Prime number sieves

Utrecht University

Guido Reitsma

Supervisor: Lasse Grimmelt

January 17th 2020

Acknowledgements

First I want to thank Lasse for being a great supervisor, especially given that it was his first bachelor's thesis that he supervised and I'm definitely not the easiest person. Secondly I want to thank all my friends, family and other people, who supported me and listened to my thoughts for the last three months and before.

Contents

Chapter 1

Introduction

In this thesis there are two applications from the Sieve of Eratosthenes The two sieves that are mentioned in this thesis are the combinatorial sieve and the General Number field Sieve. We will then compare both of the methods which each other and look how these methods coincide or differ. The first part of this introduction is about the sieve of Eratosthenes. The Sieve of Eratosthenes is first discussed in a historical manner. Secondly there will be a notion of the Sieve of Eratosthenes in a more Number Theoretical sense. Then there is an estimation of the number of primes of which the foundation have been laid by Legendre. The last part of the introduction is about the RSA, one of the main reasons the General Number Field Sieve is usefull in cryptography.

1.1 Sieve of Eratosthenes Historically

One of the first notions of a sieve, is the Sieve of Eratosthenes in the third century B.C.E. He made a list of odd numbers and decided to delete 3^2 and all the third numbers after that. After that he did the same with 5^2 and the fifth number after that. Then with 7 and so on following the primes. This process is made visual in Table (1.1)

1	3	5	7	~~9~~^{3}	11	13	~~15~~^{3}	17	19
~~21~~^{3}	23	~~25~~^{5}	~~27~~^{3}	29	31	~~33~~^{3}	~~35~~^{5}	37	~~39~~^{3}
41	43	~~45~~^{3,5}	47	~~49~~^{7}	~~51~~^{3}	53	~~55~~^{5}	~~57~~^{3}	59
61	~~63~~^{3,7}	~~65~~^{5}	67	~~69~~^{3}	71	73	~~75~~^{3,5}	~~77~~^{7}	79

Table 1.1: Example Sieve of Eratosthenes

[4] In the thirteenth century, mathematicians came to the conclusion that to get the primes till a given number X we have to do this process for all numbers until $\sqrt{X}$. The reason for this is that every number X with a prime factorization bigger than 2 has a prime factor smaller than $\sqrt{X}$

1.2 The way many people know the Sieve of Eratosthenes

An example of the Sieve of Eratosthenes that is often learned to high school students or beginning students can be found in Table 1.2. This is a process to find all the primes until one hundred. This is done by eliminating all the numbers with prime divisors till $\sqrt{100} = 10$. Those are: 2, 3, 5 and 7. First we eliminate every multiple of 2 (green) then

all the multiples of 3 not yet deleted (red), 5(blue) and 7(Yellow). Ultimately we can conclude that the white numbers till 100 are the primes.

	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
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Table 1.2: Example Sieve of Eratosthenes

1.3 Estimating primes via Eratosthenes' Sieve

After making a Sieve of Eratosthenes, we may ask ourselves how many primes there are till a number X. Additionally we want to know if this stays doable if X gets really big. For this we need a counting function $\pi(X) = |\{\text{primes } p : p \leq X\}|$ so, for example: $\pi(100) = 25$, as seen in Table 1.2.

1	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
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Table 1.3: Example Sieve of Eratosthenes

To get a functional formula for the number of primes below X, it is logical to ask how many multiples of each prime there are. Let for example look at $X = 100$, see also Table 1.3. Maybe it is possible to do something with multiple of primes. First we start with 100, and try to subtract every number till 100, which have a multiple of at least one of the primes. To find the number of numbers which have a prime divisor, first look at the multiples of the primes below $\sqrt{100}$. These are $\{2, 3, 5, 7\}$. The number of multiples of 2 from 1 to 100 are $\lfloor \frac{100}{2} \rfloor = 50$, the number of multiples of 3 from 1 to 100 are $\lfloor \frac{100}{3} \rfloor = 33$. And so, with 5, this results in 20 and with 7 we find 14 multiples. These are all coloured numbers in table 1.3. The problem here is, that there are numbers in this set, which have multiple prime factors of the set $\{2, 3, 5, 7\}$. For example the multiples of $6 = 2 \cdot 3$, will give back something twice, because those numbers have divisors 2 and 3. So secondly it is needed to add all the numbers which have 2 prime divisors of the set $\{2, 3, 5, 7\}$. So the set: $\{6, 10, 14, 15, 21, 35\}$. These are yellow and red in table 1.3. But then there is a problem with all numbers until 100, which have at least three prime divisors, which are

counted one to many time in this set. So the third step is to subtract the multiples of all the numbers which have at least three prime divisors of $\{2, 3, 5, 7\}$ until 100 hence all multiples of $\{30, 42, 70\}$ For a bigger X, see the red numbers in 1.3. This process, would have gone on with more multiples, though $2 \cdot 3 \cdot 5 \cdot 7 > 100$, so there is no need to go further for $X = 100$. But it is possible to deduct a formula for the number of primes with multiples of numbers: The Inclusion-Exclusion formula:

Lemma 1.3.1 (Inclusion-Exclusion Formula)**.** let $z = \sqrt{X}$ and p_i distinct primes lower than z. The number of primes below X is:

$$\pi(X) = X - 1 + \pi(z) - \sum_{p_1 \leq z} \left\lfloor \frac{X}{p_1} \right\rfloor + \sum_{p_1 < p_2 \leq z} \left\lfloor \frac{X}{p_1 p_2} \right\rfloor - \sum_{p_1 < p_2 < p_3 \leq z} \left\lfloor \frac{X}{p_1 p_2 p_3} \right\rfloor + \cdots \quad (1.1)$$

From this we want to define sieves in a more analytic manner. But before we do this, we need some definitions.
Let z be an arbitrary number in $\mathbb{R}$ and p a prime. The number $P(z)$ is defined as the product of the primes in $(-\infty, z)$:

$$P(z) := \prod_{p<z} p$$

$\mathcal{A}$ is the set in which we search for primes, most often this set will be used:

$$\mathcal{A} = \{n \in \mathbb{N} : 1 \leq n \leq X\} \quad (1.2)$$

Now it is useful to look at sieves in terms of sets $\mathcal{A}$ and primes up to z. So let (a, b) the greatest common divisor of a and b. Then:

$$S(\mathcal{A}, z) = |n \in \mathcal{A} : (n, P(z)) = 1|$$

This leads to an inclusion exclusion-formula for only the sieve $S(\mathcal{A}, z)$

$$S(\mathcal{A}, z) = X - \sum_{p_1 < z} \left\lfloor \frac{X}{p_1} \right\rfloor + \sum_{p_1 < p_2 < z} \left\lfloor \frac{X}{p_1 p_2} \right\rfloor - \sum_{p_1 < p_2 < p_3 < z} \left\lfloor \frac{X}{p_1 p_2 p_3} \right\rfloor + \cdots \quad (1.3)$$

