Adam Mickiewicz University in Poznań
Faculty of Mathematics and Computer Science

Master's thesis

On the twin prime conjecture

Hipoteza o liczbach pierwszych bliźniaczych

Tomasz Buchert

Supervisor
prof. dr hab. Wojciech Gajda

Poznań 2011

Poznań, dnia ....

Oświadczenie

Ja, niżej podpisany **Tomasz Buchert**, student Wydziału Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu oświadczam, że przedkładaną pracę dyplomową pt.: **On the twin prime conjecture**, napisałem samodzielnie. Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałem z pomocy innych osób, a w szczególności nie zlecałem opracowania rozprawy lub jej części innym osobom, ani nie odpisywałem tej rozprawy lub jej części od innych osób.

Oświadczam również, że egzemplarz pracy dyplomowej w formie wydruku komputerowego jest zgodny z egzemplarzem pracy dyplomowej w formie elektronicznej.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

....

Contents

Chapter 1

Introduction

1.1 Prime numbers

The prime numbers are mysterious objects. First few of them are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots$$

In Carl Sagan's novel *Contact*, the extraterrestrial race used them to write a message for humans, a message that was easily distinguishable from random noise and was a sign of intelligence. It seems that the prime numbers are universal objects, an idea that must come to minds of sufficiently intelligent beings. Also, we do not know any natural, physical phenomena generating prime numbers. At first, there is no apparent structure in the prime numbers. They seem to appear in a random fashion, constantly popping out from the natural numbers. It is not even known beforehand, if there are infinitely many of them. Of course, if the set of prime numbers would be finite, they probably would not be interesting anymore. Fortunately we know that this is not the case.

There are deep patterns in the behavior of prime numbers, patterns very intricate and subtle. They become apparent when we stop to look at each of them separately, but start to see the prime numbers as a whole entity. Then the patterns emerge in a fruitful process: the number of primes below a given bound seems to be a conceivable function, the average number of divisors and prime divisors seems to be not random after all, and so on. Instantly, the prime numbers show a deep connection to the mathematical analysis, theory of probability, topology and other fields in mathematics.

Until very recently, the number theory was considered only a pure branch of mathematics, with virtually no practical application in the real life. It is Hardy who said in his *A Mathematician's Apology*: "No one has yet discovered any warlike purpose to be served by the theory of numbers or relativity, and it seems unlikely that anyone will do so for many years." A sensible statement for a pacifist, but, as we know now, this turned out to be wrong. The number theory is practical: every day millions of people use RSA encryption scheme and complicated ciphers without even knowing about it. Finally, probably to Hardy's despair, the number theory found it's way to the military too.

People who approach the number theory for the first time, may be startled how simple the problems seem to be. For example, one may ask what is difficult in the following problem: find positive integers x, y, z and $n \geq 3$, such that

$$x^n + y^n = z^n.$$

Of course, such problems only *seem* to be simple. The history has proven that the problem above is tremendously difficult and was tantalizing the mathematical community for few centuries. As we know, only in 1995 Andrew Wiles proved that there are no numbers satisfying the equation above. It's rather common knowledge that short questions (e.g., Does God exist? What is love?) may have complicated answers or no satisfactory answer at all. Yet, people tend to think that in this particular case it is not true, and for that reason Fermat's Last Theorem was attracting so many amateurs (Fermat himself was a successful, yet not professional mathematician).

There is no way to distinguish hard problems from the easy ones. As we know, the equation

$$x^2 + y^2 = z^2$$

was solved in antiquity, just like the fact that the number of primes is infinite. But we may ask more questions: how many primes are there below a certain number, are there infinitely many prime numbers in arithmetical progressions, are there infinitely many pairs of primes whose difference is exactly two? The first two questions are already answered, the last one remains unsolved and we will talk about it later. The novelty of the approach to these problems was to use the complex analysis, a field of mathematics normally not attributed to the number theory. Probably the most famous child that this marriage gave birth to is the Riemann Hypothesis. It asks if all the non-trivial complex zeros of the Riemann Zeta Function lie on a, so called, *critical line*. This question seems to be more complicated than the statement of Fermat's Last Theorem. Also, there is no obvious connection with other number theory problems. However, the positive answer to the Riemann Hypothesis would change the number theory and mathematics as we know it.

The third problem above, i.e., are there infinitely many pairs of primes whose difference is exactly 2, is known as the Twin Prime Conjecture. The first few twin prime pairs are

$$(3,5), (5,7), (11,13), (17,19), \ldots$$

There is a strong empirical and heuristic evidence, as we will see, that this conjecture is true. One must be careful, however, to not become biased by it. The Mertens' Conjecture postulated that

$$\left| \sum_{k=1}^{n} \mu(k) \right| < \sqrt{n}$$

for any n. This claim has been checked for $n < 10^{14}$, but an indirect proof showed that it is false. Interestingly, the proof of Mertens' Conjecture would imply Riemann Hypothesis.

The Twin Prime Conjecture is possibly the most basic question one may ask, after they are satisfied with the Prime Number Theorem. There are probably no direct, practical conclusions that can be drawn from the Twin Prime Conjecture. But, just as was in the case of Fermat's Last Theorem, research toward the unproven conjecture usually yields some additional understanding and tools that can be used in other situations. The Twin Prime Conjecture already spawned a modern tool of combinatorial and analytic number theory – the sieve theory.

The sieve theory was established at the beginning of the 20th century as a simple method to count prime numbers in intervals. Today it is a powerful tool to approach problems related to the Twin Prime Conjecture, e.g., Goldbach's Conjecture. It was already used to prove countless partial results supporting many conjectures and apparently there is much more for sieves to do.

Why is that so, that the Twin Prime Conjecture resists any attempts to prove it? There is a fundamental difference between the question about the infinitude of prime numbers and the infinitude of twin prime pairs. The latter one involves not only multiplicative properties of numbers, but also additive properties. These two branches of number theory have numerous books dedicated to each of them separately. The history shows that the most difficult problems are those, which involve both domains.

It is a shame that the truth about such basic facts is hidden from us. Hopefully, one day we will understand the primes or, as Paul Erdős once said: “It will be another million years, at least, before we understand the primes.”

1.2 The goal and structure of the thesis

The main goal of this work is to present the current state of knowledge about the Twin Prime Conjecture and prime sieving algorithms – the only practical way to compute Brun's constant.

In the Chapter 2, we discuss and prove some important theorems on the prime numbers as they will be used throughout this work.

The Chapter 3 is dedicated to the description of the current knowledge about the twin prime numbers and related problems. We also show few important ways to characterize the twin prime pairs, and we simplify and generalize slightly some original proofs (Theorem 17).

In the next chapter, i.e., the Chapter 4, the exposition of the most important prime sieving algorithms is given. We prove the correctness of presented algorithms, compute their complexity and also deduce theoretical bounds on the complexity of any sieving algorithm. We also present an elementary method to prove two theorems used by Atkin's Sieve (Theorems 20 and 23).

In the Chapter 5 we introduce basic sieve methods, in particular, we consider the sieve of Brun. The goal of this chapter is to show the famous theorem of Brun about sum of reciprocals of twin primes.

In the penultimate chapter, the Chapter 6, we present a result of computation performed to obtain values of constants related to the Twin Prime Conjecture. In particular we show how fast convergent series for the twin prime constant can be obtained and we compute the twin prime constant to 15000 decimal digits. To our knowledge, this is the most precise computation so far.

The last chapter summarizes and concludes the whole thesis.

1.3 Notation

We mostly use the standard mathematical notation, but there are some exceptions. Throughout the thesis we use "log" symbol for the natural logarithm. We also use Landau's notation of "big *O*" and "small *o*". Moreover, to shorten the notation, if the sum is over p, then this implicitly means that the sum is over prime numbers. Therefore

$$\sum_{p \leq x} 1 = \sum_{\substack{p \in \mathscr{P} \\ p \leq x}} 1,$$

where $\mathscr{P}$ is the set of prime numbers. Additional notation may be introduced when needed.

1.4 References

The work presented in this thesis would not be possible without help of the literature.

The Chapter 2 is mostly backed up by [2] and other classical positions on the number theory. Some proofs are taken from the works of other authors, e.g., the proof of Theorem 1 by Erdős ([16]).

The Chapter 3 is based on publications concerning the Twin Prime Conjecture, in particular [23] and [41]. The part about related problems is a compilation of various sources ([12], [22], [41], among others), with figures computed in Sage ([39]). The current computational records were taken from [6] and [14]. The ways to characterize the twin primes come from many publications: [8], [29], [30] and [36].

The prime sieving algorithms in the Chapter 4 are presented in the literature on the computational aspects of the number theory, most notably in [11] and [41]. The algorithms themselves were presented also in [3], [33] and [34]. A big part of the discussion is backed up by additional sources ([13], [15], [37], [42]).

The Chapter 5 follows mostly [9], [24] and [40], but is built around [5].

The computation of the related constants in the Chapter 6 required the following software: [4], [20] and [28]. To obtain a feasible formula for the twin prime constant, [17] was used.

Chapter 2

Prime numbers

2.1 Basic theorems

It is a common knowledge that there are infinitely many prime numbers. We present a beautiful proof of this fact, a proof that actually shows that the series over all primes

$$\sum_p \frac{1}{p}$$

diverges. Hence there must be an infinitude of primes.

This proof was published in an article of P. Erdős from 1938 ([16]) and is considered to be a "proof from the book" ([1]).

Theorem 1 (Infinitude of primes). *The series ($\mathscr{P}$ - set of all prime numbers)*

$$\sum_{p\in\mathscr{P}} \frac{1}{p} \tag{2.1}$$

diverges.

Proof. Let p_i be the i-th prime number. Assume that (2.1) converges. Thus, there must be an index k, such that

$$\sum_{i\geq k+1} \frac{1}{p_i} < \frac{1}{2}. \tag{2.2}$$

Let N be an arbitrary natural number. Let N_1 be a number of positive integers $n \leq N$ divisible only by primes $p_1, p_2, \ldots, p_k$, and N_2 be a number of positive integers divisible by at least one p_i where $i > k$. Clearly, $N_1 + N_2 = N$.

Let's estimate N_2. There are exactly $\left\lfloor \frac{N}{p} \right\rfloor$ numbers $n \leq N$ divisible by p. This, together with (2.2), gives

$$N_2 \leq \sum_{i\geq k+1} \left\lfloor \frac{N}{p_i} \right\rfloor \leq N \sum_{i\geq k+1} \frac{1}{p_i} < \frac{N}{2},$$

as we count some numbers more than once.

Let's take a look at N_1. Every number $n \leq N$ can be written as a product $n = a_n b_n^2$ where a_n is squarefree. If n is only divisible by primes up to p_k, we may have at most 2^k different squarefree parts of n (we may either include or not include each prime). If it comes to the square part, we note that $b_n \leq \sqrt{n} \leq \sqrt{N}$. Therefore, there are at most $2^k\sqrt{N}$ such numbers, and we have $N_1 \leq 2^k\sqrt{N}$.

But if we take $N = 2^{2k+2}$, this leads to

$$N = N_1 + N_2 < \frac{N}{2} + 2^k\sqrt{N} = 2^{2k+1} + 2^{2k+1} = N,$$

a contradiction that finishes the proof. □

Knowing the divergence of this series is not enough. We want to know precisely what is the character of this divergence, i.e., how it behaves asymptotically.

2.2 Mertens' theorems

The goal of this section is to prove the following famous theorem:

Theorem 2 (Mertens' Second Theorem). *For $x \geq 2$ we have*

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{\log x}\right) \tag{2.3}$$

for some constant B.

Let's start with some basic definitions.

Definition (p-adic valuation). Let p be a prime and n a positive integer. Then

$$\nu_p(n) = k \tag{2.4}$$

if and only if $p^k \mid n$, but $p^{k+1} \nmid n$.

It is easy to see that $\nu_p(nm) = \nu_p(n) + \nu_p(m)$. Also if $\nu_p(n) = k$ then $n = p^k m$ where $p \nmid m$. In that case we will write $p^k \parallel n$.

To obtain a value of $\nu_p(n!)$ we will prove

Theorem 3 (Legendre's Theorem). *Let p be a prime and n be a positive number. Then*

$$\nu_p(n!) = \sum_{i \geq 1} \left\lfloor \frac{n}{p^i} \right\rfloor. \tag{2.5}$$

Proof.

$$\begin{aligned}
\nu_p(n!) &= \sum_{1\le i\le n} \nu_p(i) = \sum_{1\le i\le n} \sum_{\substack{j\ge 1\\ \nu_p(i)=j}} j = \sum_{j\ge 1} \sum_{\substack{1\le i\le n\\ \nu_p(i)=j}} j = \\
&= \sum_{j\ge 1} j \sum_{\substack{1\le i\le n\\ \nu_p(i)=j}} 1 = \sum_{j\ge 1} j \sum_{\substack{1\le i\le n\\ p^j \| i}} 1 = \sum_{j\ge 1} j \left(\left\lfloor \frac{n}{p^j} \right\rfloor - \left\lfloor \frac{n}{p^{j+1}} \right\rfloor \right) = \\
&= \sum_{j\ge 1} j \left\lfloor \frac{n}{p^j} \right\rfloor - \sum_{j\ge 1} j \left\lfloor \frac{n}{p^{j+1}} \right\rfloor = \sum_{j\ge 1} j \left\lfloor \frac{n}{p^j} \right\rfloor - \sum_{j\ge 2} (j-1) \left\lfloor \frac{n}{p^j} \right\rfloor = \\
&= \left\lfloor \frac{n}{p} \right\rfloor + \sum_{j\ge 2} \left\lfloor \frac{n}{p^j} \right\rfloor = \sum_{j\ge 1} \left\lfloor \frac{n}{p^j} \right\rfloor .
\end{aligned}$$

□

It will be convenient to also define the following function:

Definition (The first Chebyshev function). For $x \ge 1$, we define the first Chebyshev function as

$$\vartheta(x) = \sum_{p\le x} \log p \, . \tag{2.6}$$

P. L. Chebyshev proved in 1850 the famous Bertrand's postulate.

Theorem 4 (Bertrand's postulate, Chebyshev's Theorem). *For every $n > 1$ there exists a prime number p, such that*

$$n < p < 2n.$$

With sufficient bounds on the function ϑ, one can get a simple and elementary proof of Bertrand's postulate ([1, pages 7–10]). We will prove the following fact needed in our discourse.

Theorem 5 (Bounds on the first Chebyshev function). *There exist positive constants c_1 and c_2, such that for $x \ge 4$*

$$c_1 \le \frac{\vartheta(x)}{x} \le c_2. \tag{2.7}$$

In particular

$$\vartheta(x) = O(x).$$

Remark. Prime Number Theorem (Theorem 13) is equivalent to the fact that

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1.$$

Proof of Theorem 5. Let's start with a clever observation about $\binom{2n}{n}$. Namely, for every prime p, such that $n < p \leq 2n$, $p \mid \binom{2n}{n} = \frac{(2n)!}{(n!)^2}$ since it is present in the numerator, but absent from the denominator. Hence

$$\prod_{n<p\leq 2n} p \leq \binom{2n}{n} \leq (1+1)^{2n} = 4^n$$

or, in terms of the function ϑ,

$$\vartheta(2n) - \vartheta(n) = \log \prod_{n<p\leq 2n} p \leq n \log 4.$$

Let's set an integer k, such that $2^{k-1} < x \leq 2^k$. This leads to

$$\begin{aligned}\vartheta(x) \leq \vartheta\left(2^k\right) &= \sum_{0\leq j\leq k-1} \left(\vartheta\left(2^{j+1}\right) - \vartheta\left(2^j\right)\right) \leq \\ &\leq \log 4 \sum_{0\leq j\leq k-1} 2^j \leq 2^k \log 4 \leq \left(2\log 4\right) x. \end{aligned} \tag{2.8}$$

Therefore, in (2.7) we can take $c_2 = 2\log 4$.

To prove the lower bound, we observe that for $n \geq 1$ we have

$$2^n \leq \binom{2n}{n}, \tag{2.9}$$

a fact that can be verified by a simple induction. Moreover, from Theorem 3 we deduce that

$$\log\binom{2n}{n} = \sum_{p\leq 2n} \sum_{j\geq 1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2\left\lfloor \frac{n}{p^j} \right\rfloor\right) \log p. \tag{2.10}$$

Hence, (2.9) and (2.10) together give

$$n \log 2 \leq \sum_{p\leq 2n} \sum_{j\geq 1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2\left\lfloor \frac{n}{p^j} \right\rfloor\right) \log p.$$

At this point, we introduce an integer $\alpha \geq 1$ and split the last sum to obtain

$$\begin{aligned}&\sum_{p\leq 2n} \sum_{1\leq j\leq \alpha} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2\left\lfloor \frac{n}{p^j} \right\rfloor\right) \log p \geq \\ &\qquad \geq n \log 2 - \sum_{p\leq 2n} \sum_{j\geq \alpha+1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2\left\lfloor \frac{n}{p^j} \right\rfloor\right) \log p. \end{aligned} \tag{2.11}$$

The observation that $(\lfloor 2x \rfloor - 2\lfloor x \rfloor)$ is either 0 or 1 permits us to write

$$\sum_{p\leq 2n} \sum_{1\leq j\leq \alpha} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2\left\lfloor \frac{n}{p^j} \right\rfloor\right) \log p \leq \alpha \sum_{p\leq 2n} \log p = \alpha \cdot \vartheta(2n)$$

and (2.11) becomes

$$\alpha \cdot \vartheta(2n) \geq n \log 2 - \sum_{p \leq 2n} \sum_{j \geq \alpha+1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor \right) \log p. \tag{2.12}$$

Let's bound the sum on the right-hand side of the equation above. We have

$$\begin{aligned} \sum_{p \leq 2n} \sum_{j \geq \alpha+1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor \right) \log p \leq \\ &\leq 2n \sum_{p \leq 2n} \sum_{j \geq \alpha+1} \frac{\log p}{p^j} \leq 2n \sum_{p \leq 2n} \sum_{j \geq \alpha+1} \frac{p}{p^j} = \\ &= 2n \sum_{p \leq 2n} \frac{1}{p^\alpha - p^{\alpha-1}} \leq 4n \sum_{p \leq 2n} \frac{1}{p^\alpha}. \end{aligned}$$

Consequently

$$4n \sum_{p \leq 2n} \frac{1}{p^\alpha} \leq 4n \sum_{k \geq 2} \frac{1}{k^\alpha} \leq 4n \int_1^\infty \frac{1}{t^\alpha}\, dt = \frac{4n}{\alpha - 1}.$$

Plugging this back into (2.12) gives

$$\alpha \cdot \vartheta(2n) \geq n \log 2 - \frac{4n}{\alpha - 1} = n \left(\log 2 - \frac{4}{\alpha - 1} \right).$$

Finally, if we take $\alpha = 7$, then

$$\vartheta(2n) \geq \frac{2n}{1000}. \tag{2.13}$$

To finish the proof, take any positive real $x \geq 4$. There exists an even integer $2m$, such that $x \geq 2m$, but $x - 2m < 2$. Using (2.13) we obtain

$$\vartheta(x) \geq \vartheta(2m) \geq \frac{2m}{1000} > \frac{x-2}{1000}.$$

However, since $x \geq 4$ or $x - 2 \geq \frac{x}{2}$, we get

$$\vartheta(x) > \frac{x}{2000}. \tag{2.14}$$

Hence, if we take $c_1 = \frac{1}{2000}$, the theorem is proven. □

Before going further, let's introduce a basic tool used to work with sums involving prime numbers. We follow [2, page 77].

