251)	19374	15705	18231
1987 (300)	28380	22933	26591
2371 (351)	38552	31185	35863
2753 (402)	49951	40472	46345
3187 (451)	64763	52288	59806
3617 (506)	80766	65285	74296

effort is under way to verify results presented in this section to significantly higher ranges of prime orders. If confirmed in the further studies, the bounds pointed by numerical modeling can provide essential input to formal theoretical analysis of bounds in twin prime and other prime sequence distributions as discussed in the following sections.

3 Theoretical Approaches in Twin Prime Distribution

The analysis of prime harmonic interactions in [2] produced χ_2 distribution of tp-candidates, that is based on the assumption of uniformity of prime harmonic intersections at the higher prime orders. As the results of numerical modeling in Section 2.2 show, this estimate is somewhat higher than the actual number of twin primes and can

be conjectured to represent an upper bound on the actual distribution $N_{lp}(d)$. As has been pointed out, the challenge in establishing rigorous bounds on the number of tp-candidates at an arbitrary distance stands in estimating the number of intersections between prime harmonics at ranges significantly shorter than full prime set cycle. Before turning to possible approaches in this direction we consider the conditions on the lower bound on the number of tp-candidates that would be sufficient for the conjecture to hold true.

3.1 Sufficient Bound Condition

Lemma 1 (Sufficient bound condition). *A lower bound on the number of tp-candidates in the collision-protected range of an infinite sequence of prime orders S that is unbounded from above is sufficient for the twin prime conjecture to be true.*

Proof. First, we will note that the number of tp-candidates in the protected range, $N_{kan}(p)$ must be non-decreasing with p as all candidates in the protected range at a given order p cannot have collisions with higher harmonics and therefore, the number of candidates in that range at any order $q > p$ cannot be lower: $N_{kan}(q) \geq N_{kan}(p)$.

Now suppose there exists the last twin prime position c_{max} and the set of twin primes is finite with the cardinality of N_{max} . We will define the order p_{max} as the first prime order in S , such that $(p^2 - 2)/2 > c_{max}$. Clearly, p_{max} exists and the number of tp-candidates in the collision-protected range of p_{max} is N_{max} .

Then, by the assumption of the lemma, a bound $b(p)$ exists such that:

$$\begin{aligned} b(p_{max}) &\leq N_{max} \\ \exists \text{ prime } q \in S : b(q) &> N_{kan}(p_{max}) = N_{max} \\ b(q) &\leq N_{kan}(q) \end{aligned}$$

Then, $N_{max} = N(p_{max}) < b(q) \leq N_{kan}(q)$ and there exists a tp-candidate $c(q)$ in the collision-protected range of prime order q that is greater than the last twin prime position: $c(q) > c_{max}$. Then, by Lemma 2, [2] $c(q)$ is also protected from collisions with the harmonics of orders higher than q , and therefore must be a true twin prime.

The assumption of the existence of the maximum twin prime position is contradictory, and the condition of the lemma must be sufficient for the proof of the twin prime conjecture. \square

Evidently, a lower bound that is monotonous and unbounded would satisfy the condition of Lemma 1 and will be sufficient for the proof of the conjecture. This is true for the logarithmic bound $N_{lg}(p)$ discussed in the previous section that at the collision-protected range of order p would translate to:

$$N_{lg}(p) = \frac{A_2 p^2}{2 \log^2 p} = \frac{1}{3} \frac{p^2}{\log^2 p} \quad (5)$$

leading to the corollary of Lemma 1 of the logarithmic bound:

Corollary 1 (Sufficiency of logarithmic bound). *The logarithmic bound on the number of tp-candidates in the ranges that include collision protected zone p^2 $N_{kan}(p, d) \propto \frac{d}{\log^2 p}$ is sufficient for the twin prime conjecture to be true.*

It can be concluded then that if the validity of the logarithmic bound could be established rigorously, it would be sufficient for the proof of the infiniteness of the set of twin primes, and provide a lower estimate for the number of twin primes in a range of an arbitrary size.

3.2 Heuristic Argument for the Logarithmic Bound

An heuristic argument supporting logarithmic bound on the number of candidates $N_{lg}(d)$ can be put forward based on the analysis of intersections of prime harmonics. As was shown, the condition of intersection, or overlap between prime harmonics p , p_l splits into four separate periodic strands with the period of $3 \times p \times p_l$ (equation (12), Section 3.5).

Two of the strands are aligned i.e. have the same value of the collision interval offset, whereas two remaining ones are misaligned with one of the harmonics "ahead" or "behind" the other, producing exactly same number of intersections in the period cycle. Then the overlap conditions can be written as:

$$\begin{aligned} (2k+1)p &= (2k_l+1)p_l \\ (2k+1)p &= (2k_l+1)p_l + 2 \\ (2k+1)p &= (2k_l+1)p_l - 2 \end{aligned} \quad (6)$$

If intersections of a harmonic of a certain order p with lower ones, p_l are analyzed then its interval number k in the above can be considered fixed, and the first condition describes the two aligned strands, while the other two, the misaligned.

Now, from analyzing the first condition, it can be concluded that an intersection of h_p with the lower harmonics at the interval position identified by k is not possible if and only if the interval factor, $2k+1$ is coprime with all prime harmonics $p_l < p$.

Indeed, it is odd by definition and must be coprime with any other prime order so (6), top, cannot be satisfied, as p and p_l are coprime. Finally, it cannot divide 3 by the Corollary 3 (third intervals), [2].

One can conclude then that the number of actual collisions of the harmonic h_p with the candidate positions D_3 , that did not intersect with lower harmonics and thus need to be counted in the total number of collisions in the range d can be estimated as the number of integers in the effective range $d_e = 2d/p$ that are coprime with $p_l < p$.