Just the Inclusion-exclusion formula without 1 and $\pi(z)$ for $\mathcal{A}$ as in 1.2 Hence:

$$\pi(X) = S(\mathcal{A}, \sqrt{X}) - 1 + \pi(\sqrt{X}) \quad (1.4)$$

To make equation 1.3 a bit shorter, we make use of the Möbius function

Definition 1.3.1 (Möbius function)**.** The Möbius function $\mu(d)$ is defined as:

$$\mu(d) = \begin{cases} 1 & \text{if } d = 1 \\ (-1)^r & \text{if } d = p_1 p_2 \dots p_r \text{ for distinct primes } p_i \\ 0 & \text{if } p^2 | d \text{ for some prime } p \end{cases} \quad (1.5)$$

We can see that the Möbius function is a multiplicative function, which means that $\mu(1) = 1$ and for two coprime numbers p and q we have $\mu(pq) = \mu(p)\mu(q)$. The Möbius function is in the multiplicative way defined by the prime-powers p^a with

$$\mu(p) = -1 \text{ and } \mu(p^a) = 0 \text{ if } a \geq 2 \quad (1.6)$$

The Möbius function gives rise to a simpler equation for $S(\mathcal{A}, z)$ That combines equations 1.3 and of course 1.5 to this summation. The Möbius function will give the signs, we can

see at 1.5, so if there are an odd number of primes, it will be -1 and an even number of primes gives $+1$. With $\mathcal{A}$ as in 1.2, this gives us:

$$S(\mathcal{A}, z) = \sum_{d|P(z)} \mu(d) \left\lfloor \frac{X}{d} \right\rfloor \tag{1.7}$$

Now we want to find an equation or formula to get a better understanding of $S(\mathcal{A}, z)$ in this case.
For any number $X \in \mathbb{R}^+$ We can see that:

$$\lfloor X \rfloor = X + \{X\} = X + O(1) \tag{1.8}$$

Because $\{X\} \in [0, 1)$.
We can now estimate an upper bound with this formula, with 1.8:

$$S(\mathcal{A}, z) = X \sum_{d|P(z)} \frac{\mu(d)}{d} + \sum_{d|P(z)} O(1) \tag{1.9}$$

We can change this equation. since $\mu(d) = 0$ for numbers that have a square prime divisor: $p^2|d$. We can see that:

$$\sum_{d|P(z)} \frac{\mu(d)}{d} = 1 - \sum_{p_1 \leq z} \frac{1}{p_1} + \sum_{p_1 < p_2 \leq z} \frac{1}{p_1 p_2} - \cdots = \prod_{p \leq z} (1 - \frac{1}{p}) \tag{1.10}$$

For the right side of the $+$-sign in 1.9 we need the number of d that divide $P(z)$. We know that $P(z)$ is a multiplication of all prime numbers till z. Every d has a prime divisor either in it, or not, which gives $2^{\pi(z)}$possibilities. So

$$\sum_{d|P(z)} O(1) = O(2^{\pi(z)})$$

Hence because the number of numbers that divide $P(z)$

$$S(\mathcal{A}, z) = X \prod_{p \leq z} (1 - \frac{1}{p}) + O(2^{\pi(z)}) \tag{1.11}$$

With Mertens' theorem, if we only stick to the main term of $S(\mathcal{A}, z)$ we can see this gives us

$$X \prod_{p \leq z} (1 - \frac{1}{p}) = \frac{X \cdot e^{-\gamma}}{\log z} (1 + O(\frac{1}{\log z}))$$

While the Prime Number Theorem states:

$$\pi(X) \sim \frac{X}{\log(X)}$$

Assume now we are looking for primes, than using 1.4 there is a factor difference of(if we stick to the main term again):

$$\frac{X \cdot e^{-\gamma}}{\log(\sqrt{X})} / \frac{X}{\log(X)}$$

to finally get:

$$\frac{X \cdot e^{-\gamma}}{\log(\sqrt{X})} / \frac{X}{\log(X)} = \frac{\log(X)}{\log(\sqrt{X})} \cdot e^{-\gamma} = 2 \cdot e^{-\gamma} \approx 1,12$$

But also the second term has problems. The term $2^{\pi(z)}$ gets very large if z gets bigger. For example, for $z = 1000$: $\pi(z) = 168$ and $2^{168} \approx 10^{50} >> 1000000$ This gives notion to try to make a better estimation of the quantity of prime numbers.

About this, more will be told in Chapter 2, where we are going to talk about brun's sieve. A number theoretical way of sifting primes.

1.4 RSA

In this section the RSA(Rivest, Shamir, Adlemann) algorithm is explained. This algorithm gives rise to the importance of the General Number Field Sieve(GNFS) and is an important alogrithm for the security of data. In this part the RSA algorithm will be explained with example and some notes. Used for this section is mainly a website of David Ireland about the RSA algorithm [6].

Steps	Easy example with small primes	Notes of importance
1. Choose two large primes p and q and calculate $n = pq$	In this example we take $p = 11$ and $q = 13$. Now $n = 11 \cdot 13 = 143$	In the real cryptology case we would take large primes, often of 256, 512 or 1024 bits. Or in the range of 2^{256}, 2^{512}, 2^{1024}. In these cases calculating p and q from n will take a lot of time in most cases since, we cant go trough all primes of the form 2^{256} since that is a number bigger than the number of atoms in the universe.
2. calculate $\phi = (p-1)(q-1)$	$\phi = (p-1)(q-1) = 10 \cdot 12 = 120$	
3. Find a number e such that $1 < e < \phi$ abd $\gcd(e, \phi) = 1$	Here we choose $e = 17$, since $e = 3$ and $e = 5$ are factors of 120 $\gcd(120, 17) = 1$	e is on real algorithms often, a fermat prime. A prime of the form: $2^{2^n} + 1$ Hence the most commonly used and also first 5 fermat primes are: $3, 5, 17, 257, 65537$ These numbers will work if $p, q \not\equiv 1 \mod e$
4. find d such that $ed \equiv 1 \mod \phi$	$17 \cdot 113 = 1921 \equiv 1 \mod 120$, thus $d = 113$	Here we can use the extended Euclid Algorithm to find d
5. The public key is (n, e), the private key is (n, d)	$(n, e) = (143, 17)$ and $(n, d) = (143, 113)$	just like p and q, d should be private.
6. When the message that you want to encrypt is m: calculate the cyphertext by $c = m^e \mod n$	Let $m = 7$ $c = 7^{17} \mod 143 = 7^{16} \cdot 7^1 \mod 143 = 48 \cdot 7 \mod 143 = 50 \mod 143$	To calculate 'big' powers, we can use that $kl \mod n = k \mod n \cdot l \mod n$. Hence we can do a process where the exponent is put into binary and we can calculate the for example 7^{16} with $(((7^2)^2)^2)^2$
7. Find the message back by: $m' = c^d \mod n$	$m' = 50^{113} \mod 143 = 50^1 \mod 143 \cdot 50^{16} \mod 143 \cdot 50^{32} \mod 143 \cdot 50^{64} \mod 143 = (42 \cdot 113 \cdot 16 \cdot 50) \mod 143 = 7$	The same process of 6, can be used here also to calculate 'big' powers.