Theorem 6 (Abel's summation formula). *Let $a(n)$ be a function from the set of integers to the set of complex numbers. Let*

$$A(x) = \sum_{n \leq x} a(n),$$

where $A(x) = 0$ if $x < 1$. If f has continuous derivative on the interval $[y, x]$, where $0 < y < x$, then we have

$$\sum_{y<n\leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)\,dt. \tag{2.15}$$

Proof. The proof is easy if one uses Riemann-Stieltjes integration. We see that $A(x)$ is a step function with jumps $a(n)$ at every integer value in the sum (2.15). Thus we can write

$$\sum_{y<n\leq x} a(n)f(n) = \int_y^x f(t)\,dA(t)$$

and integrate by parts to obtain

$$\begin{aligned}\sum_{y<n\leq x} a(n)f(n) &= \left[f(t)A(t)\right]_y^x - \int_y^x A(t)\,df(t) = \\ &= f(x)A(x) - f(y)A(y) - \int_y^x A(t)f'(t)\,dt. \qquad \square\end{aligned}$$

Using Abel's summation formula it is easy to obtain

Theorem 7 (Euler's summation formula). *If f has continuous derivative on the interval $[y, x]$, for $0 < y < x$, then*

$$\sum_{y<n\leq x} f(n) = \int_y^x f(t)\,dt + \int_y^x \{t\}f'(t)\,dt + \{y\}f(y) - \{x\}f(x), \tag{2.16}$$

where $\{x\}$ stands for the fractional part of x, i.e., $\{x\} = x - \lfloor x \rfloor$.

Proof. Take $a(n) = 1$ (hence $A(x) = \lfloor x \rfloor$). Using the formula for the integration by parts, i.e.,

$$\int_y^x f'(t)\,dt = xf(x) - yf(y) - \int_y^x f(t)\,dt,$$

and by a proper rearrangement of terms, the result is obtained. □

We will use this formula to obtain a well known asymptotics for the harmonic series. We have

Theorem 8 (Asymptotic formula for the harmonic series).

$$\sum_{1\leq n\leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right), \tag{2.17}$$

where

$$\gamma \approx 0.5772156649015329$$

is a constant, known as Euler-Mascheroni constant, Euler's constant or simply – gamma constant.

Proof. We use Euler's summation formula with $f(t) = \frac{1}{t}$ and $y = 1$, to get

$$\begin{aligned}\sum_{1\le n\le x} \frac{1}{n} = \int_1^x \frac{1}{t} dt - \int_1^x \frac{\{t\}}{t^2} dt + 1 = \\ = \log x + \left(1 - \int_1^\infty \frac{\{t\}}{t^2} dt\right) + \int_x^\infty \frac{\{t\}}{t^2} dt = \\ = \log x + \left(1 - \int_1^\infty \frac{\{t\}}{t^2} dt\right) + O\left(\int_x^\infty \frac{1}{t^2} dt\right) = \\ = \log x + \left(1 - \int_1^\infty \frac{\{t\}}{t^2} dt\right) + O\left(\frac{1}{x}\right).\end{aligned}$$

If we set

$$\gamma = 1 - \int_1^\infty \frac{\{t\}}{t^2},$$

the result follows. □

Using the previous theorem, we can redefine γ as:

$$\gamma = \lim_{x\to\infty} \left(\log x - \sum_{1\le n\le x} \frac{1}{n}\right)$$

One of the greatest unsolved problems in mathematics asks if this constant is rational or not. Empirical data suggests that the latter is true: if $\gamma = \frac{a}{b}$, then it is known that b must have at least 242080 digits ([26, page 97]).

We will also need a weak version of Stirling's approximation formula.

Theorem 9 (Stirling's approximation formula).

$$\log(n!) = n \log n + O(n). \tag{2.18}$$

Proof. Using Euler's summation formula with $f(x) = \log x$ and $y = 1$, we will get

$$\begin{aligned}\sum_{1\le k\le n} \log k = \sum_{1<k\le n} \log k = \int_1^n \log t + \int_1^n \frac{\{t\}}{t} dt = \\ = \left[t(\log t - 1)\right]_1^n + O\left(\int_1^n \frac{1}{t} dt\right) = \\ = n\log n - n + O(\log n) = n \log n + O(n).\end{aligned}$$ □

We are ready to prove the following

Theorem 10 (Mertens' First Theorem). *For $x \ge 2$ we have*

$$\sum_{p\le x} \frac{\log p}{p} = \log x + O(1). \tag{2.19}$$

Proof. We write

$$n! = \sum_{p\le n} p^{\nu_p(n!)}$$

or, equivalently

$$\log(n!) = \sum_{p\le n} \nu_p\,(n!) \log p.$$

Now, from Theorem 3, we get

$$\begin{aligned}\log(n!) &= \sum_{p\le n}\sum_{k\ge 1}\left\lfloor \frac{n}{p^k}\right\rfloor \log p = \\ &= \sum_{p\le n}\left\lfloor \frac{n}{p}\right\rfloor \log p + O\left(n\sum_{p\le n}\sum_{k\ge 2}\frac{\log p}{p^k}\right) = \\ &= \sum_{p\le n}\left\lfloor \frac{n}{p}\right\rfloor \log p + O\left(n\sum_{p\le n}\frac{\log p}{p^2}\right),\end{aligned}$$

since $\sum_{k\ge 2}\frac{\log p}{p^k} = \frac{\log p}{p^2-p} \le \frac{2\log p}{p^2}$.

Moreover, $\sum_{k\ge 1}\frac{\log k}{k^2}$ converges (e.g., by the Cauchy test of convergence) and we get

$$\log(n!) = \sum_{p\le n}\left\lfloor \frac{n}{p}\right\rfloor \log p + O\,(n) = \sum_{p\le n}\frac{n}{p}\log p + O\left(\sum_{p\le n}\log p\right) + O\,(n)\,.$$

Application of Theorem 5 gives us

$$\log(n!) = \sum_{p\le n}\frac{n}{p}\log p + O\,(n)\,.$$

However, from Theorem 9 we already know that

$$\log(n!) = n\log n + O(n)$$

and we obtain

$$\sum_{p\le n}\frac{n}{p}\log p = n\log n + O(n).$$

If we divide it by n, we will finally obtain

$$\sum_{p\le n}\frac{\log p}{p} = \log n + O(1).$$

To finish the proof we have to consider any real $x \ge 2$. But then

$$\sum_{p\le x}\frac{\log p}{p} = \sum_{p\le \lfloor x\rfloor}\frac{\log p}{p} = \log\lfloor x\rfloor + O(1) = \log x + O(1),$$

since from the l'Hôpital's rule the difference $(\log\lfloor x\rfloor - \log x)$ converges to zero as x approaches infinity. □

After preparing all the necessary tools we are now ready to prove Theorem 2.

Proof of Theorem 2. We will use Abel's summation formula with $f(x) = \frac{1}{\log x}$ and

$$a(n) = \begin{cases} \frac{\log n}{n}, & \text{if n is a prime,} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$A(x) = \sum_{p \le x} \frac{\log p}{p}$$

and

$$\sum_{n \le x} a(n) f(n) = \sum_{p \le x} \frac{1}{p}.$$

So we have

$$\sum_{p \le x} \frac{1}{p} = \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2}\, dt.$$

But from Theorem 10 we know that $A(x) = \log x + O(1)$. We write $A(x) = \log x + R(x)$. Therefore

$$\begin{aligned}
\sum_{p \le x} \frac{1}{p} &= \frac{\log x + R(x)}{\log x} + \int_2^x \frac{\log t + R(t)}{t(\log t)^2}\, dt = \\
&= 1 + \frac{R(x)}{\log x} + \int_2^x \frac{1}{t \log t}\, dt + \int_2^x \frac{R(t)}{t(\log t)^2}\, dt = \\
&= 1 + \frac{R(x)}{\log x} + \log\log x - \log\log 2 + \int_2^\infty \frac{R(t)}{t(\log t)^2}\, dt - \int_x^\infty \frac{R(t)}{t(\log t)^2}\, dt = \\
&= \log\log x + B + \frac{R(x)}{\log x} - \int_x^\infty \frac{R(t)}{t(\log t)^2}\, dt = \\
&= \log\log x + B + O\left(\frac{1}{\log x} - \int_x^\infty \frac{1}{t(\log t)^2}\right) = \log\log x + B + O\left(\frac{1}{\log x}\right),
\end{aligned}$$

where

$$B = 1 - \log\log 2 + \int_2^\infty \frac{R(t)}{t(\log t)^2}\, dt. \qquad \square$$

The constant B is called the *Mertens constant* (or *Meissel–Mertens constant*) and has the value

0.2614972128476427837554268386086958590515666482612.

A formula for B with a very good numerical convergence is given by ([17])

$$B = \gamma + \sum_{k \ge 2} \frac{\mu(k)}{k} \log \zeta(k),$$

where γ is the Euler-Mascheroni constant, μ is the Möbius function and ζ is the Riemann Zeta function.

Mertens also proved the following, beautiful

Theorem 11 (Mertens' Third Theorem).

$$\prod_{p\le x}\left(1-\frac{1}{p}\right)\sim\frac{e^{-\gamma}}{\log x},$$

where e is Napier's constant and γ is the Euler-Mascheroni constant.

We will not prove this theorem here, because we will only need a much weaker version. Instead, we will prove

Theorem 12 (Weak Mertens' Third Theorem). *There exist positive constants x_0, c_1 and c_2, such that for $x > x_0$*

$$\frac{c_1}{\log x}\le\prod_{p\le x}\left(1-\frac{1}{p}\right)\le\frac{c_2}{\log x}. \tag{2.20}$$

In particular

$$\prod_{p\le x}\left(1-\frac{1}{p}\right)=O\left(\frac{1}{\log x}\right).$$

Proof. On the one hand, we have

$$\begin{aligned}\prod_{p\le x}\frac{1}{1-\frac{1}{p}}&=\prod_{p\le x}\left(1+\frac{1}{p}+\frac{1}{p^2}+\ldots\right)=\\&=\sum_{1\le k\le x}\frac{1}{k}+\sum_n{}'\frac{1}{n}\ge\sum_{1\le k\le x}\frac{1}{k}=\log x+O(1),\end{aligned} \tag{2.21}$$

where the sum $\sum'$ is over positive integers divisible only by the primes not bigger than x. Hence, if we divide (2.21) by $\log x$, we will get

$$\frac{1}{\log x}\prod_{p\le x}\frac{1}{1-\frac{1}{p}}\ge 1+O\left(\frac{1}{\log x}\right)\ge\frac{1}{2}$$

for x large enough. Thus (2.20) is bounded from above, if we take $c_2 = 2$.

On the other hand we have

$$\prod_{p\le x}\frac{1}{1-\frac{1}{p}}=\prod_{p\le x}\left(1+\frac{1}{p-1}\right)\le\prod_{p\le x}e^{\frac{1}{p-1}}=e^{\sum_{p\le x}\frac{1}{p-1}}. \tag{2.22}$$

But since

$$\sum_{p\le x}\frac{1}{p-1}-\sum_{p\le x}\frac{1}{p}=\sum_{p\le x}\frac{1}{p(p-1)}=O(1),$$

we get from Theorem 2 that

$$\sum_{p\le x}\frac{1}{p-1}=\log\log x+O(1).$$

Plugging this into (2.22) gives

$$\prod_{p\le x}\frac{1}{1-\frac{1}{p}} \le e^{\log\log x+O(1)} = O(\log x)$$

and so

$$\frac{1}{\log x}\prod_{p\le x}\frac{1}{1-\frac{1}{p}} \le O(1).$$

proving that (2.20) is bounded from below (for x large enough). This concludes the proof. □

2.3 Prime Number Theorem

Finally, we have the celebrated Prime Number Theorem, first proved in 1896 by J. Hadamard and C. J. de la Vallée-Poussin, independently. The "elementary" proof is attributed to both A. Selberg and P. Erdős, who proved it by similar methods in 1949.

Theorem 13 (Prime Number Theorem). *If we let*

$$\pi(x) = \sum_{p\le x} 1,$$

then we have

$$\pi(x) \sim \frac{x}{\log x}.$$

We will not prove this essential fact. The reader may be interested in the ingenious and intricate proof given by Donald J. Newman in [31].

2.4 Summary

In this chapter, we formulated and proved few important theorems on the prime numbers. They will be needed later to prove important facts about twin primes and prime sieving algorithms. In particular we learned that the series

$$\sum_p \frac{1}{p}$$

is divergent, just as the harmonic series over natural numbers. This will contrast with the fact that the similar series for the twin primes is convergent.

Chapter 3

Twin primes

3.1 Twin Prime Conjecture

Introduction

In 1912, E. Landau presented four problems in the number theory that he considered to be very difficult:

1. Goldbach's Conjecture: Every positive even integer is a sum of two primes.
2. Twin Prime Conjecture: 2 can be written as a difference of two primes in infinitely many ways.
3. Legendre's Conjecture: There is always a prime between n^2 and $(n+1)^2$.
4. There are infinitely many primes of the form n^2+1.

All these problems remain open, proving that Landau was right. The first three of them are related, they concern primes in some intervals. We will concentrate on the second problem, i.e.,

Conjecture 1 (Twin Prime Conjecture). *There are infinitely many twin primes, i.e., numbers p and $p+2$, such that both of them are primes.*

The first pairs of twin primes are:

$$(3,5), \quad (5,7), \quad (11,13), \quad (17,19), \quad \ldots$$

with 5 being the only prime being in two pairs. Let's denote by $\pi_2(x)$ a number of primes p, not bigger than x, such that $p+2$ is also a prime. We therefore have:

$$\begin{gathered}\pi_2(10)=2,\\ \pi_2(11)=3,\\ \pi_2(17)=4,\\ \cdots\cdots\cdots\end{gathered}$$

x	$\pi_2(x)$	$2C_2x/\log^2 x$	$\mathrm{Li}_2(x)$
10^1	2	2	5
10^2	8	6	14
10^3	35	28	46
10^4	205	156	214
10^5	1224	996	1249
10^6	8169	6917	8248
10^7	58980	50822	58754
10^8	440312	389107	440368
10^9	3424506	3074426	3425308
10^{10}	27412679	24902848	27411417
10^{11}	224376048	205808662	224368865
10^{12}	1870585220	1729364450	1870559867
10^{13}	15834664872	14735413064	15834598305

Table 3.1: The values of $\pi_2(x)$ compared.

The Table 3.1 contains values of π_2 for some powers of 10.

Whereas there is no proof that there are infinitely many twin primes, the empirical data on the function π_2 strongly suggest that this is indeed true. Hardy and Littlewood (1924) conjectured

Conjecture 2 (Strong Twin Prime Conjecture). *Let*

$$\mathrm{Li}_2(x) = 2C_2 \int_2^x \frac{dt}{\log^2 t},$$

where

$$C_2 = \prod_{p\geq 3} \left(1 - \frac{1}{(p-1)^2}\right) \approx 1.320323632$$

is the so called twin prime constant. Then we have

$$\lim_{x\to\infty} \frac{\pi_2(x)}{\mathrm{Li}_2(x)} = 1. \tag{3.1}$$

This fact would imply the infinitude of twin primes and also a simple asymptotic formula for $\pi_2(x)$, namely

$$\pi_2(x) \sim 2C_2 \frac{x}{\log^2 x},$$

which shows particular resemblance to the Prime Number Theorem (Theorem 13). Table 3.1 and Figure 3.1 show that this approximation is much worse than the representation by an integral. This resonates with the known fact that

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}$$

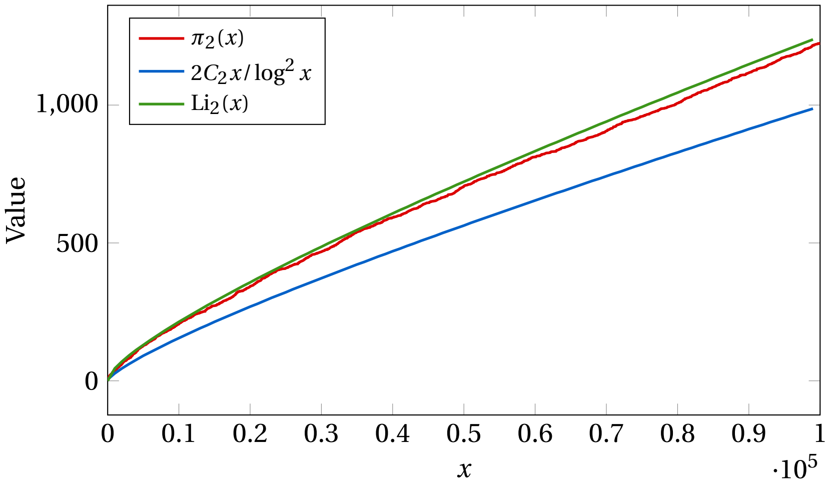

Figure 3.1: Plot of $\pi_2(x)$, $2C_2x/\log^2 x$ and $\mathrm{Li}_2(x)$ (cf. Table 3.1).

is a much better approximation to $\pi(x)$ than

$$\frac{x}{\log x},$$

although both are asymptotically equal to $\pi(x)$.

Where does C_2 come from? In [23] an intuitive, heuristic derivation is proposed and we present it below.

From the Prime Number Theorem (Theorem 13) we know that the probability of $x \geq 3$ being prime is roughly $1/\log x$. The probability that x and $x+2$ are primes at the same time is therefore $1/\log^2 x$, assuming that both events are independent. But they are not completely independent – if x is a prime, then x+2 is necessarily odd, for example. This doubles the probability of $x+2$ being a prime.

On the other hand, for any odd prime p, x will be not divisible by p with probability 1, i.e., it will belong to one of $p-1$ residue classes $\left(\mathrm{mod}\ p\right)$ with probability $\frac{p}{p-1}$. As x and $x+2$ cannot be in the same residue class, this changes the probability of $x+2$ being a prime from $\frac{p-1}{p}$ to $\frac{p-2}{p-1}$. Combining these two facts we obtain the probability of x and $x+2$ being a twin prime pair:

$$\frac{\pi_2(x)}{x} \approx \frac{1}{\log^2 x} \cdot 2 \prod_{p \geq 3} \left(\frac{p-2}{p-1}\right) \Big/ \left(\frac{p-1}{p}\right) = \frac{1}{\log^2 x} \cdot 2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2}\right) = \frac{2C_2}{\log^2 x}.$$

Even though we don't know whether Twin Prime Conjecture is true, we know some partial results.

In 1920, Viggo Brun showed that there exists a number x_0 (effectively computable), such that if $x > x_0$, then

$$\pi_2(x) < \frac{100x}{\log^2 x}.$$

This has been improved in 1966 by Bombieri and Davenport, who proved that

$$\pi_2(x) \le 8C_2 \frac{x}{\log^2 x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right).$$

The factor 8 above has been improved subsequently to 6.8325, but the Strong Twin Prime Conjecture requires it to be exactly 2.

Probably the greatest near-miss in proving the Twin Prime Conjecture is the theorem of Chen (Theorem 31).

There are interesting records related to twin primes. At the time of writing, the largest known twin prime pair is ([6])

$$65516468355 \cdot 2^{333333} \pm 1.$$

Every number from this pair has 100355 decimal digits.
It is also known that ([14])

$$\pi_2\left(2608 \cdot 10^{15}\right) = 2012314811498844.$$

Related problems

Goldbach's Conjecture

Another famous unsolved problem in number theory is the Goldbach's conjecture. One of its possible statements is

Conjecture 3 (The Extended Goldbach's Conjecture). *Let $R(n)$ be the number of representations of an even positive integer n as a sum of two primes. Then*

$$R(n) \sim 2C_2 \prod_{\substack{p|n \\ p>2}} \left(\frac{p-1}{p-2}\right) \int_2^n \frac{dt}{\log^2 t} \sim 2C_2 \prod_{\substack{p|n \\ p>2}} \left(\frac{p-1}{p-2}\right) \frac{n}{\log^2 n}.$$

See Figure 3.2 for a plot of $R(n)$.