This estimate is known and can be obtained from [3]:

$$N_{col}(p) = \frac{d_e \omega(u) - p}{\log p} + O\left(\frac{d_e}{\log^2 p}\right) \quad (7)$$

where $\omega(u)$ is Buchstab's function [4], $u = \frac{\log d_e}{\log p}$. The total number of collisions in a given range will be given by the sum of (7) over all prime harmonics p_l up to p :

$$N_{col}(d, p) = \sum_{p_l=5}^p N_{col}(d, p_l) \quad (8)$$

For most harmonics in this range, $d_e = p^2/p_l \gg p_l$ and the second term can be neglected, producing the estimate:

$$N_{col}(d, p) = A p^2 \sum_{p_l=5}^p \frac{1}{p_l \log p_l} + O \quad (9)$$

where p can be considered a constant.

Now, as discussed earlier, (9) cannot be the correct estimate of the number of collisions because it does not account for the misaligned intersections that can produce the same number of intersections with the lower harmonics as the aligned ones, that were considered in the estimate (9). However, if it could be assumed that the interval factors of h_p that are coprime with misaligned harmonics (6), 2 and 3, were distributed in the set of intervals satisfying the aligned condition with the same density, then the same argument could be applied for the number of interval factors that are coprime to both aligned and misaligned cases with the effective range given by (7), i.e.:

$$N_{col}(p, all) \sim \frac{N_{col}(p, aligned)}{\log p_l} \sim \frac{d_e}{\log^2 p_l} \sim \frac{1}{p_l \log^2 p_l}$$

A careful summation over $p_l < p$ then would yield the bound on the number of collisions at d ,

$$N_{col}(p, d) = a d - \frac{A d}{\log^2 p} + o\left(\frac{d}{\log^2 p}\right) \quad (10)$$

that is compatible, as can be shown straightforwardly, with the logarithmic bound $N_{lg}(d)$ observed in numerical modeling.

3.3 A Double Coprime Sieve

Let us return to the coprime condition on the interval factor discussed in Section 3.2. This condition can be seen as a sieve on the interval positions at the given prime order $p_l < p$. Clearly, a "clean" collision with a candidate position in D_3 happens at an p_l interval position k if none of conditions in (6) are satisfied at k .

If one considers the set $F(p_l, d)$ of all multiples of $p_s < p_l$ within the range, then the conditions (6), 2 and 3 signify that prime harmonics h_{p_l} at k "avoids" intersections with F as well as its shifted image $F \pm 2$. Then, the set of interval positions associated with clean collisions $N_{cl}(p_l, d)$ can be obtained as a result of sequential application of sieves F and $F \pm 2$ to the set of all collision ranges of the harmonic h_p at a given range.

It can be shown ([2], (12)) that the number of intersections

of h_p with F and $F \pm 2$ at a sufficiently large range is exactly the same. If an argument can be made that the sieves operate essentially independently, that is,

$$N_{cl}(p_l, d) = (F \pm 2) \otimes F N_{pos}$$

where N_{pos} is the set of collision ranges of order p in the range d . Then, application of the second sieve would produce a result similar to (7), i.e. behave as:

$$N_{cl}(p_l, d) \sim \frac{d_{eff} \tilde{\omega}(p_l, d)}{\log p_l} \quad (11)$$

where $\tilde{\omega}(p_l, d)$ is a slow varying function, and d_{eff} the product of the application of the first sieve,

$$d_{eff} \sim \frac{d}{p_l \log p_l}$$

Then,

$$N_{cl}(p_l, d) \sim \frac{d}{p_l \log^2 p_l}$$

and the bound (10) on the number of collisions up to order p at d , $N_{col}(p, d)$ could be substantiated, again leading via summation over $p_l \leq p$ to the logarithmic bound (4) on the number of tp-candidates in the given range.

4 Conclusion

The method of twin prime traversal described in this work can produce accurate results on the number of twin prime pairs at distances of arbitrary size and can be used in evaluation of hypotheses in twin prime distribution. Based on the results obtained with numerical modeling with the method, possible upper and lower bounds on the number of twin primes at arbitrary ranges were identified.

Heuristic arguments may point a direction to the rigorous proof of the conjectured bounds, and possibly, other essential questions in twin prime distribution, such as the maximum gap bounds that were established in prime distribution ([5, 6]).

It is expected that interval traversal methods based on prime harmonic analysis can be extended to investigation of distributions of prime sequences of other types [7] that will be the subject of a future study.

This work used prime and twin prime data provided by PrimePages, University of Tennessee Martin [8].

References

- [1] Tietze, H.: Prime numbers and prime twins. Famous Problems of Mathematics: Solved and Unsolved Mathematics Problems from Antiquity to Modern Times. Graylock Press, 1–20, (1965)
- [2] Dolgikh, S.: Prime harmonics and twin prime distribution. Cambridge Open Engage, doi:10.33774/coe-2020-c9x75-v2 (2020)

- [3] Tenenbaum, G.: Introduction to Analytic and Probabilistic Number Theory. Cambridge University Press, Section III.6.2 (1995)
- [4] Hildebrand, A.: Bukhstab function. Encyclopedia of Mathematics, EMS Press (2001)
- [5] Nagura, J.: On the interval containing at least one prime number. Proceedings of the Japan Academy, Series A28(4) (1952) 177–181
- [6] Schoenfeld, L.: Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. Mathematics of Computation, **30 (134)** (1976) 337–360
- [7] Dolgikh, S.: Prime sequences and distributions. Cambridge Open Engage. doi:10.33774/coe-2020-g9d1m (2020)
- [8] PrimePages, UTM: The First 1000 Primes. The First 100,000 Twin Primes.