We can see that this algorithm shows that the factorization of a prime number is very important. For small primes(like 143) This is relatively easy. But for products of two primes with 1024 bits, this gets difficult. We can't go over all the primes of 1024 bits, cause the number of primes are going to $\frac{2^{1024}}{\log 2^{1024}}$ by the Prime Number Theorem. But we have better algorithms to solve this. The best known algorithm is named the General Number Field Sieve (or GNFS) and we will discuss this in chapter 3.

Chapter 2

Combinatorial Sieve

In this chapter is shown the beginning of a notion of the Brun's sieve. First there will be some new notations and the definition of the Buchstab Identity. After that we will give a notion of Brun's sieve. In this part the book of Opera de Cribro, chapter 6 [3] will be followed thoroughly. Also parts of the first and third chapter of [4] are used.

2.1 Notation

To understand the first part of the combinatorial sieve, we need some new notations and definitions. The first notation will be the set of all numbers divisible by a number d in a set $\mathcal{A}$.

Definition 2.1.1. let d be an integer:

$$\mathcal{A}_d = \{a \in \mathcal{A} : a \equiv 0 \bmod d\} \tag{2.1}$$

We can also write if $\mathcal{A} = n \in \mathbb{N} : n \leq x$

$$A_d(x) = |\mathcal{A}_d| \tag{2.2}$$

Since $\mathcal{A}_d$ has a clear structure, it is expected that it has an approximation of the form:

$$A_d(x) = g(d)A(x) + r_d(x) \tag{2.3}$$

Here $A(x) = A_1(x)$ and $r_d(x)$ is some small number relative to $g(d)A(x)$.
From this base, it is possible to look back upon the sieve of Chapter 1. When we look at 1.7, we can change this formula with 2.1 to:

$$S(\mathcal{A}, z) = \sum_{d|P(z)} \mu(d) \left\lfloor \frac{X}{d} \right\rfloor = \sum_{d|P(z)} \mu(d)|\mathcal{A}_d|$$

To get an estimation for the sieve explained in the introduction that fits the equation 2.3, it is logical to take $g(d) = \frac{1}{d}$, $A_d(X) = X$ and $r_d = O(1)$. And we find 1.9.

2.2 Sieve weights

As we have seen with the Sieve of Eratosthenes, there is a problem when we are estimating $S(\mathcal{A}, z)$. When the set of prime number grows really large the error grows even larger. We can write the error of the estimation in multiple ways now, in the case of the Sieve of Eratosthenes:

$$\sum_{d|P(z)} r_d(x) = \sum_{d|P(z)} \mu(d) r_d(x) = \sum_{d|P(z)} \mu(d) O(1) = \sum_{d|P(z)} O(1) = O(2^{\pi(z)}) \tag{2.4}$$

Since the error term gets enormous in this sieve. The question is, if it is possible to change the $\mu(d)$ into some $\lambda(d)$ for which $\lambda(d) = 0$, for $d > D$ for some D.
A logical solution would be:

$$\lambda_D(d) = \begin{cases} \mu(d) & \text{if } d < D \\ 0 & \text{if } d \geq D \end{cases} \tag{2.5}$$

The problem with this solution is that it is not easy to see for which D, this would become a lower or an upper bound, since in this case, we just start with following the primes and than some composite numbers, but you never know if you are lower or higher than the expected sieve value if you cut it of like this at a D.
From here on a beginning of Brun's sieve will be explained. The goal is to find some lower and upperbounds for $S(\mathcal{A}, z)$ namely:

$$S^-(\mathcal{A}, z) \leq S(\mathcal{A}, z) \leq S^+(\mathcal{A}, z) \tag{2.6}$$

Now $S^{\pm}(\mathcal{A}, z)$, needs some further formulation.
First the sift weights $\mu(d)$ will be changed to new weights $\lambda^{\pm}$. We also want to try to get a lower bound that has value as high as possible as well as an upper bound with a value as low as possible. We may write:

$$S^{\pm}(\mathcal{A}, z) = XV^{\pm} + R^{\pm} \tag{2.7}$$

The big question here is what weights $\lambda^{\pm}$, we will use for different elements of $\mathcal{A}$ Hence we will look at:

$$V^- = \sum_{d|P(z)} \lambda_d^- g(d) \text{ and } V^+ = \sum_{d|P(z)} \lambda_d^+ g(d) \tag{2.8}$$

and also for the rest term:

$$R^- = \sum_{d|P(z)} \lambda_d^- r_d \text{ and } R^+ = \sum_{d|P(z)} \lambda_d^+ r_d \tag{2.9}$$

Definition 2.2.1 (Combinatorial Sieve)**.** A combinatorial sieve is a sieve λ_d where the λ's have the same value as the Möbius function or 0

In the next chapters such a combinatorial sieve with some clear upper and lower bounds is found. But first there is a notion of Buchstab's Identity.

2.3 Buchstab's Identity

In this section the Buchstab's identity is explained. So first the definition of the Buchstab's identity.