The surprising presence of twin prime constant C_2 in the conjectured formula for $R(n)$ suggests connection to the Twin Prime Conjecture. In fact, as discussed in Section 5.1 below, the methods used to attack one of these problems, usually yield results for the second as well.

de Polignac's Conjecture

The conjecture stated in 1849 by de Polignac is

Conjecture 4 (de Polignac's Conjecture). *Let n be a positive even integer. Then n can be represented as a difference of two consecutive prime numbers in infinitely many ways. Stated differently: the number of prime gaps of size n, i.e., pairs of consecutive prime numbers p_k and p_{k+1} with $p_{k+1} - p_k = n$, is infinite.*

Figure 3.3 presents plots of the number of prime gaps below some number x, for different values of n.

The special case with $n = 2$ is the Twin Prime Conjecture.

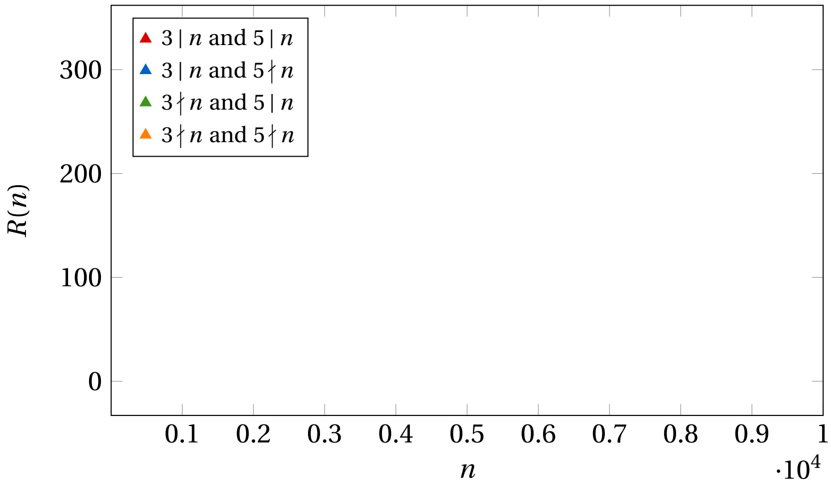

Figure 3.2: Number of representations of an even n as sum of two primes ("Goldbach's Comet"). There is an evident "clustering" of integers that share the same prime factors, here presented with primes 3 and 5.

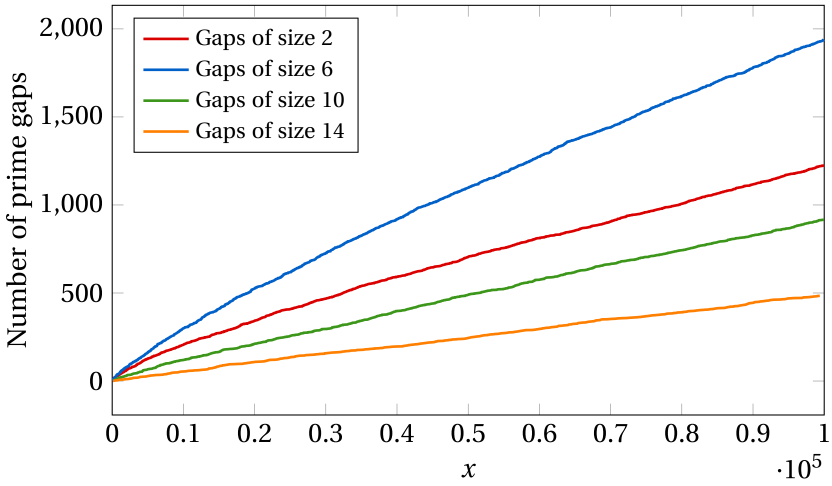

Figure 3.3: Number of prime gaps of size 2, 6, 10 and 14, such that the first prime is not greater than x (cf. Figure 3.4).

k-tuple conjecture

A very similar conjecture to de Polignac's Conjecture was proposed by Hardy and Littlewood. Consider any set of numbers $(a_1, a_2, \dots, a_k)$, called a *constellation*. We say that this constellation is *admissible* if it does not contain a complete set of residues modulo any prime number p. For example, $(0,2)$ or $(0,2,6,8,12)$ are admissible, whereas $(0,2,4)$ is not. Note also, that we have only to check primes $p \le k$.

A prime k-tuple for any admissible constellation consists of k numbers $(b_1 + a_1, b_1 + a_2, \dots, b_k + a_k)$, such that for every $i \in \{1,2,\dots,k\}$, $a_i + b_i$ is a prime. Now the reason for admissibility is obvious – otherwise it would be possible for a tuple to contain at least one number divisible by some prime p.

We may now formulate

Conjecture 5 (k-tuple conjecture). *There exist infinitely many prime k-tuples for any admissible constellation.*

It's clear now that one can assume that $a_1 = 0$.

The Twin Prime Conjecture is equivalent to the case of the constellation $(0,2)$. Furthermore, the prime pairs of the form $(p, p+4)$ are called *cousin primes* and primes of the form $(p, p+6)$ – *sexy primes*.

Hardy and Littlewood also conjectured asymptotic density for the number of primes p not greater than x, such that $p+n$ (for a fixed even n) is also a prime:

$$\pi_n(x) \sim 2C_n \int_2^n \frac{dt}{\log^2 t} \sim 2C_n \frac{x}{\log^2 x},$$

where

$$C_n = C_2 \prod_{\substack{p \mid n \\ p>2}} \frac{p-1}{p-2}$$

and C_2 is the twin prime constant. The formula suggests that the number of cousin primes is asymptotically the same as the number of twin primes ($C_4 = C_2$). This seems to be true and therefore $\pi_4(x)$ is not shown in Figure 3.4. On the other hand, $C_6 = 2C_2$, so the pairs of sexy primes should happen roughly two times more often than of size 2 or 4. This also seems to be true, as can be seen in Figure 3.4.

Small prime gaps

Let's define

$$\Delta = \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n}.$$

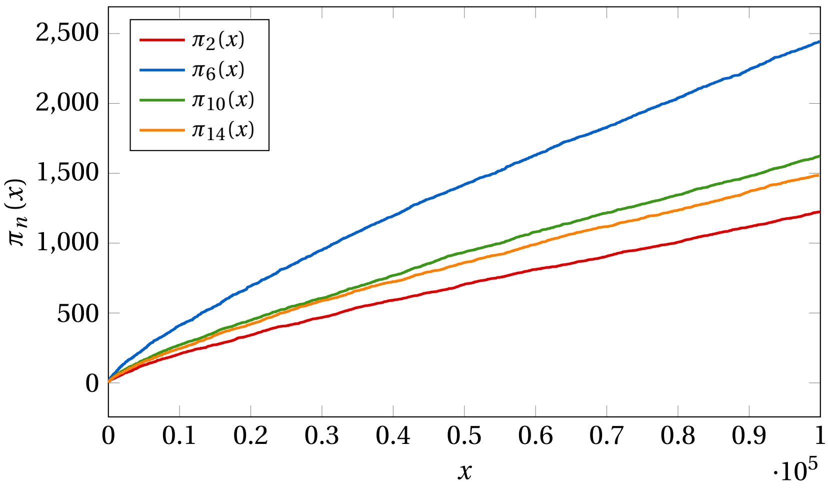

Figure 3.4: Values of $\pi_2(x)$, $\pi_6(x)$, $\pi_{10}(x)$ and $\pi_{14}(x)$ (cf. Figure 3.3).

Then the necessary condition for the Twin Prime Conjecture to be true is $\Delta = 0$. Indeed, if there are infinitely many twin prime pairs, say $(q_n, q_n + 2)$, then

$$\begin{aligned}\Delta &= \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} \\ &\leq \liminf_{n\to\infty} \frac{q_{n+1} - q_n}{\log q_n} = \liminf_{n\to\infty} \frac{2}{\log q_n} = \lim_{n\to\infty} \frac{2}{\log q_n} = 0.\end{aligned}$$

The value of Δ was subject to many improvements over the time. However, in [21] it has been established that in fact $\Delta = 0$, strongly suggesting that the Twin Prime Conjecture is true.

As a side note, note that also

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = \infty,$$

that is, *large* prime gaps also exist.

It is also conjectured (Cramér's Conjecture) that

$$p_{n+1} - p_n = O\big((\log p_n)^2\big).$$

and the numerical evidence suggests it may be true. This would also imply Legendre's Conjecture: for sufficiently large n there would always be a prime between n^2 and $(n+1)^2$.

Dickson's Conjecture

In [12], a generalization of Dirichlet's problem on the infinitude of primes in arithmetic progressions was presented. The problem is known as

Conjecture 6 (Dickson's Conjecture). *Let:*

$$a_1n+b_1, a_2n+b_2,\dots,a_kn+b_k$$

be a family of arithmetic progressions with $a_i>1$. Unless there is a prime p that divides all above values for all n, there exists infinitely many values of n, such that all above numbers are prime.

When we take $k=1$, then we obtain Dirichlet's Theorem. When the progressions are n and $n+2$, then we once again get Twin Prime Conjecture. Taking n and $2n+1$ gives conjecture about infinitude of Sophie-Germain primes. Similarly, de Polignac's Conjecture follows from Dickson's conjecture as well. Finally, the question about arbitrary long arithmetic progressions of primes is also a special case of this conjecture. As can be seen, some special cases have already been proven. The latter case is answered by the famous theorem of Green and Tao ([25]):

Theorem 14 (Green-Tao Theorem). *Primes contain arbitrarily long arithmetic progressions.*

3.2 Characterization of twin primes

Characterization by congruence relations

Let's start with a basic theorem that will be used throughout this section.

Theorem 15 (Wilson's Theorem). *Let $n>1$ be an integer. Then n is a prime if and only if*

$$(n-1)!+1\equiv 0 \pmod{n}. \tag{3.2}$$

More precisely,

$$(n-1)!+1\equiv\begin{cases}0 \pmod{n}, & \textit{if } n \textit{ is a prime,}\\ 1 \pmod{n}, & \textit{otherwise.}\end{cases} \tag{3.3}$$

Proof. Assume that n is a prime. Every number from the set $\{1,2,\dots,n-1\}$ is coprime to n and therefore has a multiplicative inverse modulo n. We know that the only solutions to the equation

$$x^2\equiv 1 \pmod{n}$$

are 1 and -1. These numbers are their own inverses and all other numbers can be grouped into pairs of mutual inverses, say (r_i,q_i) for $1\le i\le\frac{n-3}{2}$. As $r_iq_i\equiv 1 \pmod{n}$, we now have

$$\begin{aligned}(n-1)! &\equiv 1\cdot 2\cdot\ldots\cdot(n-1)\equiv 1\cdot(-1)\cdot(r_1q_1)\cdot(r_2q_2)\cdot\ldots\cdot(r_{\frac{n-3}{2}}q_{\frac{n-3}{2}})\equiv\\ &\equiv -1 \pmod{n}.\end{aligned}$$

This shows the first part of the theorem.

Assume now that n is a composite number. Hence there are two numbers $1 < a, b < n$, such that $n = ab$. If $a \neq b$ then these numbers are two distinct numbers in the factorial $(n-1)!$. Consequently $n = ab \mid (n-1)!$ and we are done.

Otherwise, $a = b$ and $n = a^2$. If $a = 2$ (and $n = 4$) the theorem can be verified by a direct calculation. Thus we may assume that $a > 2$. In that case $2a < a^2 = n$ and we see that a and $2a$ are two distinct numbers in the product $(n-1)!$. Finally we get $n = a^2 \mid (2a)a \mid (n-1)!$ and Wilson's theorem is proven. □

We will use this theorem to prove Clement's Theorem ([8]) which characterizes a pair of twin primes in the spirit of Wilson's Theorem.

Theorem 16 (Clement's Theorem on twin primes). *Let $n > 1$ be an integer. Integers n and $n+2$ are both primes if and only if*

$$4\,[(n-1)!+1] + n \equiv 0 \pmod{n(n+2)}\,. \tag{3.4}$$

Proof. The theorem can be verified by a direct computation for $n \leq 4$, so we assume that $n > 4$.

From now on, assume that n and $n+2$ are primes. From Wilson's theorem we have $(n-1)! + 1 \equiv 0 \pmod{n}$. Hence

$$4\,[(n-1)!+1] + n \equiv 4 \cdot 0 + n \equiv 0 \pmod{n}\,.$$

Since $n+2$ is a prime too, we get $(n+1)! + 1 \equiv 0 \pmod{(n+2)}$ and

$$\begin{aligned} 4\,[(n-1)!+1] + n &\equiv 2\,[2(n-1)!+2] + n \equiv 2\,[(-1)(-2)(n-1)!+2] + n \equiv \\ &\equiv 2\,[(n+1)!+1+1] + n \equiv 2 \cdot 1 + n \equiv 0 \pmod{(n+2)}\,. \end{aligned}$$

By Chinese Remainder Theorem we obtain (3.4).

Assume therefore that (3.4) is true. We have to show that n and $n+2$ are primes. If n is not a prime, then $(n-1)! \equiv 0 \pmod{n}$ and

$$0 \equiv 4\,[(n-1)!+1] + n \equiv 4 + n \equiv 4 \pmod{n}\,.$$

Hence $n \mid 4$. This implies that $n \leq 4$ - a contradiction, showing that n must be a prime.

This time assume that $n+2$ is not a prime. Then $(n+1)! \equiv 0 \pmod{(n+2)}$ and, just as before,

$$0 \equiv 4\,[(n-1)!+1] + n \equiv 2\,[(n+1)!+2] + n \equiv 4 + n \equiv 2 \pmod{(n+2)}\,.$$

Thus $(n+2) \mid 2$. In particular $n \leq 0$ which is absurd. This finishes the proof. □

In [30] a further generalization of Clement's theorem is given. It characterizes pairs of primes whose difference is an arbitrary positive even number (cf. Conjecture 5). In this thesis we present a slightly refined version of this theorem, requiring weaker assumptions and having a simpler proof.

Theorem 17 (Generalization of Clement's Theorem). *Let $n, k > 1$ be integers. Integers n and $n+2k$ are a pair of primes if and only if n has no proper prime divisors smaller than $2k$ and*

$$2k(2k)!\,[(n-1)!+1]+[(2k)!-1]\,n \equiv 0 \pmod{n(n+2k)}\,. \tag{3.5}$$

Remark. It can be easily checked that Clement's theorem is a special case of this theorem when $k = 1$.

Unfortunately, the assumption about proper prime factors of n cannot be omitted. One can see that by taking $n = 9$ and $k = 4$. These numbers satisfy the congruence (3.5), but 9 is not a prime.

Proof. Before we begin, we need a simple identity:

$$\begin{aligned}(n+2k-1)! &\equiv (n-1)!\cdot n\cdot(n+1)\cdot\ldots\cdot(n+2k-1) \equiv\\ &\equiv (n-1)!\cdot(-2k)\cdot(-2k+1)\cdot\ldots\cdot(-1) \equiv\\ &\equiv (n-1)!(2k)! \pmod{(n+2k)}\,.\end{aligned}$$

Assume that n and $n+2k$ are primes. It follows easily that n, being a prime, has no proper prime divisors and, in particular, no prime factors smaller than $2k$. From Wilson's theorem $(n-1)!+1 \equiv 0 \pmod{n}$ and

$$2k(2k)!\,[(n-1)!+1]+[(2k)!-1]\,n \equiv 0 \pmod{n}\,.$$

Also $n+2$ is a prime so $(2k)!(n-1)! \equiv (n+2k-1)! \equiv -1 \pmod{(n+2)}$. This leads to

$$\begin{aligned}2k(2k)!\,[(n-1)!+1]+[(2k)!-1]\,n &\equiv 2k(2k)!(n-1)!+2k(2k)!+n(2k)!-n \equiv\\ &\equiv -2k-2k(2k)!+n(2k)!-n \equiv\\ &\equiv [(2k)!-1]\,(n+2k) \equiv 0 \pmod{(n+2k)}\,.\end{aligned}$$

Application of Chinese Remainder Theorem shows that (3.5) holds.

Next, let's prove that n and $n+2k$ are primes when (3.5) is true and n has no proper prime factors smaller than $2k$. Assume that $2n+k$ is not a prime. Another application of Wilson's theorem gives $(2k)!(n-1)! \equiv (n+2k-1)! \equiv 0 \pmod{(n+2k)}$. We easily see that

$$\begin{aligned}0 &\equiv 2k(2k)!\,[(n-1)!+1]+[(2k)!-1]\,n \equiv\\ &\equiv 2k(2k)!(n-1)!+2k(2k)!+n(2k)!-n \equiv\\ &\equiv (2k)!(n+2k)-n \equiv 2k \pmod{(n+2k)}\,.\end{aligned}$$

In particular $n+2k \le 2k$ or $n \le 0$. This contradiction shows that $n+2k$ is a prime.

It remains to show that n is a prime. Assume it is not. We already know that $n+2k$ is a prime, so $2k$ and $n+2k$ are coprime. From this we deduce that $2k$ and n are also coprime:

$$(n,2k) = (n+2k,2k) = 1.$$

Since n is not a prime or $(n-1)! \equiv 0 \pmod{n}$, we obtain from (3.5)

$$\begin{aligned} 0 &\equiv 2k(2k)!\,[(n-1)!+1] + [(2k)!-1]\,n \equiv \\ &\equiv 2k(2k)! \pmod{n}. \end{aligned}$$

But $(n, 2k) = 1$ and we can divide by $(2k)^2$ to get

$$(2k-1)! \equiv 0 \pmod{n}.$$

Let p be a proper prime factor dividing n. From our assumptions we know that $p \geq 2k$. But from the above relation $p \mid n \mid (2k-1)!$, so p divides at least one number smaller than $2k$. This means that $p < 2k$ - a contradiction that shows that n must be a prime. The proof is concluded. □

Characterization by multiplicative functions

Another way to characterize twin prime pairs is the theorem given in [29]. Beforehand, however, we need to define functions σ and φ.

Definition (Divisor function σ). For a positive integer n, we define

$$\sigma(n) = \sum_{d|n} d,$$

a sum of all positive divisors of n.

Definition (Euler's totient function φ). For a positive integer n, we define

$$\sigma(n) = \sum_{\substack{1 \leq a \leq n \\ (a,n)=1}} 1,$$

a number of positive integers not bigger than n and coprime to n.

It is easy to check that these functions are *multiplicative*, i.e., whenever a and b are coprime,

$$\sigma(ab) = \sigma(a)\sigma(b) \quad \text{and} \quad \varphi(ab) = \varphi(a)\varphi(b).$$

Using this property, if we write an integer n as

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_k^{a_k},$$

where p_i are distinct primes and a_i are non-negative integers for $i \in \{1, 2, \ldots, k\}$, then

$$\sigma(n) = \prod_{1 \leq i \leq k} \frac{p^{a_i+1} - 1}{p-1} \tag{3.6}$$

and

$$\varphi(n) = \prod_{1 \leq i \leq k} p^{a_i - 1}\left(p-1\right). \tag{3.7}$$

Now we can state the following

Theorem 18 (Characterization of twin primes by multiplicative functions). *A number n is a product of two primes that differ by 2 (i.e., twin primes) if and only if*

$$\varphi(n)\sigma(n) = (n-1)^2 - 4. \tag{3.8}$$

Proof. Let's proof the necessity of the condition. Assume therefore that $n = p(p+2)$ and both p and $p+2$ are primes. Then, by multiplicative properties of φ and σ

$$\varphi(n)\sigma(n) = (p-1)(p+1)\cdot(p+1)(p+3) = (p(p+2)-1)^2 - 4.$$

Assume now that (3.8) holds. We may write n as

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_k^{a_k},$$

where $(p_1 > p_2 > \ldots > p_k)$. Then, by (3.6) and (3.7),

$$\varphi(n)\sigma(n) = \prod_{1\le i\le k} p_i^{a_i-1}\left(p_i^{a_i+1}-1\right)$$

and our assumption can be written as

$$2n+3 = n^2 - \prod_{1\le i\le k} p_i^{a_i-1}\left(p_i^{a_i+1}-1\right). \tag{3.9}$$

First, let's notice that k cannot be 1. To see that, let's assume that $k=1$. Then we have

$$\begin{aligned} 2p^a+3 &= p^{a-1} \\ p^{a-1}\left(2p-1\right) &= 3. \end{aligned}$$

This implies that $p=1$, a contradiction and $k \ge 2$.