Lemma 2.3.1 (Buchstab's identity)**.** It holds that:

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p|P(z)} S(\mathcal{A}_p, p) \tag{2.10}$$

For clarity of this identity we will again consider the special case of $\mathcal{A} = \{x \in \mathbb{N} : 1 \leq x \leq N\}$ and start to put in the first three primes, beginning with $p = 2$

$$S(\mathcal{A}_2, 2) = \lfloor \frac{1}{2} N \rfloor \tag{2.11}$$

The number of even numbers. Now secondly p=3:

$$S(\mathcal{A}_3, 3) \tag{2.12}$$

We will look at every number which is in $\mathcal{A}_3$ Hence $x \equiv 0 \mod 3$, but if they have a same prime factor as $P(z) = 2$ we will not do anything. And this holds for all numbers which are even. Hence we only eliminate the numbers that are 0 mod 3 but not 0 mod 2 Hence all the numbers that are 3 mod 6 Now thirdly p=5:

$$S(\mathcal{A}_5, 5) \tag{2.13}$$

We can see $P(5) = 3 \cdot 2 = 6$. Hence we only eliminate all numbers that are 0 mod 5 but not 0 mod 2 or 0 mod 3 Hence they must also be 1 mod 6 or 5 mod 6.
We will go on with this process until z hence we are left with exactly the number of elements in $\mathcal{A}$ minus once the number of elements which have at least 1 prime factor smaller than z. Which is exactly $S(\mathcal{A}, z)$ And this looks like the way we eliminated numbers in our first example at table 1.2, but there we didn't eliminate the primes p and we did eliminate 1.

2.4 Brun's sieve

In this part we will talk about the main example of combinatorial sieves. We will use the inclusion-exclusion principle to get to a lower and upper bound for $S(\mathcal{A}, z)$ This construction is called Brun's construction, named after Viggo Brun a Norwegian Mathematician. We will first make a construction for the upper bound of $S(\mathcal{A}, z)$. Starting with Buchstab's Identity:

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1 | p(z)} S(\mathcal{A}_{p_1}, p_1) \tag{2.14}$$

Where p_1 runs over the divisors of $P(z)$ To make an upper bound we may cut of all the big primes higher than some number y_1. We don't know what happens with the solution nor the error yet, but we will find an upper bound, hence:

$$S(\mathcal{A}, z) \leq |\mathcal{A}| - \sum_{p_1 < y_1} S(\mathcal{A}_{p_1}, p_1) \tag{2.15}$$

We can than use Buchstab's identity again to solve for $S(\mathcal{A}_{p_1}, p_1)$. We will use:

$$S(\mathcal{A}_{p_1}, p_1) = |\mathcal{A}_{p_1}| - \sum_{p_2 < p_1} S(\mathcal{A}_{p_1 p_2}, p_2) \tag{2.16}$$

To get to the new inequality:

$$S(\mathcal{A}, z) \leq |\mathcal{A}| - \sum_{p_1 < y_1} |\mathcal{A}_{p_1}| + \sum \sum_{p_2 < p_1 < y_1} S(\mathcal{A}_{p_1 p_2}, p_2) \tag{2.17}$$

There is now no possibility do erase some numbers for p_2 since doing that would decrease or limit, but after 3 iterations, this is again a possibility. We can write:

$$S(\mathcal{A}, z) \leq |\mathcal{A}| - \sum_{p_1 < y_1} |\mathcal{A}_{p_1}| + \sum \sum_{p_2 < p_1 < y_1} |\mathcal{A}_{p_1 p_2}| - \sum \sum \sum_{p_3 < p_2 < p_1 < y_1} S(\mathcal{A}_{p_1 p_2 p_3}, p_3) \tag{2.18}$$

Here we can give a new constrain by letting $p_3 < y_3$

$$S(\mathcal{A}, z) \leq |\mathcal{A}| - \sum_{p_1 < y_1} |\mathcal{A}_{p_1}| + \sum_{p_2 < p_1 < y_1} \sum |\mathcal{A}_{p_1 p_2}| - \sum_{\substack{p_3 < p_2 < p_1 < y_1 \\ p_3 < y_3}} \sum \sum \sum S(\mathcal{A}_{p_1 p_2 p_3}, p_3) \tag{2.19}$$

When we do this more times, this gives rise to a generalized upper bound.
We can also to a similar thing for the lower bound. We start again with Buchstab's Identity and than iterate it. Hence,

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1} S(\mathcal{A}_{p_1}, p_1) = |\mathcal{A}| - \sum_{p_1} |\mathcal{A}_{p_1}| + \sum_{p_1} \sum_{p_2 < p_1} S(\mathcal{A}_{p_1 p_2}, p_2) \tag{2.20}$$

To get a lower bound for $S(\mathcal{A}, z)$ here, we can give a constraint to p_2 such that there is less added on and get a lower bound.

$$S^-(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1} |\mathcal{A}_{p_1}| + \sum_{\substack{p_2 < p_1 \\ p_2 < y_2}} \sum S(\mathcal{A}_{p_1 p_2}, p_2) \tag{2.21}$$

Now we can continue this, just in a similar way as for S^+, so iterate the buchstab identity twice and than give an upper bound for even n in p_n
We will define a combinatorial sieve λ^+ and λ^- on sets $\mathcal{D}^+$ and $\mathcal{D}^-$ as

Definition 2.4.1. Let:

$$\mathcal{D}^+ = \{d = p_1 \cdots p_l : p_m < y_m \text{ for } m \text{ odd}\} \tag{2.22}$$

$$\mathcal{D}^- = \{d = p_1 \cdots p_l : p_m < y_m \text{ for } m \text{ even}\} \tag{2.23}$$

where d is a product of decreasing distinct primes p_i and $1 \in \mathcal{D}^{\pm}$.
We can now see that continuing the process of 2.22 gives us an upper bound:

$$S^+(\mathcal{A}, z) = \sum_{d | P(z)} \lambda_d^+ S(\mathcal{A}, z) := \sum_{\substack{d | P(z) \\ d \in \mathcal{D}^+}} \mu(d) |\mathcal{A}_d| \geq S(\mathcal{A}, z) \tag{2.24}$$

and of course the continued process around 2.24 gives us a similar lower bound:

$$S^-(\mathcal{A}, z) = \sum_{d | P(z)} \lambda_d^- S(\mathcal{A}, z) := \sum_{\substack{d | P(z) \\ d \in \mathcal{D}^-}} \mu(d) |\mathcal{A}_d| \leq S(\mathcal{A}, z) \tag{2.25}$$

The choices we make for y_m can decide how good this process works. Brun used:

$$y_m = D^{\alpha \gamma^m} \tag{2.26}$$

for constants $0 < \alpha, \gamma < 1$. But we mostly nowadays choose:

$$y_m = (D_1 / p_1 \cdots p_m)^{\frac{1}{\beta}} \tag{2.27}$$

with $\beta \geq 1$.
In both cases $\max\{d : d \in D^{\pm}\} \leq D$, hence there is a combinatorial sieve that have a lower and upper bound.

Chapter 3

The General Number Field Sieve

This chapter will be about the General Number Field Sieve (Short GNFS). This algorithm is one of the fastest algorithms to solve the factorization of a general composite number. So the goal is to factorize a big composite number n to two unknown primes p_1 and p_2. Used and sometimes cited is A Beginner's Guide To The General Number Field sieve by Michael Case[2].

3.1 Difference of Squares Factorization method

The GNFS makes use of the Difference of squares Factorization method. Let n be a composite number with $n = pq$ and $s, r \in \mathbb{Z}, s^2 \equiv r^2 \mod n$. Then $pq|(s^2 - r^2)$, thus $pq|(s-r)(s+r)$ Hence, $p|(s-r)(s+r)$ and $q|(s-r)(s+r)$. Thus p and q have to devide **at least one** of $s-r$ and $s+r$. This gives rise to 9 divisiblity Scenarios of which 6 give back at least one of p and q. If one of p, q is found, the other one will be $\frac{n}{q}$ or $\frac{n}{p}$

$p\|(s+r)$	$p\|(s-r)$	$q\|(s+r)$	$q\|(s-r)$	$\gcd(pq, s+r)$	$\gcd(pq, s-r)$	Succes?
Yes	Yes	Yes	Yes	pq	pq	No
Yes	Yes	Yes	No	pq	p	Yes
Yes	Yes	No	Yes	p	pq	Yes
Yes	No	Yes	Yes	pq	q	Yes
Yes	No	Yes	No	pq	1	No
Yes	No	No	Yes	p	q	Yes
No	Yes	Yes	Yes	q	pq	Yes
No	Yes	Yes	No	q	p	Yes
No	Yes	No	Yes	1	pq	No

So if all divisibility scenario's have equal chance we have a $\frac{2}{3}$ chance that if we know a difference of squares scenario, it is a successful factorization method.
This step will be the last one of the GNFS. First the GNFS will try to find s and r, after which this method is used to get the p or q. With a failure of this method, the algorithm can be done again.

3.2 Free parameters in GNFS

There are two free parameters that must be chosen. The first parameter is a polynomial $f : \mathbb{R} \to \mathbb{R}$ with integer coefficients. The second parameter is a number $m \in N$ such that $f(m) \equiv 0 \mod n$.
This is equivalent to finding the base-m expansion of n:

$$n = a_d m^d + a_{d-1} m^{d-1} + a_{d-2} m^{d-2} + \ldots + a_1 m^1 + a_0$$

So finding f and m is writing down a number n in base-m, with arbitrarily chosen m. Thus for instance let $n = 713$, we first choose m the base we want to write our number in. This time 7 is chosen. Then we write this as a polynomial:

$$f(7) = 2 \cdot 7^3 + 0 \cdot 7^2 + 3 \cdot 7 + 6 = 713$$

Hence $f(m)$ is in this case:

$$f(m) = 2 \cdot m^3 + 0 \cdot m^2 + 3 \cdot m + 6 \tag{3.1}$$

We can see that $f(m) = n$ and so $f(m) \equiv 0 \mod n$.

3.3 $\mathbb{Z}[\theta]$

In this section the ring $\mathbb{Z}[\theta]$ is explained. This ring will be used thoroughly throughout the GNFS. Let $\theta \in \mathbb{C}$ a root of polynomial f and let d be the degree of f. The space $\mathbb{Z}[\theta]$ is defined as follows.

$$\mathbb{Z}[\theta] = \{x : x = a_{d-1}\theta^{d-1} + a_{d-2}\theta^{d-2} + \ldots + a_1\theta^1 + a_0 \text{ for } \{a_j\} \subset \mathbb{Z}\}$$

The addition on this space to make this space a ring, is just the addition of polynomials. Now is shown hat the product of two elements of this ring is again in the ring $\mathbb{Z}[\theta]$.
Let $A = a(\theta), B = b(\theta) \in \mathbb{Z}[\theta]$ With the algorithm for polynomial division, we can divide $a(\theta) \cdot b(\theta)$ by $f(\theta) = 0$ to get:

$$AB = a(\theta)b(\theta) = g(\theta)f(\theta) + c(\theta) = g(\theta) \cdot 0 + c(\theta) = c(\theta) = C$$

, where C is again a polynomial, with a degree less than d. Hence $C \in \mathbb{Z}[\theta]$.
Below is an example that goes further on the above polynomial 3.1. Let $A = 4\theta^2 + 2$ and $B = \theta^2 + 1$. Then $AB = 4\theta^4 + 6\theta^2 + 2$. We can see:

$$AB = 4\theta^4 + 6\theta^2 + 2 = 2\theta \cdot (2\theta^3 + 3\theta^2 + 6) + (-12(\theta) + 2) = (-12(\theta) + 2) \in \mathbb{Z}[\theta]$$

So products are also well-defined. In conclusion the space $\mathbb{Z}[\theta]$ is a ring.

3.4 Finding two squares

The reason we can find two squares such that $x^2 \equiv y^2 \mod n$ is the use of the following result:

Lemma 3.4.1. Given a polynomial $f(x)$ with integer coefficients a root $\theta \in \mathbb{C}$, and an $m \in \mathbb{Z}/n\mathbb{Z}$ such that $f(m) \equiv 0 \mod n$, there exist a unique mapping $\phi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ satisfying:

1. $\phi(ab) = \phi(a)\phi(b)\ \forall a, b \in \mathbb{Z}[\theta]$
2. $\phi(a + b) = \phi(a) + \phi(b)\ \forall a, b \in \mathbb{Z}[\theta]$
3. $\phi(1) = 1 \mod n$
4. $\phi(\theta) = m \mod n$.

The GNFS will try to find a difference of squares congruences in the following way: Let U be a finite set of integers (a, b), $\beta \in \mathbb{Z}[\theta]$ and $y \in \mathbb{Z}$ such that

$$\prod_{(a,b)\in U} (a + b\theta) = \beta^2 \text{ and } \prod_{(a,b)\in U} (a + bm) = y^2$$

Let $x = \phi(\beta)$ then working congruent modulo n we can find:

$$\begin{aligned} x^2 &= \phi(\beta)\phi(\beta) && \text{with rule 1,} \\ &= \phi(\beta^2) \\ &= \phi(\prod_{(a,b)\in U} (a + b\theta)) && \text{with rule 1 again,} \\ &= \prod_{(a,b)\in U} \phi(a + b\phi) && \text{with rule 4,} \\ &= \prod_{(a,b)\in U} (a + bm) \\ &= y^2 \end{aligned}$$

So now we have found a relation $x^2 \equiv y^2 \mod n$ and by 3.1 we have a $\frac{2}{3}$ chance for a solution.