Now it's easy to see that n must be odd. Indeed, if n is even, then the left hand side of (3.9) is odd. Also, since $k \ge 2$ there is another, odd prime, say p_j, dividing n. But then $p_j^{a_j+1} - 1$ is even and the right hand side of (3.9) is even. This contradiction shows that $p_1 \ge 3$.

Now, we will prove that n must be squarefree. To see that, let's assume that $p^2 \mid n$ for some prime p. Reducing (3.9) $\pmod p$ we immediately see that $p \mid 3$ and p must be 3. So we have $3^2 \mid n$. But since $k \ge 2$ there exists a prime p_j, distinct from p, such that $p_j \mid n$ and $p_j^2 \nmid n$ (we just have proven the last fact). Hence $p_j \equiv \pm 1 \pmod 3$ or $p_j^{a_j+1} - 1 \equiv 0 \pmod 3$, since $a_j = 1$. Therefore the right hand side of (3.9) is divisible by 9, but the left side is not. Thus n is squarefree and (3.9) becomes

$$2n+3 = n^2 - \prod_{1\le i\le k}\left(p_i^2-1\right). \tag{3.10}$$

Consequently, let's rule out the case $k \geq 3$. Since $p_1 \geq 3, p_2 \geq 5, p_3 \geq 7$, we have

$$\begin{aligned} n^2 - \prod_{1\leq i\leq k} \left(p_i^2 - 1\right) &= p_1^2 p_2^2 \dots p_k^2 - (p_1^2-1)(p_2^2-1)\dots(p_k^2-1) > \\ &> p_1^2 p_2^2 \dots p_k^2 - (p_1^2-1)p_2^2 \dots p_k^2 = p_2^2 \dots p_k^2 = \\ &= \frac{p_2 p_3 \dots p_k}{p_1} \cdot n > p_3 p_4 \dots p_k \cdot n \geq 7n > \\ &> 2n+3, \end{aligned}$$

a contradiction which shows that n is a product of two distinct primes. Let's write $n = pq$ and plug it into (3.10). This gives

$$\begin{aligned} 2pq + 3 &= p^2 q^2 - (p^2-1)(q^2-1) \\ p^2 - 2pq + q^2 - 4 &= 0 \\ (p-q)^2 - 2^2 &= 0 \\ (p-q+2)(p-q-2) &= 0. \end{aligned}$$

This means that either $p = q-2$ or $p = q+2$. In both cases (p,q) is a twin prime pair. This finishes the proof of the theorem. □

Actually, in [36] the authors show

Theorem 19 (Sergusov's Theorem). *A number n is a product of two primes that differ by 2 if and only if*

$$\sigma(n) = n+1+2\sqrt{n+1} \qquad \text{or} \qquad \varphi(n) = n+1-2\sqrt{n+1}. \tag{3.11}$$

Proof. Let $n = p(p+2)$ for some prime p, such that $p+2$ is also prime. Then we have

$$\begin{aligned} \sigma\left(p(p+2)\right) &= (p+1)(p+3) = p(p+2)+1+2(p+1) = \\ &= p(p+2)+1+2\sqrt{(p+1)^2} = p(p+2)+1+2\sqrt{p(p+2)+1} = \\ &= n+1+2\sqrt{n+1}, \\ \varphi\left(p(p+2)\right) &= (p-1)(p+1) = p(p+2)+1-2(p+1) = \\ &= n+1-2\sqrt{n+1}. \end{aligned}$$

Now we will use two basic facts:

$$\begin{aligned} \sigma(n) &\geq n+1, \\ \varphi(n) &\leq n-1. \end{aligned}$$

Here the equality holds if and only if n is a prime.

We may now put $n = m^2 - 1$, because otherwise the right hand sides in (3.11) would not be integers. Then the equation becomes:

$$\sigma((m-1)(m+1)) = m(m+2) \qquad \text{or} \qquad \varphi((m-1)(m+1)) = m(m-2).$$

Let's assume that $(m-1, m+1) = 1$. Then

$$m(m+2) = \sigma\left((m-1)(m+1)\right) \geq m(m+2).$$

Therefore both $m-1$ and $m+1$ must be primes.

If $m-1$ and $m+1$ are not coprime, then $(m-1, m+1) = (2, m+1) = 2$ and m is odd. We put $m = 2k+1$ to get

$$4k^2 + 8k + 3 = \sigma\left(2k(2k+2)\right) = \sigma\left(4k(k+1)\right).$$

Now if k is odd then

$$4k^2 + 8k + 3 = \sigma\left(4(k+1)\right)\sigma\left(k\right) \geq (4k+5)(k+1) = 4k^2 + 9k + 5,$$

a contradiction. Therefore k is even and

$$4k^2 + 8k + 3 = \sigma\left(k+1\right)\sigma\left(4k\right) \geq (k+2)(4k+1) = 4k^2 + 9k + 2.$$

So $k = 1$, but this means that $n = 3 \cdot 5$.

Again, let's assume that $(m-1, m+1) = 1$. Thus we obtain

$$m(m-2) = \varphi\left((m-1)(m+1)\right) \leq m(m-2)$$

and once more both $m-1$ and $m+1$ must be primes.

If $(m-1, m+1) = 2$, we put $m = 2k+1$ to obtain

$$4k^2 - 1 = \varphi\left(4k(k+1)\right).$$

If k is odd, then:

$$4k^2 - 1 = \varphi\left(4(k+1)\right)\varphi\left(k\right) \leq (4k+3)(k-1) = 4k^2 - k - 3,$$

with no solutions in positive k.

Finally, if k is even then

$$4k^2 - 1 = \varphi\left(k+1\right)\varphi\left(4k\right) \leq k(4k-1) = 4k^2 - k,$$

which implies that $k = 1$ and $n = 3 \cdot 5$. □

3.3 Summary

In this chapter, the current knowledge about the Twin Prime Conjecture was presented. Moreover, we showed few related problems in the number theory. The positive answer to some of them would imply the Twin Prime Conjecture. We also provided couple of ways to characterize twin primes with the proofs.

Chapter 4

Prime sieving algorithms

The term *sieve* in mathematics is ambiguous – there are at least two different, yet connected ideas bearing this name. Most people will think about a sieve of Eratosthenes – the famous ancient method to produce a list of all primes up to a specific number. We will present this algorithm and refined algorithms that can be used for the same purpose.

However, this is only a tip of the iceberg. In 20th century a truly beautiful and powerful sieve methods were introduced. It dates back to 18th century when Legendre, using a sieve of Eratosthenes, developed an idea now confusingly known as the sieve of Eratosthenes, too. We have to stress a distinction between the sieve of Eratosthenes, *the algorithm* and the sieve of Eratosthenes, *the combinatorial tool.*

4.1 Sieve of Eratosthenes

The most basic and historically the first method used to obtain a list of prime numbers up to some limit is the sieve of Eratosthenes. Eratosthenes was the director of the library of Alexandria, famous not only because of his sieve, but also for performing high precision measurement of the circumference of Earth. However, historically the first appearance of sieve of Eratosthenes is known to us from the work of Nicomedes, his only complete work that survived to our times, entitled *Introduction to arithmetic.*

Assume we want to make a list of prime numbers up to n. We start with a list of numbers from 2 to n and follow these steps:

1. Take the smallest number i that is not yet crossed out. This is a prime.

2. If $i^2 > n$ the algorithm finishes. Primes are the numbers that where not crossed out.

3. Otherwise cross out the numbers $2i$, $3i$, $4i$, etc. from the list.

Iteration 0:	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Iteration 1:	$\boxed{2}$	3	$\underline{4}$	5	$\underline{6}$	7	$\underline{8}$	9	$\underline{10}$	11	$\underline{12}$	13	$\underline{14}$	15	$\underline{16}$	17	$\underline{18}$	19	$\underline{20}$	21	$\underline{22}$	23	$\underline{24}$	25	$\underline{26}$	27	$\underline{28}$	29	$\underline{30}$
Iteration 2:	$\boxed{2}$	$\boxed{3}$	$\underline{4}$	5	$\underline{\underline{6}}$	7	$\underline{8}$	$\underline{9}$	$\underline{10}$	11	$\underline{\underline{12}}$	13	$\underline{14}$	$\underline{15}$	$\underline{16}$	17	$\underline{\underline{18}}$	19	$\underline{20}$	$\underline{21}$	$\underline{22}$	23	$\underline{\underline{24}}$	25	$\underline{26}$	$\underline{27}$	$\underline{28}$	29	$\underline{\underline{30}}$
Iteration 3:	$\boxed{2}$	$\boxed{3}$	$\underline{4}$	$\boxed{5}$	$\underline{\underline{6}}$	7	$\underline{8}$	$\underline{9}$	$\underline{\underline{10}}$	11	$\underline{\underline{12}}$	13	$\underline{14}$	$\underline{\underline{15}}$	$\underline{16}$	17	$\underline{\underline{18}}$	19	$\underline{\underline{20}}$	$\underline{21}$	$\underline{22}$	23	$\underline{\underline{24}}$	$\underline{25}$	$\underline{26}$	$\underline{27}$	$\underline{28}$	29	$\underline{\underline{\underline{30}}}$
Result:	$\boxed{2}$	$\boxed{3}$	$\underline{4}$	$\boxed{5}$	$\underline{\underline{6}}$	$\boxed{7}$	$\underline{8}$	$\underline{9}$	$\underline{\underline{10}}$	$\boxed{11}$	$\underline{\underline{12}}$	$\boxed{13}$	$\underline{14}$	$\underline{\underline{15}}$	$\underline{16}$	$\boxed{17}$	$\underline{\underline{18}}$	$\boxed{19}$	$\underline{\underline{20}}$	$\underline{21}$	$\underline{22}$	$\boxed{23}$	$\underline{\underline{24}}$	$\underline{25}$	$\underline{26}$	$\underline{27}$	$\underline{28}$	$\boxed{29}$	$\underline{\underline{\underline{30}}}$

Figure 4.1: Consecutive steps in the sieve of Eratosthenes

4. Go to the Step 1.

The example of this process is presented in Figure 4.1. The pseudocode for this algorithm is given in Algorithm 4.1.

The algorithm relies on a basic fact that if a number d is composite then it has a prime factor not greater than $\lfloor\sqrt{d}\rfloor$. This precisely is the stopping condition in Step 2.

We immediately see that the space needed to run this algorithm is $O(n)$.

Let's now compute the running time of this algorithm as the number of operations required to finish the sieving. In our model every arithmetical operation will take a bounded time what is not true in practice.

```
Require: $A[i]$ for $2 \le i \le n$ - an array of numbers to sieve for primes
Ensure: $A[i] = 1$ iff $i$ is a prime number
for $i \leftarrow 2 \ldots n$ do
    $A[i] \leftarrow 1$
end for
for $i \leftarrow 2 \ldots \lfloor\sqrt{n}\rfloor$ do
    if $A[i] = 1$ then
        for $j \in \{2i, 3i, 4i, \ldots\}$, $j \le n$ do
            $A[j] \leftarrow 0$
        end for
    end if
end for
```

Algorithm 4.1: Sieve of Eratosthenes.

The algorithm has $\pi(\sqrt{n})$ iterations. For every iteration we have to cross out all multiples of prime p (without p itself). There are $\left\lfloor\frac{n}{p}\right\rfloor - 1 = O\left(\frac{n}{p}\right)$ of them at each stage. Therefore the running time is of order

$$\sum_{p \le \sqrt{n}} O\left(\frac{n}{p}\right) = O\left(n \sum_{p \le \sqrt{n}} \frac{1}{p}\right) = O\left(n \log\log\sqrt{n}\right) = O\left(n \log\log n\right),$$

by using the Mertens' Theorem (Theorem 2). This is a little more than linear in respect to the size of n and we can say that on average $\log\log n$ operations need to be performed for every number to tell if it is a prime. This function diverges to infinity, but very slowly.

By using more sophisticated arguments one can reduce the complexity of the sieve of Eratosthenes. In [11, Section 3.2.7] authors present a refinement of the above algorithm which runs in sublinear time $O\big(n/\log\log n\big)$. We, instead, will describe in Section 4.5 how the factor $\log\log n$ can be easily removed leading to a sieving algorithm with linear complexity.

There is also a less known variant of the sieve of Eratosthenes, called Euler's sieve where each composite number is removed exactly once (Figure 4.1 shows that in the original algorithm numbers are often crossed out multiple times). The algorithm starts with a list of numbers from 2 to n and goes as follows:

1. Take the smallest number i that is not yet crossed out. This is a prime.
2. If $i^2 > n$ the algorithm finishes. Primes are the numbers that remain.
3. Otherwise build a new list by multiplying every element of original list by i. Remove every element from this list from the original list.
4. Go to the Step 1.

As we can see, only the Step 3 is substantially different. We *remove* numbers instead of crossing them out and they are not considered afterwards. Nevertheless, the number of iterations will be the same as before. The algorithm is presented as Algorithm 4.2.

```
Require: $A = \{2, 3, \ldots, n\}$ - a set of numbers to sieve for primes
Ensure: $A$ contains only prime numbers
while $\exists i \le \lfloor\sqrt{n}\rfloor, i \in A$ do
    $B \leftarrow \{ij : j \in A,\ j \ge i,\ ij \le n\}$
    $A \leftarrow A - B$
end while
```

Algorithm 4.2: Sieve of Euler.

The correctness of the algorithm can be proven easily. First, observe that primes will not be sieved during the process as only the composite numbers are removed. Now, let's take any composite number m from the list. Thus $m = pd$ where $1 < p, d < m$ and p is the smallest prime factor of m. It follows that m will be sieved during the iteration where $a = p$ because d (having no prime factors smaller than p) was not removed during previous iterations.

Let's compute the running time for that sieve. We assume that we can remove a number from a list in a constant time. There will be $n - \pi(n)$ removals in total, because we are left with $\pi(n)$ primes at the end. Therefore the running time is

$$n - \pi(n) = n - O\left(\frac{n}{\log n}\right) = O(n),$$

Iteration 0: 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Iteration 1: $\boxed{2}$ 3 5 7 9 11 13 15 17 19 21 23 25 27 29
Iteration 2: $\boxed{2}$ $\boxed{3}$ 5 7 11 13 17 19 23 25 29
Iteration 3: $\boxed{2}$ $\boxed{3}$ $\boxed{5}$ 7 11 13 17 19 23 29

Result: $\boxed{3}$ $\boxed{5}$ $\boxed{7}$ $\boxed{11}$ $\boxed{13}$ $\boxed{17}$ $\boxed{19}$ $\boxed{23}$ $\boxed{27}$ $\boxed{29}$

Figure 4.2: Consecutive steps in the sieve of Euler

by the Prime Number Theorem (Theorem 13). This is a small improvement over the original algorithm. In practice the assumption about constant time removal of elements is not feasible. The standard data structures implementing an ordered set of numbers (e.g., red-black trees) are of complexity $O(\log n)$ ([10, chapters 13, 14]).

As a result, implementations of Euler's sieve tend to be slower than the sieve of Eratosthenes.

4.2 Sieve of Sundaram

In [33] authors present an interesting prime sieve called Sundaram's sieve. It was discovered by an Indian student S. P. Sundaram in 1934.

The algorithm, as usual, starts with a list of numbers from 1 to n. This time, however, we will sieve numbers up to $2n+2$ as this sieve explicitly does not consider even numbers. Now, we cross out all numbers of the form $i + j + 2ij \leq n$ where $1 \leq i \leq j \leq \frac{n-1}{3}$. The primes are obtained by taking all the numbers that were not crossed out, multiplying them by 2 and incrementing by 1. Notice that 2 will not be on the list. The pseudocode for this algorithm is presented as Algorithm 4.3.

```
Require: $A[j]$ for $1 \leq j \leq n$ - an array of numbers to sieve for primes
Ensure: $A[j] = 1$ iff $2j+1$ is a prime number
for $j \leftarrow 1 \ldots n$ do
    $A[j] \leftarrow 1$
end for
for $j \leftarrow 1 \ldots \left\lfloor \frac{n-1}{3} \right\rfloor$ do
    for $i \leftarrow 1 \ldots \left\lfloor \frac{n-j}{1+2j} \right\rfloor$ do
        $A[i+j+2ij] \leftarrow 0$
    end for
end for
```

Algorithm 4.3: Sieve of Sundaram.

The algorithm works because of the following identity:

$$2(i+j+2ij)+1 = (2i+1)(2j+1). \tag{4.1}$$

Therefore an odd number will be crossed out if and only if it is a product of two odd numbers greater than 1, i.e., it is composite.

The memory requirement is of order $O(n)$ and the running complexity of this method is

$$\begin{aligned}\sum_{1\le j\le (n-1)/3} \left\lfloor \frac{n-j}{1+2j} \right\rfloor = \sum_{1\le j\le (n-1)/3} \frac{n-j}{1+2j} + O(n) = \\ = n \sum_{1\le j\le (n-1)/3} \frac{1}{1+2j} - \sum_{1\le j\le (n-1)/3} O(1) + O(n) = \\ = nO\big(\log n\big) + O(n) = O\big(n \log n\big),\end{aligned}$$

since $\frac{1}{2j} < \frac{1}{2j+1} < \frac{1}{2j+2}$ and by asymptotic behavior of harmonic series. This is slightly worse than the sieve of Eratosthenes, but nevertheless the algorithm is notable for its simplicity.

4.3 Sieve of Pritchard

In [34] the author shows a simple refinement of the sieve of Eratosthenes. The basic observation is that the order of two nested loops in Algorithm 4.1 can be reversed. If d is a composite number we can write $d = pf$, where p is the smallest prime dividing d (of course $p \le \sqrt{d}$ and $f > 1$). In the sieve of Eratosthenes we basically iterate over p and then over f. If we are to reverse this order, we have to have a list of primes beforehand. How can we do that if we actually sieve to obtain this list?

In fact we only need the primes up to $\sqrt{n}$. Then for every possible value of f we take every prime p and we cross out numbers of the form pf. We observe that f must be bigger than 1 and not bigger than $\frac{n}{2}$ (because $f = \frac{n}{p} \le \frac{n}{2}$). The last, but crucial observation is that if p is the smallest prime diving $d = pf$, then the smallest prime dividing f must be at least as big as p. Therefore p ranges from 2 to the smallest prime that divides the given f.

We therefore get a method presented as Algorithm 4.4. The list of primes up to $\sqrt{n}$ can be obtained using the classical sieve of Eratosthenes.

It's easy to prove that the algorithm is valid. We directly see that no prime numbers are crossed out. Hence we must convince ourselves only that every composite number is crossed out. But, as we analyzed above, every composite number d is of the form $d = pf$, where p is the smallest prime dividing d, so it will be crossed out.

In fact every composite number will be sieved out exactly once. To see that assume that a number $d = p_1 f_1$ is sieved also as $d = p_2 f_2$, where p_1, p_2 are distinct primes and p_1 is the smallest prime number diving d. We therefore have $p_1 < p_2$ and it follows that $p_1 \mid f_2$ since p_1 and p_2 are coprime. But this is impossible: when $f = f_2$ in the algorithm, the inner loop will finish as soon as $p = p_1 < p_2$. This shows that every composite number will be crossed out exactly once.

```
Require: $A[i]$ for $2 \le i \le n$ - an array of numbers to sieve for primes
Ensure: $A[i] = 1$ iff $i$ is a prime number
$P \leftarrow$ a set of primes from the set $\{1, 2, \ldots, \lfloor\sqrt{n}\rfloor\}$
for $i \leftarrow 2 \ldots n$ do
    $A[i] \leftarrow 1$
end for
for $f \leftarrow 2 \ldots \lfloor \frac{n}{2} \rfloor$ do
    for $p \in P$ (in ascending order) do
        $d \leftarrow pf$
        if $d > n$ then
            break
        end if
        $A[d] \leftarrow 0$
        if $f \bmod p = 0$ then
            break
        end if
    end for
end for
```

Algorithm 4.4: Sieve of Pritchard.