3.5 Smoothness over a factor base

In this section smoothness over a factor base will be described. Secondly a process to find such smooth elements in both sets $\mathbb{Z}$ and $\mathbb{Z}[\theta]$ is described. It is used to find square numbers in $\mathbb{Z}$ and $\mathbb{Z}[\theta]$. For $\mathbb{Z}$ finding a square is relatively easy, though we can define a square number as some number $p_1^{a_1} \cdot p_2^{a_2} \cdot p_3^{a_3} \cdot \ldots \cdot p_n^{a_n}$ with $\forall i \ a_i = 0 \mod 2$.

Definition 3.5.1. A Rational factor base $\mathcal{R}$ is a finite collection of primes.

We can see here that $\mathcal{A}$ is a rational factor base, though we want to call it in this section $\mathcal{R}$ Because of the R in Rational. The next definition is about smoothness on a rational factor base $\mathcal{R}$

Definition 3.5.2. A number N is smooth on a rational factor base if the number only has prime factors of that factor base.

We will look at an example now. Let the factor base $\mathcal{R}$ be $\{2, 3, 5, 7, 11, 13, 17\}$ We can see that $14365 = 17 \cdot 13^2 \cdot 5$ is smooth on the factor base $\mathcal{R}$, but $3542 = 23 \cdot 7 \cdot 2 \cdot 11$ isn't, because $23 \notin \mathcal{R}$. In the ring $\mathbb{Z}[\theta]$, this can be done in a similar way.

Definition 3.5.3. An algebraic factor base is a finite subset $\mathcal{A} \subset \mathbb{Z}[\theta]$, where every element of $\mathcal{A}$ is of the form $\{p\theta + q\}$ such that there are not two elements a, b in $\mathbb{Z}[\theta]$ with $a \cdot b = x$ These elements generate a subset of the prime ideals of $\mathbb{Z}[\theta]$.

The Rational Factor Base also has a smoothness definition, which logically follows from definition 3.5.3.

Definition 3.5.4. An element $l \in \mathbb{Z}[\theta]$ is smooth over an algebraic factor base $\mathcal{A}$ if it is writable as a product of factors from $\mathcal{A}$

To write such elements down in program language we need this theorem:

Theorem 3.5.1. Let $f(x)$ be a polynomial with integer coefficients and let $\theta \in \mathbb{C}$ be a root of $f(x)$. Then the set of pairs (r, p) where p is a prime integer and $r \in \mathbb{Z}/p\mathbb{Z}$ with $f(r) \equiv 0 \mod p$ is in bijective correspondence with the set of $a + b\theta \in \mathbb{Z}[\theta]$ that satisfy the criteria for being in an algebraic factor base

To find smooth elements in $\mathbb{Z}$ and $\mathbb{Z}[\theta]$ means to find (a, b) such that $a + b\theta$ is smooth in some Algebraic factor base $\mathcal{A}$ and $a + bm$ is smooth in some Rational factor base $\mathcal{R}$ To find these smooth numbers sieve arrays will be used. This section is very well described in and almost verbatim taken from [2, p. 9]

1. Fix $b \in \mathbb{Z}$ and let N be an arbitrary positive integer.

2. Let a vary from $-N$ to N and create two arrays for the algebraic and rational factor base. One for the values $a + b\theta$ and the other one $a + bm$.

3. For q_i in $\mathcal{R}$, q_i divides $a + bm$ if and only if $a \equiv -bm \mod q_i$. Find values of a for which $a = -bm + kq_i$ for some $k \in \mathcal{Z}$ and for each value a make note of this factor of $a + bm$ in the sieve array. Repeat this for every $q_i \in \mathcal{R}$ Now make a note of all $a + bm$ in de sieve array that are completely factored by this method. These $a + bm$ are smooth in $\mathcal{R}$.

4. Now proceed in the same way for the algebraic factor base. So an $(r_i, p_i) \in \mathcal{A}$ divides $a + b\theta$ if and only if $a \equiv -br_i \mod p_i$. Find values of a satisfying $a = br_i + kp_i$ for some $k \in \mathbb{Z}$. For each a found, make note of this (r_i, p_i) factor of $a + b\theta$ in the sieve erray. When finished there will be a list of (r_i, p_i) factors. If $\prod p_i = (-b)^d f(-a/b)$ then this list of factors is a complete factorization and hence $a + b\theta$ is smooth over the given algebraic factor base $\mathcal{A}$.

5. Compare the two arrays by entry. If at any position $a + bm$ and $a + b\theta$ are smooth, then this (a, b) is what is sought after, so save them.

$-N + b\theta$	$-N + bm$
$-(N-1) + b\theta$	$-(N-1) + bm$
$\vdots$	$\vdots$
$(N-1) + b\theta$	$(N-1) + bm$
$N + b\theta$	$N + bm$

Table 3.1: example of a Sieve Array

We can repeat this process for different b to find enough (a, b) that we need.

3.6 Verifying squares in $\mathbb{Z}$ and $\mathbb{Z}[\theta]$

It's relatively easy to find out if numbers in $\mathbb{Z}$ are squares. Let $s \in \mathbb{Z}$ and its prime factorization $s = p_1^{q_1} p_2^{q_2} \cdots p_n^{q_n}$, than we can say that s is a square if $\forall i, q_i \equiv 0 \mod 2$. For $\mathbb{Z}[\theta]$ the equivalent of this test, is just one have of the solution. An element $t \in \mathbb{Z}[\theta]$ is square if its factorization is: $t = (a_1 + b_1\theta)^{q_1}(a_2 + b_2\theta)^{q_2} \cdots (a_n + b_n\theta)^{q_n}$ and $\forall i, q_i \equiv 0 \mod 2$.
So, this is the first condition that must hold to find that $t \in [\theta]$ is a perfect square. For the second test the Legendre symbol is needed:

Definition 3.6.1. The Legendre symbol $\left(\frac{a}{p}\right)$ for $a \in \mathbb{Z}$ and p a prime integer is defined as:

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{if } x^2 \equiv a \mod p \text{ has a solution} \\ -1 & \text{if } x^2 \equiv a \mod p \text{ has no solution} \\ 0 & \text{if } p|a \end{cases}$$

The following theorem will be used to increase the chances that t will really be a square element of $\mathbb{Z}[\theta]$

Theorem 3.6.1. Let U be a set of (a, b) pairs such that $\prod_{(a,b)\in U}(a + b\theta)$ is a perfect square in $\mathbb{Z}[\theta]$. Then for any (s, q) with q prime and s given as in theorem 3.5.1 with $(s, q) \nmid a + b\theta$ for any $(a, b) \in U$

$$\prod_{(a,b)\in U} \left(\frac{a + bs}{q}\right) = 1$$

To find such elements we only have to use some q that are higher than p.
So we can test if $l \in \mathbb{Z}[\theta]$ is a square element in the following way:

1. Verify that for a factorization $t = (a_1+b_1\theta)^{q_1}(a_2+b_2\theta)^{q_2} \cdots (a_n+b_n\theta)^{q_n}$ has $\forall i, q_i \equiv 0 \mod 2$.

2. Let $\mathcal{Q}$ be a set of pairs of numbers (s, q) with q prime and s as in theorem 3.5.1. Let $(s, q) \nmid a + b[\theta]$ for every $a + b\theta$ occurring in the factorization of t. Verify

$$\prod_{(a,b)\in U} \left(\frac{a + bs}{q}\right) = 1$$

   For U as in Theorem 3.6.1. $\mathcal{Q}$ is called the quadratic character base.

3. If the two above are satisfied, than t is probably a perfect square in $\mathbb{Z}[\theta]$. If $\mathcal{Q}$ is enlarged, the probability of t being a square will increase.

3.7 Combine everything

In this section the goal is to combine everything to find $V \subset U$. Such that:

1. $\prod_{(a,b)\in V} a + bm$ is square in $\mathbb{Z}$.

2. $\prod_{(a,b)\in V} a + b\theta$ is square in $\mathbb{Z}[\theta]$.

In section 3.5 the arrays that were constructed in the end of the paragraph gave us a set $\{(a_i, b_i)\} \in U$ that had the following properties:

1. $a + bm$ is smooth over the rational factor base $\mathcal{R}$.

2. $a + b\theta$ is smooth over an algebraic factor base $\mathcal{A}$.

Hence all the possible products of elements of $V \subset U$ defined by $\prod_{(a,b)\in V} a + bm$ and $\prod_{(a,b)\in V} a + b\theta$ are smooth over $\mathcal{R}$ respectively $\mathcal{A}$. Since these products are now smooth it is needed to check these four things:

1. $\prod_{(a,b)\in V}(a + bm) > 0$, Since a square has to be a positive number in $\mathbb{Z}$.

2. $\prod_{(a,b)\in V}(a+bm)$ is writable as $p_1^{e_1}p_2^{e_2}\cdots p_n^{e_n}$, where $\mathcal{R}=\{p_1,p_2,...,p_n\}$ form the rational factor base $\mathcal{R}$. And $\forall i, e_i \equiv 0 \mod 2$.

3. $\prod_{(a,b)\in V}(a+b\theta)$ is writable as $r_1^{f_1}r_2^{f_2}\cdots r_m^{f_m}$, where $\mathcal{A}=\{r_1,r_2,...,r_m\}$ form the algebraic factor base $\mathcal{A}$. And $\forall i, f_i \equiv 0 \mod 2$.

4. $\prod_{(a,b)\in V}\left(\frac{a+bs_i}{q_i}\right)=1$ for all $(s_1,q_1),\ldots,(s_k,q_k)=\mathcal{Q}$ defined in section 3.6.

We can now give every element $(a,b)\in U$ a vector of length $1+n+m+k$ to help us chech the 4 things mentioned above.

1. where the first element will be 0 if $a+bm>0$ and 1, else. A sum of these elements will now give 0 $\mod 2$ if $\prod_{(a,b)\in V}(a+bm)>0$ and 1 $\mod 2$ else.

2. The next n elements in this vector for (a,b) will be the e_i named at the list above. A product $\prod_{(a,b)\in V}(a+bm)$ will be square if all e_i are even, so if the sum for each e_i in the product are 0 $\mod 2$.

3. The next m elements in this vector for (a,b) will be the f_i named at the list above. A product $\prod_{(a,b)\in V}(a+b\theta)$ will be square if all f_i are even, so if the sum for each f_i in the product are 0 $\mod 2$.

4. The last k elements in this vector will be used to check if for any $(s_i,q_i)\in\mathcal{Q}$, $\prod_{(a,b)\in V}\left(\frac{a+bs_i}{q_i}\right)=1$. We can define for every element $(a,b)\in U$ the number 0 if $\left(\frac{a+bs_i}{q_i}\right)=1$ and 1 if $\left(\frac{a+bs_i}{q_i}\right)=-1$ In this case the sum of multiple elements will give 0 $\mod 2$ if $\prod_{(a,b)\in V}\left(\frac{a+bs_i}{q_i}\right)=1$ and 1 $\mod 2$ if $\prod_{(a,b)\in V}\left(\frac{a+bs_i}{q_i}\right)\neq 1$.

So a vector for $(a,b)\in U$ looks like this:

$$\Big(\underbrace{a_1}_{1.}\underbrace{a_2\cdots a_{1+n}}_{2.}\underbrace{a_{1+n+1}\cdots a_{1+n+m}}_{3.}\underbrace{a_{1+n+m+1}\cdots a_{1+n+m+k}}_{4.}\Big)$$

We have now defined for every $(a,b)\in U$ a vector with $1+n+m+k$ elements. So if we have found enough U such that: $|U|>1+n+m+k$ we can make a $|U|\times 1+n+m+k$-matrix C.
Now a solution will be a product, for which every element of the matrix will be 0. Hence there must be a vector $a=(a_1,a_2,...,a_{|U|})^T$ such that $Ca=0 \mod 2$. The $(a,b)\in U$ for which a_1=1 are the $(a,b)\in V$ which make all 4 given properties above 0 $\mod 2$. And thus both $\mathbb{Z}$ and $\mathbb{Z}[\theta]$ square.
Then, with paragraph 3.4 it is possible to find a relation $x^2\equiv y^2 \mod n$
And after that there is a $\frac{2}{3}$ chance to have found a prime factorization following the Difference of Squares Factorization method in paragraph 3.1

Chapter 4

Discussion and Conclusion

In this thesis we considered the combinatorial sieve and the general number field sieve. Both can be seen as successors of the Sieve of Eratotsthenes. In this final chapter we look at further developments as well as compare these two modern sieve. Some further developments after the combinatorial sieve are Chen's theorem and the result of Zhang-Maynard. Both of these results say something about twin primes and make a beginning of proving the twin prime conjecture. Beside the combinatorial Sieve also Selberg's sieve is used for these progressions. Selberg's sieve[3, ch 7] is changing the sift weights on a different way than the combinatorial sieve does. There the weight of the sieve is slowly decreasing depending on the bigness of the number with a logarithmic scale. In the second part the consequences of the GNFS on RSA are discussed. There is a discussion on how the GNFS makes it possible to crack the RSA-algorithm. Then we may question ourselves is RSA really safe, or are there further developments for which we have to watch out.