The memory complexity of this algorithm is again $O(n)$. The running time consists of the first sieving and the main loop, that is,

$$O\left(\sqrt{n}\log\log\sqrt{n}\right) + O(n - \pi(n)) = O(n)\,.$$

It means that the elementary observations made above gave rise to a sieving algorithm that is linear. Additional work can improve the time to $O\left(n/\log\log n\right)$. We discuss it in Section 4.5.

4.4 Sieve of Atkin

In [3], the authors propose a completely different approach. They use quadratic forms and the following three theorems to separate primes from composite numbers:

Theorem 20 (On the quadratic form $x^2 + 4y^2$). *Let n be a squarefree positive integer, such that $n \equiv 1 \pmod 4$. Then n is a prime if and only if the set*

$$\left\{(x, y) : x, y > 0,\ x^2 + 4y^2 = n\right\}$$

has an odd number of elements,

Theorem 21 (On the quadratic form $x^2 + 3y^2$). *Let n be a squarefree positive integer, such that $n \equiv 1 \pmod 6$. Then n is a prime if and only if the set*

$$\{(x, y) : x, y > 0,\ x^2 + 3y^2 = n\}$$

has an odd number of elements,

Theorem 22 (On the quadratic form $3y^2 - x^2$). *Let n be a squarefree positive integer, such that $n \equiv 11 \pmod{12}$. Then n is a prime if and only if the set*

$$\{(x, y) : y > x > 0,\ 3y^2 - x^2 = n\}$$

has an odd number of elements.

The authors prove these facts using properties of Euclidean domains obtained by extending $\mathbb{Z}$ by roots of unity. We will show Theorem 20 using an ingenious, "one-sentence" proof of Fermat's theorem on sum of squares given in [42] (see [1] for a "proof-from-The-Book" version). But before, we will need an obvious, but powerful lemma:

Lemma 1. *Let S be a finite set and involutions f, g on this set, i.e., functions from S to S, such that for every $x \in S$*

$$f(f(x)) = x \quad and \quad g(g(x)) = x.$$

Moreover, let f_1 and g_1 be numbers of fixed points of f and g, respectively. Then f_1 and g_1 have the same parity, that is,

$$f_1 \equiv g_1 \pmod 2.$$

Remark. The finite cardinality of S (assuming that f and g have a finite number of fixed points) is important as the following counterexample shows:

$$f : \mathbb{N}_+ \mapsto \mathbb{N}_+, \quad f(x) = \begin{cases} n-1, & \text{if } n \text{ is even,} \\ n+1, & \text{if } n \text{ is odd.} \end{cases}$$

$$g : \mathbb{N}_+ \mapsto \mathbb{N}_+, \quad g(x) = \begin{cases} 1, & \text{if } n = 1, \\ n+1, & \text{if } n > 1 \text{ and } n \text{ is even,} \\ n-1, & \text{if } n > 1 \text{ and } n \text{ is odd.} \end{cases}$$

Both f and g are involutions, but f has no fixed points and g has precisely one.

Proof of Lemma 1. First note that f and g are bijections from S to itself (they are their own inverses). Assume that $x \in S$ is not fixed by f. For such an element x we have: $f(x) = y \neq x$. But f is an involution, so this implies also that $f(y) = f(f(x)) = x \neq y$. Hence also y is not a fixed point. Consequently, the number of points that are not fixed by f is even. Therefore we have

$$|S| \equiv f_1 \pmod 2.$$

The same is true for g_1 and the lemma follows. □

Proof of Theorem 20. Let p be a number, such that $p \equiv 1 \pmod 4$. Consider the set

$$S = \left\{(x,y,z) \in \mathbb{N}^3 : x^2 + 4yz = p\right\}.$$

It is of finite cardinality and has an obvious involution given by the transformation $(x,y,z) \mapsto (x,z,y)$. Fixed points of this involution are precisely the representations of p as a sum of the considered quadratic form. Indeed, if $(x,y,z) = (x,z,y)$, then $y = z$ and $x^2 + 4y^2 = p$. Now we have another involution:

$$(x,y,z) \mapsto \begin{cases} (x+2z, z, y-x-z), & \text{if } x < y-z, \\ (2y-x, y, x-y+z), & \text{if } y-z < x < 2y, \\ (x-2y, x-y+z, y), & \text{if } x > 2y. \end{cases}$$

One can easily check that this function maps solutions of $x^2+4yz = p$ to different solutions. Also the boundaries above are not attained. Indeed, if $x = y-z$, then p would be a square - a contradiction. If $x = 2y$, then p would be even - again a contradiction.

Moreover, triples from the first case map onto the set from the third case and vice versa, and the triples from the second case map onto itself. Therefore the only fixed points of this involution satisfy the relation:

$$(2y-x, y, x-y+z) = (x,y,z).$$

Hence $x = y$ and we have

$$\begin{aligned} x^2 + 4xz &= p \\ x(x+4z) &= p. \end{aligned} \tag{4.2}$$

If p is a prime then the representation above is unique and necessarily $x = y = 1$, $z = \frac{p-1}{4}$ and there is exactly one fixed point. It follows that the first involution has an odd number of fixed points, i.e., there is an odd number of representations of p as a quadratic form $x^2 + 4y^2$, just as required by the assumptions of Theorem 20.

If, on the other hand, p is a composite number then there are $\frac{d(p)}{2}$ solutions to (4.2). But since p is a squarefree number divisible by more than one prime, $d(n)$ must be a power of 2 greater than 2. Then $\frac{d(p)}{2}$ is an even number too. This implies that in this case the number of fixed points is even and the theorem is proven. □

In [15] it has been shown that this method of proof can be generalized. In particular, Zagier's proof is a simple application of the presented technique. Moreover, in the article, the following theorem is also proven (actually only the necessity is proved, but the sufficiency follows as easily):

Theorem 23 (On the quadratic form $4x^2+3y^2$). *Let n be a squarefree positive integer, such that $n \equiv 7 \pmod{12}$. Then n is a prime if and only if the set*

$$\left\{(x,y) : x,y>0,\ 4x^2+3y^2=n\right\}$$

has an odd number of elements.

Theorem 23 is a weaker version of Theorem 21, but it will suffice. We give only an outline of a proof – the details are left to the reader as an exercise as they are very easy to check manually.

Proof of Theorem 23. Consider the set $S = \left\{(x,y,z)\in\mathbb{N}^3 : 3x^2+4yz=p\right\}$. We have a trivial involution $(x,y,z)\mapsto(x,z,y)$ whose fixed points map to representations of p as a quadratic form $3x^2+4y^2$. We also have the following, less trivial involution:

$$(x,y,z)\mapsto\begin{cases}(x-2y,3x-3y+z,y), & \text{if } y\in(0,\frac{x}{2}),\\ (-x+2y,y,3x-3y+z), & \text{if } y\in(\frac{x}{2},x+\frac{z}{3}),\\ (5x-4y+2z,6x-4y+3z,-3x+3y-z), & \text{if } y\in(x+\frac{z}{3},\frac{5}{4}x+\frac{z}{2}),\\ (-5x+4y-2z,-3x+3y-z,6x-4y+3z), & \text{if } y\in(\frac{5}{4}x+\frac{z}{2},\frac{3}{2}x+\frac{3}{4}z),\\ (7x-4y+4z,6x-3y+4z,-6x+4y+3z), & \text{if } y\in(\frac{3}{2}x+\frac{3}{4}z,\frac{7}{4}x+z),\\ (-7x+4y-4z,-6x+4y-3z,6x-3y+4z), & \text{if } y\in(\frac{7}{4}x+z,2x+\frac{4}{3}z),\\ (5x-2y+4z,3x-y+3z,-6x+3y-4z), & \text{if } y\in(2x+\frac{4}{3}z,\frac{5}{2}x+2z),\\ (-5x+2y-4z,-6x+3y-4z,3x-y+3z), & \text{if } y\in(\frac{5}{2}x+2z,3x+3z),\\ (x+2z,z,-3x+y-3z), & \text{if } y\in(3x+3z,\infty).\end{cases}$$

Note that this involution can be obtained in an almost algorithmic manner, as has been showed in [15].

Just as before, one can check, with easy but tedious calculations, that the boundaries are not attained and that it is actually an involution on the set S. Only 5 cases (precisely, the cases 2nd, 4th, 5th, 6th and 8th) above are actually involutions on their own, so we need to check only those for fixed points.

If $(-x+2y,y,3x-3y+z)=(x,y,z)$ then $x=y$ and

$$\begin{aligned} 3x^2+4xz &= p\\ x(3x+4z) &= p. \end{aligned} \tag{4.3}$$

If $p=ab$ $(a<b)$ is any factorization of p to distinct (n is squarefree) numbers then it must be that $a=x$ and $b=3x+4z$. It follows that $x=a$ and $z=\frac{b-3a}{4}$ (z is an integer if you consider it $\pmod 4$). Therefore we must have $b>3a$.

If $(-5x+4y-2z,-3x+3y-z,6x-4y+3z)=(x,y,z)$ then $3x-2y+z=0$ and $p=3x^2+4y(2y-3x)=3x^2+8y^2-12xy$. If we reduce $\pmod 3$ one gets that $p\equiv 2 \pmod 3$ – a contradiction.

If $(7x-4y+4z,6x-3y+4z,-6x+4y+3z)=(x,y,z)$ then $3x=2(y-z)$, so x is even. This would imply that p is even too – a contradiction.

If $(-7x+4y-4z, -6x+4y-3z, 6x-3y+4z) = (x, y, z)$ then $2x-y+z=0$ and $p = 3x^2+4z(2x+z) = (3x+2z)(x+2z)$. If, as before, $p = ab$ for any $a < b$, we must have $x+2z = a$ and $b = 3x+2z$. It follows that $x = \frac{b-a}{2}$ and $y = \frac{3a-b}{4}$, so $3a > b$ (again, a reduction $\pmod 4$ of $3a-b$ tells us that y is an integer). As we see, the cases counted here are exactly those that were not counted in (4.3). Therefore the total number of fixed points (so far) is a number of divisors of p divided by 2 (because we require $a < b$).

Finally, if $(-5x+2y-4z, -6x+3y-4z, 3x-y+3z) = (x, y, z)$ then $p = 3x^2+12xz+8z^2$. Reduction $\pmod 3$ implies that $p \equiv 2 \pmod 3$ – again a contradiction.

We see that there are exactly $\frac{d(p)}{2}$ fixed points. But from the proof of Theorem 20 we know that it is odd if and only if p is a prime. Therefore also the number of fixed points of $(x, y, z) \mapsto (x, z, y)$, i.e., representations of p as a quadratic form $3x^2+4y^2$, is odd if and only if p is a prime. The proof is finished. □

We will not prove Theorem 22 here, because it requires a slightly different approach. However, if the reader is familiar with the algebraic number theory, the proof in the article [3] is approachable.

Now we are ready to state the Algorithm 4.5.

The algorithm correctly identifies primes. We argue as follows. The first loop (x, y changing from 1 to $\lfloor\sqrt{n}\rfloor$) uses three theorems shown above. According to the theorems, all squarefree numbers of the form $12k+1$, $12k+5$, $12k+7$ and $12k+11$ will be marked as primes if and only if they are primes. The numbers from other residue classes are not considered, because (with the only two exceptions of 2 and 3) they must be composite. Hence at the end of the main loop all squarefree numbers will be correctly sieved.

In the second phase we deal with the remaining numbers. The numbers divisible by 2 or 3 need not be considered, because we already ruled them out by properly partitioning them according to their residue class $\pmod{12}$. Later, for every prime greater than 3, we sieve out the numbers that are divisible by this prime squared. Obviously, a prime will not be removed because it is squarefree. Consider now a number m that is divisible by a prime squared and therefore is not squarefree. We have $m = p^2k$ where p is a prime so this number will be crossed out when $i = p$ in the last loop. This shows that the algorithm is correct.

Let's analyze the complexity of Algorithm 4.5. The main loop runs for time $O\left(\lfloor\sqrt{n}\rfloor^2\right) = O(n)$. The auxiliary loop runs for time

$$\sum_{5 \le p \le \sqrt{n}} \frac{n}{p^2} = n \cdot O(1) = O(n),$$

so the total time is $O(n)$. Also, from Theorem 19, we know that the number of squarefree numbers smaller than n is asymptotically $\left(1-\frac{6}{\pi^2}\right)n$, so the auxiliary loop (which sieves out square numbers) cannot be further optimized. However, with some additional work the running time of this algorithm can be improved to be $O\left(\frac{n}{\log\log n}\right)$ (see Section 4.5).

```
Require: $A[i]$ for $5 \le i \le n$ - an array of numbers to sieve for primes
Ensure: $A[i] = 1$ iff $i$ is a prime number
for $i \leftarrow 5 \ldots n$ do
    $A[i] \leftarrow 0$
end for
for $x \leftarrow 1 \ldots \lfloor \sqrt{n} \rfloor$ do
    for $y \leftarrow 1 \ldots \lfloor \sqrt{n} \rfloor$ do
        if $4x^2 + y^2 \equiv 1$ or $5 \pmod{12} \ \wedge \ 4x^2 + y^2 \le n$ then
            $A[4x^2 + y^2] = 1 - A[4x^2 + y^2]$
        end if
        if $3x^2 + 4y^2 \equiv 7 \pmod{12} \ \wedge \ 3x^2 + 4y^2 \le n$ then
            $A[3x^2 + 4y^2] = 1 - A[3x^2 + 4y^2]$
        end if
        if $3x^2 - y^2 \equiv 11 \pmod{12} \ \wedge \ 3x^2 - y^2 \le n \ \wedge \ x > y$ then
            $A[3x^2 - y^2] = 1 - A[3x^2 - y^2]$
        end if
    end for
end for
for $i \leftarrow 5 \ldots \lfloor \sqrt{n} \rfloor$ do
    if $A[i] = 1$ then
        for $k \in \{i^2, 2i^2, 3i^2, \ldots\} \ \wedge \ k \le n$ do
            $A[k] = 0$
        end for
    end if
end for
```

Algorithm 4.5: Sieve of Atkin.

The authors of [3] also give a highly optimized implementation of their algorithm. It is available at [4] and will be used to generate primes in Section 6.3.

4.5 Possible improvements

It is easy to obtain a theoretical bound on the number of steps needed to obtain a list of primes up to the number n. Let's start, like always, with a list A of numbers from 2 to n. We require at the end of an algorithm to be able to distinguish prime numbers from the rest. It means that $A[i] \neq A[j]$ for every pair of numbers where either i or j (but not both!) is prime. But this means that the algorithm has to change the value of at least $\pi(n)$ (number of primes) or $(n - \pi(n))$ (number of composites) elements. Therefore the running time of *any* prime sieving algorithm is at least

$$\min(\pi(n), n - \pi(n)) = \min\left(n/\log n, n - n/\log n\right) = n/\log n,$$

for a big enough n and from Prime Number Theorem (Theorem 13). We don't know any algorithm running in such a time. This also shows why the previous algorithms were at best linear – they cross out composites. There are asymptotically n of them in the range from 1 to n, so this is the minimal number of steps if one sieves out *every* composite number.

However, as we will later see, the number of sieved composites can be made smaller by simple techniques.

Wheel data structure

A *wheel* is a data structure that allows to explicitly ignore numbers that must be composite numbers. For example, there is no need to consider even numbers (apart from 2). We could ignore as well multiples of 3, of 5, etc. This leads to the idea of considering M_k, a product of first k primes

$$M_k = p_1 \cdot p_2 \cdot \ldots \cdot p_k.$$

The wheel is a list of M_k elements indexed from 0 to $M_k - 1$. The k-th element represents numbers congruent to k mod M_k. We set, for $i \in \{0, 1, \ldots, M_k - 1\}$:

$$M[i] = \begin{cases} 0, & \text{if } i \text{ is not coprime to } M_k, \\ d_i, & \text{if } i \text{ is coprime to } M_k, \end{cases} \tag{4.4}$$

where d_i is the smallest positive integer, such that $i + d_i$ is coprime to M_k.

We can compute this data structure for the given k in the time proportional to M_k ([13]). With this additional information we can "skip" the numbers that cannot be primes. To see how many numbers we will skip, let's compute $\varphi(M_k)$, namely

$$\varphi(M_k) = \varphi(p_1) \cdot \varphi(p_2) \cdot \ldots \cdot \varphi(p_k) = (p_1 - 1) \cdot (p_2 - 1) \cdot \ldots \cdot (p_k - 1).$$

Consequently, we observe that the ratio of numbers coprime to M_k to all numbers is

$$\frac{\varphi(M_k)}{M_k} = \prod_{p \le p_k} \frac{p-1}{p} = \prod_{p \le p_k} \left(1 - \frac{1}{p}\right).$$

From Mertens' Third Theorem (Theorem 12) we obtain positive constants c_1 and c_2, such that

$$\frac{c_1}{\log p_k} < \frac{\varphi(M_k)}{M_k} < \frac{c_2}{\log p_k}, \tag{4.5}$$

for k large enough.

In practice M_k is taken to be between $n^{1/3}$ and $n^{1/2}$. This assumption and Theorem 5 imply that

$$\frac{c_3}{\log\log n} < \frac{1}{\log p_k} < \frac{c_4}{\log\log n} \tag{4.6}$$

for some positive contants c_3 and c_4. The inequalities (4.5) and (4.6) together give

$$\frac{c_1 c_3}{\log\log n} < \frac{\varphi(M_k)}{M_k} < \frac{c_2 c_4}{\log\log n}.$$

Therefore, the fraction of integers that we will have to check for the primality is bounded from above and below by the function $\frac{1}{\log\log n}$ multiplied by some positive constants. Informally, the work to be done will be reduced by a factor of $\frac{1}{\log\log n}$ (the upper bound) and cannot be improved upon anymore by this method (the lower bound).

The method will reduce the number of steps in sieve of Eratosthenes, sieve of Pritchard or sieve of Atkin, so that the running complexity for each algorithm can be:

- Sieve of Eratosthenes – $O(n)$,
- Sieve of Pritchard – $O\big(n/\log\log n\big)$,
- Sieve of Atkin – $O\big(n/\log\log n\big)$.

For more details see [37].

Segmented sieve

All the presented methods use memory of size $O(n)$ to store the whole list of numbers. Another approach is to split the whole range of the numbers to sieve into “segments”. For example, to sieve primes in the range $[1,100]$, one can first get primes in the range $[1,25]$ then, using the primes just sieved, find primes in the range $[26,50]$ and so on.

There are two reasons to do that:

- Smaller memory usage – this can effectively lead to an algorithm using $O(\sqrt{n})$ of space.
- Better locality of the memory – smaller segments improve the locality of the memory and can significantly improve the speed of computation. Currently processors have very fast cache memories that are by an order of magnitude faster than RAM memory (cf. Table 4.1). Therefore fitting the working set memory into processor’s cache may dramatically accelerate the process of sieving.

In [19] the sieve of Atkin’s is improved to use only space of order $O(n^{1/3+\epsilon})$. In [38] the other researchers show an algorithm with the running time of only $O\big(n(\log n)^2/\log\log n\big)$, but with a conjectured memory consumption of order $O\big((\log n)^3/\log\log n\big)$. The conjectured complexity depends on the validity of Extended Riemann Hypothesis.

Memory type	Number of cycles
Register	≤ 1
L1 cache	~ 3
L2 cache	~ 14
Main Memory	~ 240

Table 4.1: Memory access speed in Pentium M

4.6 Summary

This chapter gave an exposition of important prime sieving algorithms. Although the most popular is the Sieve of Eratosthenes, we showed that it is not the most optimal algorithm for this problem. Apart from presentation of this algorithms, we also showed possible improvements that can be made to improve both running and space complexity. Finally, we proved two out of three theorems needed in the Sieve of Atkin in a different way than the authors.

Chapter 5

Sieve methods

5.1 History and results

After the exposition of sieving algorithms, we will describe sieving techniques used to obtain powerful theorems on the distribution of primes. In particular, with Brun's sieve we will be able to obtain the famous theorem proved in 1916 by Viggo Brun ([5]), stating that the sum

$$\sum_{\substack{p,p+2 \\ \text{are primes}}} \left(\frac{1}{p} + \frac{1}{p+2}\right)$$

converges.

A motivating example is to see problems in number theory which were attacked by sieve methods (we mostly follow [24]). We have of course the famous problem communicated by Goldbach to Euler:

Conjecture 7 (Goldbach's Conjecture). *Let n be an even integer greater than 2. Then n can be represented as a sum of two prime numbers.*

We have also a generalization of Bertrand's postulate (cf. Theorem 4):

Conjecture 8 (On primes in the interval $(n, n+\sqrt{n})$). *For n big enough, the interval $(n, n+\sqrt{n})$ contains a prime.*

Finally there is the already mentioned conjecture on the infinitude of prime twins (Conjecture 1).

All of them are long standing problems, tantalizing the mathematicians for centuries. For a very long time there was virtually no method to approach them. That was till around 1920 when Viggo Brun showed the following theorems:

Theorem 24. *Every sufficiently large even integer can be represented as a sum of two numbers each of which has at most nine prime factors,*

Theorem 25. *If n is large enough, then the interval $(n, n+\sqrt{n})$ contains a number with at most eleven prime factors,*

Theorem 26. *There are infinitely many pairs of numbers of difference 2, such that both of them have at most nine prime factors.*

He also showed

Theorem 27. *For sufficiently large x, the number of prime twins not exceeding x, denoted $\pi_2(x)$, is*

$$\pi_2(x) \leq \frac{100x}{\log^2 x}.$$

Quite unjustly, Brun's methods were not recognized immediately. It seems that mathematicians did not believe that such elementary methods (Brun's sieve is basically a combinatorial tool) could be used to approach such difficult conjectures like those given above. There is an anecdote that E. Landau did not read Brun's paper for a decade because of this superstition. This skepticism was partially overcome when in 1933 L. G. Shnirelman proved his weak statement of Goldbach's conjecture ([35]):

Theorem 28. *There exists a positive integer s, such that every sufficiently large integer is the sum of at most s primes.*

Another major milestone was set in 1947 by A. Selberg. Selberg's sieve method is simpler to understand and quite often leads to better results. This again is the example of the *upper bound sieve.*

The methods of Brun and his successors work with numbers smaller than N, which are then sieved using primes not exceeding a certain threshold N^c. If we could set $c = \frac{1}{2}$, then the remaining numbers would be primes, of course, and we could estimate and bound precisely the number of primes in this range. But this is in general beyond the reach. One can see that all theorems of Brun above refer to numbers with a bounded number of prime factors. Some work was done to overcome this limitation. For example P. Kuhn in 1941 realized that one can obtain better bounds for the number of prime factors by "weighting" the sieve in a certain way, relaxing the restriction.

These ideas were used by J. R. Chen who in 1975 established

Theorem 29 (Chen's Theorem I). *If n is large enough, then the interval $(n, n+\sqrt{n})$ contains an integer with at most two prime factors.*

He also showed ([7])

Theorem 30 (Chen's Theorem II). *Every sufficiently large even number can be written as the sum of either two primes, or a prime and an integer that is a product of at most 2 primes,* and

Theorem 31 (Chen's Theorem III). *There are infinitely many pairs of numbers of difference 2, such that the smaller number in the pair is a prime and the larger is a product of at most two primes.*

These results are proven using basically the same approach, it seems that all these problems are deeply connected.

Sieve theory is a very exciting area of research. One of the relatively recent results that used deep sieve methods was a result of H. Iwaniec and J. Friedlander, who showed ([18])

Theorem 32. *There are infinitely many primes of the form $x^2 + y^4$,*

and a result of D. R. Heath-Brown ([27]):

Theorem 33. *There are infinitely many primes of the form $x^3 + 2y^3$.*

These striking results show that the sieve theory can continuously provide interesting and better results in number theory.

In the following sections of this chapter we generally follow [40].

5.2 Sieve of Eratosthenes

Let's write

$$P = \prod_{p \le \sqrt{x}} p.$$

An integer n, such that $\sqrt{x} < n \le x$ is a prime number if and only if P and n are coprime or $(P, n) = 1$. To formalize it, we can write

$$\pi(x) - \pi(\sqrt{x}) + 1 = \sum_{n \le x} \delta((n, P)),$$

where

$$\delta(n) = \begin{cases} 1, & \text{if } n = 1, \\ 0, & \text{otherwise.} \end{cases}$$

But we have

$$\delta((n,P)) = \sum_{d|(n,P)} \mu(d) = \sum_{\substack{d|n \\ d|P}} \mu(d). \tag{5.1}$$

Hence:

$$\pi(x) - \pi(\sqrt{x}) + 1 = \sum_{n \le x} \sum_{\substack{d|n \\ d|P}} \mu(d) = \sum_{de \le x} \sum_{d|P} \mu(d) = \sum_{d|P} \mu(d) \sum_{e \le \frac{x}{d}} 1 =$$

$$= \sum_{d|P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor. \tag{5.2}$$

In fact this is a concealed application of inclusion-exclusion principle. Indeed, if we let $P = p_1 p_2 \dots p_k$, then

$$\sum_{d|P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = \sum_{\substack{d|P \\ \nu(d) \text{ even}}} \left\lfloor \frac{x}{d} \right\rfloor - \sum_{\substack{d|P \\ \nu(d) \text{ odd}}} \left\lfloor \frac{x}{d} \right\rfloor = n - \left\lfloor \frac{n}{p_1} \right\rfloor - \left\lfloor \frac{n}{p_2} \right\rfloor - \dots - \left\lfloor \frac{n}{p_k} \right\rfloor +$$
$$+ \left\lfloor \frac{n}{p_1 p_2} \right\rfloor + \left\lfloor \frac{n}{p_1 p_3} \right\rfloor + \dots + \left\lfloor \frac{n}{p_1 p_k} \right\rfloor + \left\lfloor \frac{n}{p_2 p_3} \right\rfloor + \dots + \left\lfloor \frac{n}{p_{k-1} p_k} \right\rfloor -$$
$$- \left\lfloor \frac{n}{p_1 p_2 p_3} \right\rfloor - \dots - \left\lfloor \frac{n}{p_{k-2} p_{k-1} p_k} \right\rfloor +$$
$$+ \dots + (-1)^k \left\lfloor \frac{n}{p_1 p_2 \cdot \dots \cdot p_k} \right\rfloor .$$

This formula can be used to obtain a value of $\pi(x)$ for very large values of x ([41, 288–292]).

Let us go back to (5.2) and estimate its value by taking $\lfloor x/d \rfloor = x/d + O(1)$. We obtain

$$\pi(x) - \pi(\sqrt{x}) + 1 = x \sum_{d|P} \frac{\mu(d)}{d} + O\left(2^{\pi\sqrt{x}}\right) = x \prod_{p|P} \left(1 - \frac{1}{p}\right) + O\left(2^{\pi\sqrt{x}}\right) =$$
$$= x \prod_{p \le \sqrt{x}} \left(1 - \frac{1}{p}\right) + O\left(2^{\pi\sqrt{x}}\right).$$

This can be estimated by means of Mertens' Theorem (Theorem 11). Then the main term is

$$O\left(\frac{x}{\log x}\right)$$

and agrees with the Prime Number Theorem (Theorem 13). Sadly, the error term $O\left(2^{\pi\sqrt{x}}\right)$ is actually bigger and this approximation is useless.

To overcome this, let's take a parameter y instead of $\sqrt{x}$ above. Then, by the same computation, we will arrive at:

$$\pi(x) - \pi(y) + 1 = x \prod_{p \le y} \left(1 - \frac{1}{p}\right) + O\left(2^{\pi(y)}\right) = x \frac{e^{-\gamma} + o(1)}{\log y} + O\left(2^y\right),$$

where e is the Napier's constant and γ is the Euler-Mascheroni constant. Thus

$$\pi(x) \le x \frac{e^{-\gamma} + o(1)}{\log y} + O\left(2^y\right).$$

To equate the two terms, the optimal choice is to take $y = \log x$. This gives

$$\pi(x) \le O\left(\frac{x}{\log\log x}\right), \tag{5.3}$$

which is obviously inferior to the approximation from the Prime Number Theorem. However, the generality of this method is amazing, as we will see.

Let's find the basic ingredients in the recipe above. First, we have a set that we sieve – a set of numbers not greater than x in our case. We also have a characteristic function (*sifting function*) of the subset that we want to sieve out (or not to sieve out) – this is (5.1). Finally, by representing the sifting function in a convenient way and using a derived approximation (i.e., Mertens' Theorem) we get the result.

Therefore, let us use our new tool in a different setting. We will prove

Theorem 34 (Asymptotic formula for the number of squarefree numbers). *Let $Q(x)$ be the number of squarefree numbers not bigger than x. Then*

$$Q(x) = \frac{6}{\pi^2}x + O\left(\sqrt{x}\right). \tag{5.4}$$

Proof. The set we are about to sieve is again a set of numbers not greater than x. We must now find a characteristic function for the set of squarefree numbers. Obviously, the absolute value of Möbius function is such a function. One can try this, but the obtained sum will not be easy to work with. We instead observe that the following function is also a characteristic function for squarefree numbers:

$$s(n) = \sum_{d^2|n} \mu(d).$$

First observation is that this function is multiplicative. Indeed, if $n = ab$ and $(a,b) = 1$ then

$$\sum_{d^2|ab} \mu(d) = \sum_{e^2|a}\sum_{f^2|b} \mu(ef) = \left(\sum_{e^2|a} \mu(e)\right)\left(\sum_{f^2|b} \mu(f)\right).$$

We used the fact that if ab is square, for a, b coprime, then both a and b are squares. Consequently, it's enough to check only the value of $s(n)$ for powers of prime numbers. Hence, let $n = p^k$. If $k \leq 1$ then the only term in the sum is $\mu(1) = 1$. If $k \geq 2$, we get

$$s\left(p^k\right) = \mu(1) + \mu(p) + \mu\left(p^2\right) + \ldots + \mu\left(p^{\lfloor k/2 \rfloor}\right) = 1 - 1 + 0 = 0.$$

We now have

$$\begin{aligned} Q(x) &= \sum_{n\leq x} s(n) = \sum_{n\leq x}\sum_{d^2|n} \mu(d) = \sum_{d^2e=n\leq x} \mu(d) = \\ &= \sum_{d\leq\sqrt{x}}\sum_{e\leq\frac{x}{d^2}} \mu(d) = \sum_{d\leq\sqrt{x}} \mu(d)\left\lfloor\frac{x}{d^2}\right\rfloor = x\sum_{d\leq\sqrt{x}} \frac{\mu(d)}{d^2} + O\left(\sqrt{x}\right) = \\ &= x\left(\frac{1}{\zeta(2)} - \sum_{d>\sqrt{x}} \frac{\mu(d)}{d^2}\right) + O\left(\sqrt{x}\right) = \frac{x}{\zeta(2)} + O\left(x\sum_{d>\sqrt{x}} \frac{1}{d^2}\right) + O\left(\sqrt{x}\right). \end{aligned}$$

But, using Euler's formula (Theorem 7) we may deduce that

$$x\sum_{d>\sqrt{x}}\frac{1}{d^2}=O\left(x\frac{1}{\sqrt{x}}\right)=O\left(\sqrt{x}\right).$$

This finishes the proof. □

This is a very interesting fact - roughly 61% of the numbers are squarefree.

We may know state the *sieve problem*. Let $\mathscr{A}$ be a set of any positive integers. The general goal is to estimate the number of primes in $\mathscr{A}$. Ideally, an asymptotic formula is obtained.

In practice, it is convenient to introduce the following definitions and notation:

$$\begin{aligned}\mathscr{A}_d &= \text{card }\{n\in\mathscr{A}:n\equiv 0 \pmod{d}\},\\ P_y &= \prod_{p\le y}p,\\ S(\mathscr{A},\mathscr{P},y) &= \text{card }\left\{n\in\mathscr{A}:(n,P_y)=1\right\},\end{aligned}$$

because it's easier to work with $S(\mathscr{A},\mathscr{P},y)$. Assume further that $\mathscr{A}_d$ can be written as

$$\mathscr{A}_d=X\frac{\omega(d)}{d}+R_d,$$

for some real X, some multiplicative function ω and a remainder term R_d, hopefully small. Finally, let $I(n)$ be a characteristic function of primes in the set $\mathscr{A}$.

We can now deduce that

$$\begin{aligned}S(\mathscr{A},\mathscr{P},y) &= \sum_{n\in\mathscr{A}}I(n)=\sum_{n\in\mathscr{A}}\delta((n,P_y))=\sum_{n\in\mathscr{A}}\sum_{d|(n,P_y)}\mu(d)=\\ &=\sum_{n\in\mathscr{A}}\sum_{\substack{d|n\\ d|P_y}}\mu(d)=\sum_{de\in\mathscr{A}}\sum_{d|P_y}\mu(d)=\\ &=\sum_{d|P_y}\mu(d)\sum_{ed\in A}1=\sum_{d|P_y}\mu(d)\mathscr{A}_d=\\ &=\sum_{d|P_y}\mu(d)\left(X\frac{\omega(d)}{d}+R_d\right)=X\sum_{d|P_y}\frac{\mu(d)\omega(d)}{d}+\sum_{d|P_y}\mu(d)R_d=\\ &=X\prod_{p\le y}\left(1-\frac{\omega(p)}{p}\right)+\sum_{d|P_y}\mu(d)R_d.\end{aligned}$$

Clearly, if the remainder R_d is sufficiently small and y is appropriately chosen, then a precise approximation can be obtained.

Let's go to the example at the beginning of this chapter. To estimate $\pi(x)$ we used

$$\begin{aligned}\mathscr{A} &= \{n : n \le x\},\\ \mathscr{A}_d &= \left\lfloor \frac{x}{d} \right\rfloor = \frac{x}{d} + O(1) \qquad (X = 1,\, \omega(n) = 1,\, R_d = O(1)),\\ y &= \log x.\end{aligned}$$

Unfortunately, the unrefined sieve of Eratosthenes is not enough to get estimates about twin prime numbers. Although it is possible to improve the sieve of Eratosthenes so that it is almost as powerful as the sieve of Brun ([9]), we do not follow this path. Instead, in the next section, we present the sieve of Brun.

5.3 Brun's sieve

Whereas the sieve of Eratosthenes is based on the equality between multiplicative functions of the form $\mu * \mathbf{1} = \delta$, i.e.,

$$\sum_{d|n} \mu(d) = \delta(n).$$

The idea of Brun was to bound δ function by two functions μ_1 and μ_2, such that

$$\mu_1 * \mathbf{1} \le \delta \le \mu_2 * \mathbf{1}.$$

The choice of μ_1 and μ_2 leading to Brun's sieve is

$$\begin{aligned}\mu_1(n) &= \mu(n)\chi_{2h+1}(n),\\ \mu_2(n) &= \mu(n)\chi_{2h}(n),\end{aligned}$$

for any $h \ge 0$, where χ_t is a characteristic function of numbers having at most t prime factors. Actually Brun showed the following

Theorem 35 (Brun's Theorem). *For any $h \ge 0$ we have*

$$\mu_1(n) * \mathbf{1} \le \delta \le \mu_2(n) * \mathbf{1}. \tag{5.5}$$

Proof. When n is not squarefree then (5.5) is true, because all sides of the inequality are zero. Let's therefore consider n that is squarefree and let n be a product of k different primes, that is $\omega(n) = k$. For any $i \le k$ there are exactly $\binom{k}{i}$ numbers with i prime divisors dividing n. We have

$$\begin{aligned}(\mu \cdot \chi_t) * \mathbf{1}\,(n) &= \sum_{\substack{d|n\\ \omega(d)\le t}} \mu(d) = \sum_{i\le t} (-1)^i \binom{k}{i} =\\ &= \sum_{i\le t} (-1)^i \left(\binom{k-1}{i} + \binom{k-1}{i-1}\right) = (-1)^t \binom{k-1}{t},\end{aligned}$$

since the sum is telescopic. But this means that $\mu \cdot \chi_t$ is positive for even t and negative otherwise. This is exactly what we wanted to show. □

Keeping the same notation as in the previous section, we get

$$\begin{aligned}S(\mathscr{A},\mathscr{P},y) &= \sum_{n\in\mathscr{A}} \delta((n,P_y)) \le \sum_{n\in\mathscr{A}} \sum_{d|(n,P_y)} \mu(d)\chi_{2h}(d) = \\ &= \sum_{n\in\mathscr{A}} \sum_{\substack{d|n \\ d|P_y}} \mu(d)\chi_{2h}(d) = \sum_{de\in\mathscr{A}} \sum_{d|P_y} \mu(d)\chi_{2h}(d) = \\ &= \sum_{d|P_y} \mu(d)\chi_{2h}(d) \sum_{ed\in A} 1 = \sum_{d|P_y} \mu(d)\chi_{2h}(d)\mathscr{A}_d = \\ &= \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} \mu(d)\mathscr{A}_d.\end{aligned}$$

Similarly we may obtain the lower bound. These two bounds together give

$$\sum_{\substack{d|P_y \\ \omega(d)\le 2h+1}} \mu(d)\mathscr{A}_d \le S(\mathscr{A},\mathscr{P},y) \le \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} \mu(d)\mathscr{A}_d. \tag{5.6}$$

The integer parameter h gives an additional degree of freedom. By setting it to a proper value, one can improve the results of the sieve of Eratosthenes. Let's see how it can be used to obtain an improved bound on the prime counting function $\pi(x)$.

We have, just as in the derivation of (5.3), that

$$\begin{aligned}\pi(x) &\le \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} \mu(d)\left\lfloor \frac{x}{d} \right\rfloor + y = \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} \mu(d)\frac{x}{d} + O\left(y + \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} 1 \right) = \\ &= x\sum_{d|P_y} \frac{\mu(d)}{d} - x \sum_{\substack{d|P_y \\ \omega(d)>2h}} \frac{\mu(d)}{d} + O\left(y + \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} 1 \right) = \\ &= x\prod_{p\le y}\left(1-\frac{1}{p}\right) + O\left(y + \sum_{\substack{d|P_y \\ \omega(d)\le 2h}} 1 + x\sum_{\substack{d|P_y \\ \omega(d)>2h}} \frac{1}{d} \right).\end{aligned} \tag{5.7}$$

The second error term is bounded by y^{2h} since this is a bound for any d in the sum. To estimate the third term, let's take any $u \ge 1$. We have

$$\sum_{\substack{d|P_y \\ \omega(d)\le 2h}} \frac{1}{d} \le \sum_{d|P_y} \frac{u^{\omega(d)-2h}}{d} = u^{-2h}\prod_{p\le y}\left(1+\frac{u}{p}\right) \le \exp\left(-2h\log u + u\sum_{p\le y}\frac{1}{p}\right).$$

By taking

$$u = \frac{2h}{\sum_{p\le y}\frac{1}{p}},$$

we get (by Theorem 10)

$$2h = u\sum_{p\le y}\frac{1}{p} = u\log\log y + B\cdot u + O\left(\frac{u}{\log y}\right) \ll_u u\log\log y.$$

Consequently, one obtains that the third error term is

$$\ll_u \exp\left(2h(1-\log u)\right) = \left(\log y\right)^{u(1-\log u)} = \left(\log y\right)^{u-u\log u}.$$

Now, it's easy to see that $u - u\log u$ is greater than 3 if $u > 5$. Moreover, for sufficiently large y there must be an u (depending on y), such that $5 < u < 6$ and

$$h = \frac{1}{2}u\sum_{p\le y}\frac{1}{p}$$

is an integer. We now know that the third term is smaller than

$$x(\log y)^{-3}.$$

Similarly, the second term is smaller than

$$y^{2h} = y^{u\log\log y} = \exp\left(6\log y\log\log y\right).$$

If we finally set

$$\log y = \frac{\log x}{10\log\log x},$$

then (5.3) becomes

$$y^{2h} \le \exp\left(6\frac{\log x}{10\log\log x}\log\frac{\log x}{10\log\log x}\right) = x^{\frac{3}{5\log\log x}}\frac{\log x}{10\log\log x} < x^{\frac{3}{4}},$$

for x big enough.