4.1 How to go further with sieves

This section will be about Combinatorial sieves of Chapter 2 and looks at the twin prime conjecture and how we are getting closer to solve it using a combinatorial sieve and sift weights.
Let p_n be the n-th prime. The twin conjecture is:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) = 2 \tag{4.1}$$

The first proven result that will be discussed in this paragraph is the one of Chen's Theorem. Chen's theorem is described below.
Let p and $p+2$ be both primes or one a prime and the other one writable as a multiplication with two prime factors. Then there are infinitely many p that hold.
The mathematician Chen used the combinatorial sieve to solve this problem in his paper. The other important sieve with sieve weights is Selberg's sieve. This result is also proven by Ross[1] in his paper with Selberg's sieve.
Another newer result that is proven has just yet been solved.
Let p_n be the n-th prime. then:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) < N \text{ for } N = 7 \cdot 10^6 \tag{4.2}$$

This result has been proven by the mathematician named Zheng in this decade, using Selberg's sieve. After this result, Maynard tried to lower this limit and succeeded by getting it to 645.

4.2 Number of calculations using GNFS

In this part the beginning of [5], will be used to find a way to calculate how fast we can crack the code of RSA. To find the number of calculations needed for a given x to get a given $x^2 \equiv y^2 \mod n$

$$\exp\left(\left(\sqrt[3]{\frac{64}{9}} + o(1)\right)(\log n)^{\frac{1}{3}}(\log\log n)^{\frac{2}{3}}\right) \tag{4.3}$$

Now let in equation 4.3, $n = 2^k$ and $o(1) = 0$. LogPlot in Wolphram Alpha for $k = 1$ to 2048 gives this this graph.

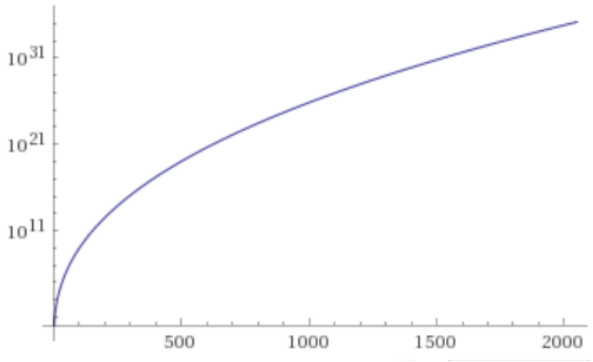

Figure 4.1: (k, number of calculations)

,

Right now the biggest supercomputer that will be build in 2021 will have 1.5 exaflops.[7] Or can do about $1.5 \cdot 10^{18}$ calculations per second. So let us take this number and look at $k = 512, 1024, 2048$. The number of calculations we have to do for each k that are mentioned are about: $1.76 \cdot 10^{19}$ for $k = 512$, $1.32 \cdot 10^{26}$ for $k = 1024$ and $1.53 \cdot 10^{35}$ for $k = 2048$. To calculate the number of seconds, we just have to divide this by the number of calculations we can do in a second, which give these results: for $k = 512$, $t \approx 11s$. for $k = 1024$ $t \approx 4y$. for $k = 2048$ $t \approx 4,9$ billion years. It is easy to see that there are big differences for different n that are often used. But $n = 2^{2048}$ seems safe for now, via the GNFS.
Quantum computers may fasten the process in the future a lot, due to Shor's algorithm, though, that is not yet a problem for RSA.

4.3 Is RSA really safe?

For immediate security RSA-1024 and RSA-2048 are definitely safe. The methods to crack these algorithms are not yet good enough and take some years to find the message. The problem is that we can save the encrypted messages and maybe uncover these messages in a later time, when RSA-2048 can be cracked(By most probably Quantum Computers.) When this will happen is a big question.

4.4 Conclusion

In this paragraph there is a discussion about the differences and similarities between the two sieves discussed in this thesis.

The first similarity is that both sieves are about primes. In Chapter 2, there was a discussion over an upper- or lower bound for the number of primes in a set. While in Chapter 3.7, the goal is finding the prime factorization of a big composite number.
The second similarity, is that both have some kind of sifting process. In the Combinatorial sieve we start with a set $\mathcal{A}$ and try to find any element in this set that is a prime or at least are following the rules that are set. In the GNFS we have sieve arrays, where we start with all $a + bm$ and $a + b\theta$ for a and b in between some numbers. After the process, we have a set U over which we than find some smooth products.
The third similarity is that there are in both sieves a sifting process that uses a relatively small set of primes. In both the sieve of Eratosthenes and the combinatorial sieve, we always sift over the primes $p_i \leq z$, while in the GNFS there is a similar thing with the Rational Factor Base and Algebraic Factor Base.
The main difference is that we are looking for different things. The GNFS is looking for a prime factorization, while the combinatorial Sieve is counting the primes in a set. The consequence is that for the combinatorial Sieve we are using more analytical number theory, while for the GNFS more algebraic number theory is used. This makes that however both are modern applications of the Sieve of Eratosthenes, the two solutions really differ a lot. To conclude the Sieve of Eratosthenes has a lot of modern applications in mathematics and those are in multiple fields of number theory. That doesn't mean that if they are called sieves, the same procedures are used in both the sieves. In the examples of this thesis, we have found that the sieves are applications in algebraic number theory and in Analytic number theory and have a completely different way of going to the solution.

Whole bibliography

[1] PM Ross. "On Chen's theorem that each large even number has the form p1+ p2 or p1+ p2p3". In: *Journal of the London Mathematical Society* 2.4 (1975), pp. 500–506.

[2] Michael Case. "A beginner's guide to the general number field sieve". In: *Oregon State University, ECE575 Data Security and Cryptography Project* (2003).

[3] John B Friedlander and Henryk Iwaniec. *Opera de cribro*. Vol. 57. American Mathematical Soc., 2010.

[4] George Greaves. *Sieves in number theory*. Vol. 43. Springer Science & Business Media, 2013.

[5] Jonathan D Lee and Ramarathnam Venkatesan. "Rigorous analysis of a randomised number field sieve". In: *Journal of Number Theory* 187 (2018), pp. 92–159.

[6] David Ireland. *RSA Algorithm*. URL: https://www.di-mgt.com.au/rsa_alg.html#note3.

[7] James Vincent. *World's fastest supercomputer will be built by AMD and Cray for US government*. URL: https://www.theverge.com/2019/5/7/18535078/worlds-fastest-exascale-supercomputer-frontier-amd-cray-doe-oak-ridge-national-laboratory.