By collecting all the bounds we obtained before, (5.7) becomes

$$\begin{aligned}\pi(x) &\le x\prod_{p\le y}\left(1-\frac{1}{p}\right) + O\left(x^{\frac{3}{4}} + x\frac{(\log\log x)^3}{(\log x)^3}\right)\\ &= O\left(x\frac{\log\log x}{\log x}\right) + o(x).\end{aligned}$$

So the final result is

$$\pi(x) \ll \frac{x\log\log x}{\log x}.$$

This is still inferior to the Prime Number Theorem, but is a substantial improvement over (5.3). But almost the same reasoning allows us to prove the theorem of Brun.

Theorem 36 (Brun's Theorem II). *The number of primes p, such that $p < x$ and $p+2$ is a prime too, denoted $\pi_2(x)$, satisfies*

$$\pi_2(x) \ll \left(\frac{\log\log x}{\log x}\right)^2.$$

Proof. The set of numbers to sieve is

$$\mathscr{A} = \{m(m+2) : m \leq x\}.$$

As before, we have

$$\pi_2(x) \leq S(\mathscr{A}, \mathscr{P}, y) + y \leq \sum_{\substack{d \mid P_y \\ \omega(d) \leq 2h}} \mu(d) \mathscr{A}_d,$$

where $\mathscr{A}_d$ is the number $\rho(d)$ of solutions $m \leq x$ to the congruence

$$m(m+2) \equiv 0 \ (\mathrm{mod}\ d).$$

We can solve it only for primes and then use the Chinese Remainder Theorem to get the final result. We obviously have $\rho(2) = 1$. Also, for odd prime p, $\rho(p) = 2$ since either $p \mid m$ or $p \mid (m+2)$. Now, each interval of length d contains $\rho(d)$ numbers m counted in the set $\mathscr{A}_d$. This means that one can write

$$\mathscr{A}_d = x\frac{\rho(d)}{d} + O\big(\rho(d)\big).$$

Performing calculations parallel to those above, we will get

$$\pi_2(x) \leq x \sum_{d \mid P_y} \frac{\mu(d)\rho(d)}{d} + O\left(y + \sum_{\substack{d \mid P_y \\ \omega(d) \leq 2h}} \rho(d) + x \sum_{\substack{d \mid P_y \\ \omega(d) > 2h}} \frac{\rho(d)}{d} \right).$$

The main term is

$$x \sum_{d \mid P_y} \frac{\mu(d)\rho(d)}{d} = x\left(1 - \frac{1}{2}\right) \prod_{3 \leq p \leq y} \left(1 - \frac{2}{p}\right) \leq 2x \prod_{3 \leq p \leq y} \left(1 - \frac{1}{p}\right)^2 \sim C\frac{x}{(\log y)^2},$$

for some constant C. Now, by the same method as above, we will obtain

$$h = c_1 \log\log y + O(1)$$
$$\log y = c_2 \frac{\log x}{\log\log x},$$

with appropriate constants c_1 and c_2, such that the error term is smaller than the main term. But then the main term is

$$C\frac{x}{(\log y)^2} = C' x \left(\frac{\log\log x}{\log x}\right)^2,$$

for another constant C'. This is exactly what we wanted to prove. □

From this result we obtain as an easy corollary, the main result of this chapter.

Theorem 37 (Brun's Theorem III).

$$\sum_{\substack{p,p+2 \\ \textit{are primes}}} \left(\frac{1}{p}+\frac{1}{p+2}\right) < \infty.$$

Proof. We use Abel's summation formula (Theorem 6) with

$$a(n) = \begin{cases} 1, & \text{if both } n, n+2 \text{ are primes,} \\ 0, & \text{otherwise,} \end{cases}$$
$$f(x) = \frac{1}{x}.$$

Then we get

$$\begin{aligned} \frac{1}{2} \sum_{\substack{p,p+2 \\ \text{are primes}, p\le x}} \left(\frac{1}{p}+\frac{1}{p+2}\right) \le \sum_{\substack{p,p+2 \\ \text{are primes}, p\le x}} \frac{1}{p} = \sum_{3\le n\le x} a(n) f(n) = \\ = \frac{\pi_2(x)}{x} + \int_3^x \frac{\pi_2(t)}{t^2}\, dt \ll \\ \ll \left(\frac{\log\log x}{\log x}\right)^2 + \int_3^x \frac{1}{t}\left(\frac{\log\log t}{\log t}\right)^2 dt. \end{aligned}$$

The first term converges to zero as x goes to the infinity. Moreover, the integral

$$\int_3^x \frac{1}{t}\left(\frac{\log\log t}{\log t}\right)^2 dt$$

converges as x approaches the infinity. This means that the sum we started with is bounded and therefore converges. □

5.4 Summary

In this chapter we proved a weak statement of Brun's theorem. From this fact, it follows directly that the sum of inverses of twin primes converges to a finite value. To achieve this remarkable result we used Brun's combinatorial sieve – a relatively modern tool in number theory.

Chapter 6

Related constants

There are two constants related to the Twin Prime Conjecture:

- the twin prime constant – defined already in Chapter 3,
- the Brun's constant.

As we will see, the former one is much easier to compute.

6.1 Description of the environment

The experiments were performed on a PC computer with Intel Core i5-2410M (2.3 GHz) processor and 8GB of DDR3 operating memory. To generate primes, *primegen* program is used ([4]). It is written in C language and was compiled using GCC compiler (version 4.5.2). High precision computation was performed using *mpmath* library ([28]) at version 0.17 with GMP library ([20]) at version 4.3.2 as a backend. The version of Python distribution was 2.7.1+.

All programs written to obtain results from this thesis can be downloaded from `https://bitbucket.org/thinred/twinprimes/`.

6.2 Computation of C_2

Analysis

In this section the problem of calculating C_2 constant is described. The goal is to obtain a computationally feasible formula and to calculate the constant with a high precision.

Let's start with the infinite product that defines C_2, i.e.,

$$C_2 = \prod_{p\geq 3}\left(1 - \frac{1}{(p-1)^2}\right). \tag{6.1}$$

Primes needed	Correct digits
3	$\underline{0.6}601618158468$
6	$\underline{0.66}01618158468$
31	$\underline{0.660}1618158468$
305	$\underline{0.6601}618158468$
1019	$\underline{0.66016}18158468$
23378	$\underline{0.660161}8158468$
45599	$\underline{0.6601618}158468$
624284	$\underline{0.66016181}58468$

Table 6.1: Convergence of the infinite product (6.1).

The convergence of this product is very slow. Table 6.1 shows how many primes are needed to obtain the first, the second, etc., digit of C_2.

For the sake of completeness, we provide the following formulas for C_2 ([23]):

$$C_2 = \sum_{\substack{n\geq 1 \\ n \text{ odd}}} \frac{\mu(n)}{\varphi^2(n)} = \frac{1}{8} \sum_{\substack{n\geq 1 \\ n \text{ odd}}} \frac{\mu(n) 2^{\nu(n)} \log^2 n}{n}.$$

Unfortunately, they are equally impractical, because of the required computation of arithmetic functions $\varphi(n)$ (Euler's totient function) and $\nu(n)$ (number of prime factors of n) and slow convergence.

To derive a better formula, we will use a method described in [17]. Let's start with some helpful definitions.

Definition (Truncated zeta function). Let q be a prime number. The truncated zeta function $\zeta_{\geq q}$ is defined as

$$\zeta_{\geq q}(s) = \prod_{p\geq q} \left(1 - p^{-s}\right)^{-1} = \zeta(s) \prod_{p<q} \left(1 - p^{-s}\right). \tag{6.2}$$

Definition (Truncated prime zeta function). Let q be a prime number. The truncated prime zeta function $P_{\geq q}$ is defined as

$$P_{\geq q}(s) = \sum_{p\geq q} p^{-s}. \tag{6.3}$$

We have

$$\begin{aligned} \log \zeta_{\geq q}(s) &= -\sum_{p\geq q} \log\left(1 - p^{-s}\right) = \sum_{p\geq q} \sum_{m\geq 1} \frac{1}{m p^{sm}} = \\ &= \sum_{m\geq 1} \frac{1}{m} \sum_{p\geq q} p^{-sn} = \sum_{m\geq 1} \frac{1}{m} P_{\geq q}(sm). \end{aligned}$$

Using Möbius inversion formula, we obtain another representation of $P_{\geq q}$, that is,

$$P_{\geq q}(s) = \sum_{m\geq 1} \frac{\mu(m)}{m} \log \zeta_{\geq q}(sm). \tag{6.4}$$

By applying the logarithm on both sides of (6.1), we get

$$\begin{aligned}
\log C_2 &= \sum_{p\geq 3} \log\left(1-\frac{1}{(p-1)^2}\right) = \sum_{p\geq 3} \log \frac{1-\frac{2}{p}}{\left(1-\frac{1}{p}\right)^2} = \\
&= \sum_{p\geq 3} \left(\log\left(1-\frac{2}{p}\right) - 2\log\left(1-\frac{1}{p}\right)\right) = \\
&= \sum_{p\geq 3} \left(\sum_{m\geq 1} \frac{2}{mp^m} - \sum_{m\geq 1} \frac{2^m}{mp^m}\right) = \\
&= \sum_{p\geq 3} \sum_{m\geq 1} \frac{2-2^m}{mp^m} = \sum_{p\geq 3} \sum_{m\geq 2} \frac{2-2^m}{mp^m} =
\end{aligned}$$

$$= \sum_{m\geq 2} \frac{2-2^m}{m} \sum_{p\geq 3} p^{-m} = \sum_{m\geq 2} \frac{2-2^m}{m} P_{\geq 3}(m) = \tag{6.5}$$

$$\begin{aligned}
&= \sum_{m\geq 2} \frac{2-2^m}{m} \sum_{k\geq 1} \frac{\mu(k)}{k} \log\zeta_{\geq 3}(mk) = \\
&= \sum_{k\geq 1} \sum_{m\geq 2} \frac{2-2^m}{m} \frac{\mu(k)}{k} \log\zeta_{\geq 3}(mk) = \\
&= \sum_{\substack{k,m\geq 1 \\ (k,m)\neq(1,1)}} \frac{2-2^m}{m} \frac{\mu(k)}{k} \log\zeta_{\geq 3}(mk),
\end{aligned}$$

by applying (6.4) and noting that $2-2^m$ is zero for $m=1$.

Now, taking $n=mk$ and $d=k$, we observe that as m and k change, n runs over all natural numbers greater than 2 and d runs over divisors of n. Therefore

$$\begin{aligned}
\log C_2 &= \sum_{n\geq 2} \sum_{d|n} \frac{2-2^{n/d}}{n} \mu(d) \log\zeta_{\geq 3}(n) = \\
&= 2\sum_{n\geq 2} \frac{\log\zeta_{\geq 3}(n)}{n} \sum_{d|n} \mu(d) - \sum_{2\geq n} \log \frac{\log\zeta_{\geq 3}(n)}{n} \sum_{d|n} \mu(d) 2^{n/d} = \\
&= -\sum_{n\geq 2} \log \frac{\log\zeta_{\geq 3}(n)}{n} \sum_{d|n} \mu(d) 2^{n/d} = \\
&= \log \prod_{n\geq 2} [\zeta_{\geq 3}(n)]^{-\sum_{d|n} \mu(d) 2^{n/d}},
\end{aligned}$$

since $\sum_{d|n} \mu(n)$ is zero for $n>1$. Finally, we obtain an interesting formula for C_2:

$$C_2 = \prod_{n\geq 2} [\zeta_{\geq 3}(n)]^{-I_n} = \prod_{n\geq 2} \left[\zeta(n)\left(1-2^{-n}\right)\right]^{-I_n}, \tag{6.6}$$

where

$$I_n = \frac{1}{n} \sum_{d|n} \mu(d) 2^{n/d}. \tag{6.7}$$

To see whether the convergence of this formula is improved we first must analyze I_n. Quite surprisingly, I_n is the number of irreducible monic polynomials of degree n with coefficients from GF(2). In fact, it is easy to prove even a more general theorem, originally due to Gauss. The proof below comes from [9, pages 49–50] and shows a remarkable resemblance to the arguments above, justifying the presence of I_n in (6.6).

Theorem 38 (On polynomials with coefficients from $\mathrm{GF}(p)$**).** *Let I_n be the number of irreducible monic polynomials from* $\mathrm{GF}(p)[x]$ *(p - prime) of degree n. Then*

$$I_n = \frac{1}{n}\sum_{d|n} \mu(d) p^{n/d}. \tag{6.8}$$

Proof. There are p^n monic polynomials of degree n. Let's write a power series with a formal parameter T for these numbers:

$$\sum_f T^{\deg f} = \sum_{n\geq 0} p^n T^n = \frac{1}{1-pT},$$

where f runs over all monic polynomials.

On the other hand, $\mathrm{GF}(p)[x]$ is a Euclidean domain, so every monic polynomial has a unique representation as a product of monic irreducible polynomials. Thus we can write an Euler product for the above series as

$$\frac{1}{1-pT} = \prod_v \left(1 - T^{\deg v}\right)^{-1} = \prod_{m\geq 1} \left(1 - T^m\right)^{-I_m},$$

where v runs over all irreducible monic polynomials.

If we take a logarithm on both sides of the above equation we will have

$$\log(1-pT)^{-1} = \sum_{n\geq 1} \frac{1}{n} p^n T^n$$

and

$$\begin{aligned}\log \prod_{m\geq 1} \left(1 - T^m\right)^{-I_m} &= \sum_{m\geq 1} -I_m \log\left(1 - T^m\right) = \\ &= \sum_{m\geq 1}\sum_{n\geq 1} \frac{I_m}{n} T^{nm} = \sum_{m\geq 1}\sum_{n\geq 1} \frac{mI_m}{nm} T^{nm} = \\ &= \sum_{n\geq 1}\sum_{d|n} \frac{dI_d}{n} T^n.\end{aligned}$$

If we equate the coefficients, then immediately

$$p^n = \sum_{d|n} dI_d.$$

To get the value of I_n and finish the proof, we invert the above identity and get

$$I_n = \frac{1}{n}\sum_{d|n} \mu(d) p^{n/d}.$$

This finishes the derivation. □

n	I_n	n	I_n
1	2	11	2046
2	2	12	4020
3	6	13	8190
4	12	14	16254
5	30	15	32730
6	54	16	65280
7	126	17	131070
8	240	18	261576
9	504	19	524286
10	990	20	1047540

Table 6.2: Values of I_n.

Using (6.7) it is easy to show an asymptotic formula, an analogue of prime number theorem. Namely

$$\left|\frac{1}{n}\sum_{d|n}\mu(d)2^{n/d}-\frac{2^n}{n}\right| = \left|\frac{1}{n}\sum_{\substack{d|n\\ d>1}}\mu(d)2^{n/d}\right| \leq \frac{1}{n}\sum_{\substack{d|n\\ d>1}}2^{n/d} = \frac{1}{n}\sum_{\substack{d|n\\ d\leq\frac{n}{2}}}2^{d} = O\left(\frac{2^{n/2}}{n}\right),$$

so

$$I_n = \frac{2^n}{n} + O\left(\frac{2^{n/2}}{n}\right). \tag{6.9}$$

From this we see that I_n roughly doubles with n (cf. Table 6.2), since

$$\frac{I_{n+1}}{I_n} = \frac{\frac{2^{n+1}}{n+1}+O\left(\frac{2^{(n+1)/2}}{n+1}\right)}{\frac{2^n}{n}+O\left(\frac{2^{n/2}}{n}\right)} = \frac{\frac{2}{n+1}+O\left(\frac{2^{(1-n)/2}}{n+1}\right)}{\frac{1}{n}+O\left(\frac{2^{-n/2}}{n}\right)} \xrightarrow[n\to\infty]{} 2.$$

The main term of $\zeta(n)\,(1-2^{-n})$ in (6.6) is 3^{-n}. The exponent $-I_n$ slows down the convergence by a factor of 2. Therefore the product converges like $\left(\frac{2}{3}\right)^n$ or gives about $\log_{10}\left(\frac{2}{3}\right) \approx 0.18$ decimal digits per term.

This is satisfactory, but we can do better than that. In (6.5) we can introduce an arbitrary prime number $q \geq 3$ to obtain

$$\begin{aligned}\log C_2 &= \sum_{m\geq 2}\frac{2-2^m}{m}\sum_{p\geq 3}p^{-m} = \sum_{m\geq 2}\frac{2-2^m}{m}\left(\sum_{p\geq q}p^{-m}-\sum_{3\leq p<q}p^{-m}\right) = \\ &= \sum_{m\geq 2}\frac{2-2^m}{m}P_{\geq q}(m) - \sum_{m\geq 2}\sum_{3\leq p<q}\frac{2-2^m}{mp^m}.\end{aligned}$$

Using the same method as before, the first term will be

$$\log\prod_{n\geq 2}\left[\zeta_{\geq q}(n)\right]^{-I_n}.$$

On the other hand, the second term is

$$
\begin{aligned}
-\sum_{m\geq 2}\sum_{3\leq p<q}\frac{2-2^m}{mp^m} = \sum_{3\leq p<q}\sum_{m\geq 1}\frac{2^m-2}{mp^m} &= \\
&= \sum_{3\leq p<q} 2\log\left(1-\frac{1}{p}\right) - \sum_{3\leq p<q}\log\left(1-\frac{2}{p}\right) = \\
&= \sum_{3\leq p<q}\log\left(1-\frac{1}{(p-1)^2}\right) = \log\prod_{3\leq p<q}\left(1-\frac{1}{(p-1)^2}\right).
\end{aligned}
$$

This gives us the following:

$$
C_2 = \prod_{3\leq p<q}\left(1-\frac{1}{(p-1)^2}\right)\prod_{n\geq 1}\left[\zeta_{\geq q}(n)\right]^{-I_n}, \tag{6.10}
$$

or even more concisely:

$$
\prod_{p\geq q}\left(1-\frac{1}{(p-1)^2}\right) = \prod_{n\geq 1}\left[\zeta_{\geq q}(n)\right]^{-I_n},
$$

which is a formula for a value of a “tail” in the product defining C_2.

By the same analysis as before, the convergence of (6.10) is of order $\frac{2}{q}$. For example, if we take $q = 23$, then

$$
\begin{aligned}
C_2 = \frac{1836515055375}{2751882854400}\prod_{n\geq 1}&[\zeta(n)\left(1-2^{-s}\right)\left(1-3^{-s}\right)\left(1-5^{-s}\right)\left(1-7^{-s}\right) \\
&\left(1-11^{-s}\right)\left(1-13^{-s}\right)\left(1-17^{-s}\right)\left(1-19^{-s}\right)]^{-I_n},
\end{aligned} \tag{6.11}
$$

which produces more than one decimal digit per each term.

The *mpmath* library already has an implementation of this algorithm using $q = 11$. We use formula with $q = 59$ to compute C_2 to a high precision. With this parameter we observed a noticeable acceleration of calculation. This can be seen in Figure 6.1. The new implementation runs almost twice as fast as the old one, but nevertheless the time required is exponential with respect to the required precision.

The twin prime constant C_2 was computed to 15000 decimal places. It took 210 hours (almost 9 days) to perform this computation.

Constant value

$$
\begin{aligned}
C_2 = 0.&6601618158468695739278121100145557784326233602847334133194484233354056423044 \\
&952771437600314138398679117790052266933040029658477551233662277471657132 13 \\
&986968741097620630214153735434853131596097803669932135255299767199302474590 \\
&593101082978291553834469297505205916657133653611991532464281301172462306379 \\
&341060056466676584434063501649322723528968010934966475600478812357962789459 \\
&842433655749375581854814173628678098705969498703841243363386589311969079150 \\
&040573717814371081810615401233104810577794415613125444598860988997585328984 \\
&038108718035525261719887112136382808782349722374224097142697441764455225265 \\
&548994829771790977784043757891956590649994567062907828608828395990394287082 \\
&529070521554595671723599449769037800675978761690802426600295711092099633708 \\
&272559284672129858001148697941855401824639887493941711828528382365997050328 \\
&725708087980662201068630474305201992394282014311102297265141514194258422242
\end{aligned}
$$

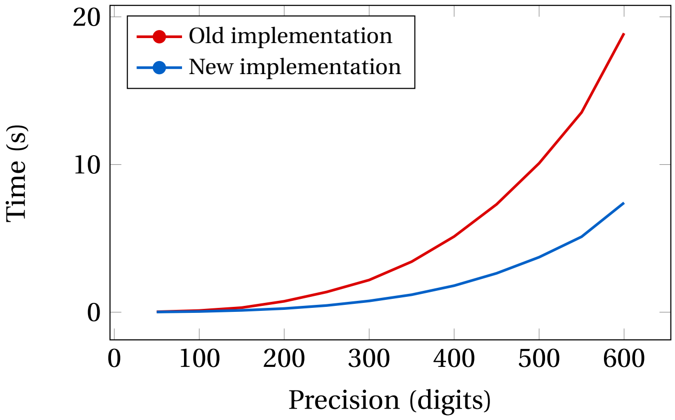

Figure 6.1: Comparison of the two implementations to compute the twin prime constant C_2. The new one runs almost two times faster.

```
375342296879836738796224286600285358098482833679152235700192585875285961205
994728621007171131607980571923988095560228270488107816765934282355475496309
266825229003591924631717409983012936531429813259810097377748558297744909551
479925068704626219528880108327386502657532264815237703850519708981859957858
102930103360383840786300680226342387945883514752520423084151976921757294137
976302727886083571553421342447340900744924989792428713014151542470640538846
819669255393305383337675341486088187109672998796506019368339988582819107010
139986258913749399696455959940669511059347457946288941314048415837718601561
511991160543925262584828223510322074268176656529243640839449903887458328691
325381818695303531525918917794833001207188010738575528896587449009039110194
984711174161475536478121096630049540953037469214431518138879622464609817971
063319157425900544807791177914563624369625019838249339038198884298030963857
792046864034702620737268681424992264085056125700592803722826653495328150202
675152135799091428248686302435481244448569763017448648766901295141649572521
923264561205245805264352158094845568557036029247430939429502731560139282997
364635235946336023366292461154544071634825853545013885887986311104493175989
942987639920400068052981695551896662139795794093718622171189160670972887760
785873933028636258405215686921094336264347389320817912725550057264982392433
726044195398506695724553828662017606916260929375951140690636190477119654881
631689313413370485036757426117439822722686884298347046912120174071588271617
950531782335442713415348445102220870438179760668395537024416722186976713202
755352818996091847759730724239432121368612650284219814551963243384601955926
909677877389306546159280704303825412314262425589531427134127742206223929036
154350634767054033350055990971228337187307076449070588256688416443243896602
425424674130561160058380001598892283985329418976711349937346633239339879708
180186675000350279072194449473465466508576810703544143682411062455559413387
888332246046283737092052506936531139406272785753221798836639400413857927012
313782357305681219145864623063531550453320785368134943078304409216484822247
769379264523164200937124344762755462254616718641907997486601976566131973326
363511210160035429079643004405510175585116924587038982978591442972806154708
408665773429545704852358860773177975944376892082598244502861992996242738693
689885998853872521565198543335024965136141609291023282300497224344360960160
670813651144905509166668263771279165786220586524807159002355580203663001120
986500275474656136044396280105595466063855032830343646387906521549265865539
025901693633231962094845661464509890681217341234677215365380244861004521102
878513796469666257823761843719534578409400242656093051354821135716499943566
313547169981928479003542042748249422696243116955091888870749842519990799292
763973120461025640877181163825531370585142016456201046976158966784865426508
655517325425692388304669815371829170165845787388955603886611072111012453527
160632610054606323139700221041847537678417637958364499876034086414602633789
733441668599037177704645035374356034708996482010754880929837557685746425644
727150601516676972591173460575270901033610926479884288553661186622139955693
340376914161474252105555328760598966822810148276784197461029924402411262000
270579503391447643548112469135104563591913837321765311511971878389051081379
604784632227364457461678685768056950844991389175978436482013670304642505936
978153776065450426299060132139222324088012645351757417875350793998161841468
129391797756853152810174314408491463413663350533827251141889757066198252277
157475248712364571833236796667280301007284702711012721258417999415109568060
079617710245533410948931234846061101015964774108932218054830567637161161996
184775961043639469118342627514615985728236339739735306175286594023032743399
348095611916047435400902808666360859874985835439282723498951415942509194954
```

```
605646534588125713580522553267437868292924825682312717909808480529148162044
```

```
437148107086352032291049160448086180867382271177641442764640594448388391220
823524918095389035859451283219730774824531484200273652962462957688579595887
439089361956539399338483027657676799803575494062529429556561580077971901109
954172599509032028149959527877484664068442532277739512520042634334542681825
907361022508495326455871223149924785813792372904065433388934975465929827566
977233189667000818132130221665320214445220032713125007042184391953754518969
527146643508604987165400651334903563755676083588147200415609265243812633877
993218366859013647137133949268421799203775924251259279783740011346590282844
959888226069725067385019359510293346365593648883053366508750028332281991514
159918758071131595716567969249412500068674428741337069203280263355480403915
932630451082270624385705753452296936797717696763816811924760617690690340045
474181619644067614425339610454768457794938585751054543027416255007403812762
118532657666305463428144352133399649510083229793075567600291389640713810451
427382290030518692538452915351925951993085959330585550814250307843224368690
013823201496760173155900292184023151317223562008412155153839578321943229957
941065768120373285945021075491645868499368293947809828114078527322209061309
312276787421994051170051337307751443444259467578669699101718856413584653100
423813884432189737960006953331294678522982404523408504984374325700734851591
664268332460114112759527390523951564445917376819787510704819557076817737283
907057151020602607595134583874342216837997375177929839544814982847872704878
136128480692207844493795757129155858877039702657222426779442073538735351551
449466666385385504615364105173309352671644174370754540301968163294804592484
145343078371036045325149483280624239558828189572040237689353573186415969557
738388366438919360692888649601881756214335954458511321433232726450255014935
078873851090708140622131516423557758541587095923903745777135970184857002960
497190897423511773557333572826976083748307747761095437071173785846712486678
699924554693817204603556241279118272415316814241019742518575837998933764284
817658519327933815319849964018732437747788307556784092203153322549618710377
523712857427488363638087422841606119810755239082237925909619407034020695933
237047617370980350415214440154403455073993060797014171020660552906195521679
662045789334990918696265399555036718524145141532182907338960289153632949573
072965077378726693774912781850450858703440653096741053296895599456012314780
451513867234465258139651446445220098806377996520260274121454248647607057095
342703619301901062974666239146127884361627373482094519818296456333568004257
271898392424690983136498422732113049280751322146061737598660744452597191150
064133689193098543582205089991843065581789948223130626852919268562596914259
063720392487545484642009450577776706753849826320638144326308455305374564451
438445160104884626056413987655766366900989762612989978462991671822671734148
656499458623300491502548558603624854348369534324987961315778773614917968310
734044819196648568200636203092733974337856288902269181469278859282476179524
511445711388232474789051473439084108904095650630978018661862275419914010590
690882639095114697037984433009652136195327293388151063584231477597319201919
705050861594534281183430168005334140640038459217490120076555556781899671680
659512760638674135011967905424457239442200019836352004370200288446974244838
740739746586274066398389252361750621305592951173541652703502582020075813230
600484403823868232618187371802900286931647467995542259657432161429133416640
503683826677856747169738797047143506821212724326239761899993802226014864896
703723306027240458370671177952906829748986231184455173651652157529933675764
068487667559079911936198340028068811720443565683726947732608538063171477983
027323515955278237117790480772424988060695516315953276042237958363908690307
517960560299066230697818006756241374504932012792259826508690430366972266728
980909961931776668001821262806937488584576864539953164277147953477966905012
484731895002207213730497641354965201232878228622327273221991142575636736816
041878338228202948501269806836946160054467589294499872215545933922041844875
629685027461706445173809456890496562786955404627388317339930056974892165279
```

6.3 Computation of Brun's constant

Analysis

We don't know a formula for Brun's constant apart from the very series that define it, that is,

$$\sum_{\substack{p,p+2\\\text{primes}}} \left(\frac{1}{p} + \frac{1}{p+2}\right).$$

This sum converges very slowly and we cannot be sure even about the first places after the decimal point. The contribution, even from the terms with large primes, is too huge.

We generated pairs of twin prime numbers up to 10^{10} and computed their inverses. The computation was done using 1000 decimal places of working precision. It took few hours to get the result.

Pentium FDIV bug

In 1994, Thomas Nicely has found a bug in a model of Pentium processor while computing the Brun's constant. The bug called "Pentium FDIV bug" became apparent when Nicely found some inconsistencies in his computations. The problem was indeed very rare, but possible. It manifested itself when specific pairs of floating pairs of numbers very divided by each other. For example when diving 4195835 by 3145727, one obtains

$$1.333820449136241.$$

However, the flawed processor would return

$$1.333739068902037,$$

which is incorrect at the fourth digit after the decimal point. Intel, the manufacturer of the processor, was reluctant to resolve the issue, but finally decided to launch a total recall of the flawed processors. See [32] for the whole story and more information.

Constant value

$$\begin{aligned}
B_2 = {} & \texttt{1.78747850271924154746273348811292230518634708287049027644388 6}\\
& \texttt{78513430462794470500693498335796250458088979186433470895533 5}\\
& \texttt{79976320959546383694561409684927526396418934512951601864596 5}\\
& \texttt{21288094621967963366054966141564347163766326108175503201320 6}\\
& \texttt{79895496907483271147805035968642759226716229822503264530925 5}\\
& \texttt{01839853272976646824681703161321418973300535542492681450229 4}\\
& \texttt{86291824083081325695612750536506916604722389266662752878792 8}\\
& \texttt{48391454435150052153840107134696898386335011221464218294869 4}\\
& \texttt{34093668451706156870450317557649353639687277447292954868486 0}\\
& \texttt{02110010883504083627868862364024015594803362874650047404059 7}\\
& \texttt{92427438803579073726663394884878407755345578476997246305324 9}\\
& \texttt{32907746828373939203292692828485838424610983633082147724445 0}\\
& \texttt{65081960329345985854902153721424774906154695014051828508534 5}\\
& \texttt{25678743629665003478293863101268525290610765784674425112630 2}\\
& \texttt{64730722482439616947534851704098211055842925956268315294894 5}\\
& \texttt{06271338900625893158330025562840738135567593809507691463668 4}
\end{aligned}$$

6.4 Summary

We introduced the problem of computing constants related to the Twin Prime Conjecture. The twin prime constant has a convenient series that can be used to compute it to a desired precision, yet the computation still takes the exponential time to finish.

A much more complicated problem is to compute Brun's constant B_2. To the author's knowledge, there are no known formulas other than the definition itself.

Chapter 7

Summary

In this thesis we presented the Twin Prime Conjecture and prime sieving algorithms. We proved some theorems on the twin primes, their characterization and distribution, and relation to other, mostly also unsolved, problems in the number theory. In particular, we showed Brun's theorem that states that the sum of reciprocals of twin primes converges. Moreover, we contrasted the current knowledge about this conjecture with state-of-the-art algorithms for prime sieving. Finally, we performed analysis of approaches used to compute related numerical constants and time consuming computations thereafter. As a result, we obtained the value of the twin prime constant to a very high precision.

For the prime sieving algorithms, the question remains open if there exists an algorithm that can achieve the theoretical bound on the running complexity of $O(n/\log n)$. There is no algorithm known that achieves that bound. If it is not possible to achieve this bound, a proof of this fact would be a step forward.

We also are not aware of a practical way to obtain value of Brun's constant to a high precision. The series that define the constant are very slowly convergent and there is no obvious way to accelerate the convergence as it is possible wit the twin prime constant.

The Twin Prime Conjecture remains unproven, but definitely there are serious attempts to prove it. It's not clear, however, if the problem can be ultimately resolved with the help of sieve methods. The deep and complicated work of Chen resulted in a near miss attack on the conjecture, but nothing substantially important was proven since then in the domain of twin primes. There is however serious interest in the research about small prime gaps, a subject that may result in the result related to the Twin Prime Conjecture. For example, it has been conditionally proven (assuming Elliott-Halberstam Conjecture) that there exist infinitely many primes whose difference is 16 or less ([22]).

There has been a lot of successful research on the Twin Prime Conjecture and empirical evidence confirms it, but the main question remains open:

Are there infinitely many twin primes?

List of Theorems

Bibliography

[1] M. Aigner, G. M. Ziegler, and K. H. Hofmann. *Proofs from the book*. Springer, 2009.

[2] T. M. Apostol. *Introduction to Analytic Number Theory*. Number v. 1 in Undergraduate Texts in Mathematics. Springer-Verlag, 1976.

[3] A. O. L. Atkin and D. J. Bernstein. Prime Sieves Using Binary Quadratic Forms. *Mathematics of Computation*, 73:2004, 1999.

[4] D. J. Bernstein. primegen. `http://cr.yp.to/primegen.html`.

[5] V. Brun. Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare. *Archiv for Math. og Naturvid.*, 34(8), 1915.

[6] C. K. Caldwell. The Top Twenty: Twin Primes. `http://primes.utm.edu/top20/page.php?id=1#records`.

[7] J. Chen. On the Representation of a Large Even Integer as the Sum of a Prime and the Product of at Most Two Primes. *II. Sci. Sinica*, 21(4):421–30, 1978.

[8] P. A. Clement. Congruences for Sets of Primes. *The American Mathematical Monthly*, 56(1):23–25, 1949.

[9] A. C. Cojocaru and M. R. Murty. *An Introduction to Sieve Methods and Their Applications*. Cambridge University Press, Cambridge, 2005.

[10] T. H. Cormen. *Introduction to Algorithms*. MIT Electrical Engineering and Computer Science. MIT Press, 2001.

[11] R. Crandall and C. Pomerance. *Prime Numbers – A Computational Perspective*. Springer, New York, second edition, 2005.

[12] L. E. Dickson. A new extension of Dirichlet's theorem on prime numbers. *The Messenger of Mathematics*, 33, 1903.

[13] B. Dunten, J. Jones, and J. Sorenson. A Space-Efficient Fast Prime Number Sieve. In *Information Processing Letters 59*, pages 79–84, 1996.

[14] T. O. e Silva. Tables of values of $\pi(x)$ and of $\pi_2(x)$. `http://www.ieeta.pt/~tos/primes.html`.

[15] C. Elsholtz. Kombinatorische Beweise des Zweiquadratesatzes und Verallgemeinerungen. *Mathematische Semesterberichte*, 50:77–93, 2003. 10.1007/s00591-003-0060-3.

[16] P. Erdős. Über die Reihe $\sum \frac{1}{p}$. *Mathematica, Zutphen*, B 7:1–2, 1938.

[17] P. Flajolet and I. Vardi. Zeta Function Expansions of Classical Constants. unpublished manuscript, `http://algo.inria.fr/flajolet/Publications/landau.ps`, 1996.

[18] J. Friedlander and H. Iwaniec. Using a Parity-Sensitive Sieve to Count Prime Values of a Polynomial. *Proc Natl Acad Sci U S A*, 94(4):1054, 1997.

[19] W. F. Galway. Dissecting a Sieve to Cut Its Need for Space. In *Proceedings of the 4th International Symposium on Algorithmic Number Theory*, pages 297–312, London, UK, 2000. Springer-Verlag.

[20] *The GNU Multiple Precision Arithmetic Library (version 4.3.2)*, October 2011. `http://gmplib.org/`.

[21] D. A. Goldston, Y. Motohashi, J. Pintz, and C. Y. Yıldırım. Small Gaps between Primes Exist, May 2005.

[22] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in Tuples I. *Annals of Mathematics*, 170(2):819–862, 2009.

[23] S. W. Golomb. The Twin Prime Constant. *The American Mathematical Monthly*, 67(8):767–769, 1960.

[24] G. Greaves. *Sieves in number theory*. Number v. 43 in Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, 2001.

[25] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions, February 2006.

[26] Julian Havil. *Gamma : exploring Euler's constant*. Princeton University Press, March 2003.

[27] D. R. Heath-Brown. Primes represented by $x^3 + 2y^3$. *Acta Mathematica*, 186:1–84, 2001. 10.1007/BF02392715.

[28] F. Johansson et al. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.14)*, February 2010. `http://code.google.com/p/mpmath/`.

[29] W. G. Leavitt and A. A. Mullin. Primes Differing by a Fixed Integer. *Mathematics of Computation*, 37(156):581–585, 1981.

[30] H. Lee and Y. Park. The Generalization of Clement's Theorem on Pairs of Primes. *Journal of Applied Mathematics & Informatics*, 27(1–2):89–96, 2009.

[31] D. J. Newman. Simple Analytic Proof of the Prime Number Theorem. *The American Mathematical Monthly*, 87(9):693–696, 1980.

[32] T. R. Nicely. Pentium FDIV Flaw. `http://www.trnicely.net/#PENT`.

[33] C. S. Ogilvy and J. T. Anderson. *Excursions in number theory.* Dover Books Explaining Science Series. Dover Publications, 1988.

[34] P. Pritchard. A Sublinear Additive Sieve for Finding Prime Numbers. *Commun. ACM*, 24:18–23, January 1981.

[35] L. Schnirelmann. Über additive Eigenschaften von Zahlen. *Mathematische Annalen*, 107:649–690, 1933. 10.1007/BF01448914.

[36] I.S.A. Sergusov. On the problem of prime twins. *Jaroslav. Gos. Ped. Inst. Ucen. Zap.*, 82:85–86, 1971.

[37] J. Sorenson. Trading Time for Space in Prime Number Sieves. In *Proceedings of the Third International Algorithmic Number Theory Symposium (ANTS III*, pages 179–195, 1998.

[38] J. Sorenson. The Pseudosquares Prime Sieve. In *ANTS*, pages 193–207, 2006.

[39] W. A. Stein et al. *Sage Mathematics Software (Version 4.7.2).* The Sage Development Team, 2011. `http://www.sagemath.org`.

[40] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1995.

[41] S. Y. Yan. *Number theory for computing.* Springer, 2002.

[42] D. Zagier. A One-Sentence Proof That Every Prime $p \equiv 1 \pmod 4$ Is a Sum of Two Squares. *The American Mathematical Monthly*, 97:144, February 1990.

Adam Mickiewicz University in Poznań
Faculty of Mathematics and Computer Science

Typeset using LaTeX.

BibTeX:

```
@mastersthesis{ key,
    author = "Tomasz Buchert",
    title = "{On the twin prime conjecture}",
    school = "Adam Mickiewicz University",
    address = "Pozna{\'n}, Poland",
    year = "2011",
}
```