Public Key Cryptography over the Gaussian and Eisenstein Integers

Stefanie Fazekas

A thesis submitted for the degree of
Master of Science

Maynooth University
Department of Mathematics and Statistics
July, 2023

Head of Department: Stephen M. Buckley
Supervisor: Stephen M. Buckley

Abstract.

Public key encryption has been studied in great detail mainly in the context of the ring of integers. In this thesis, we develop analogues of popular encryption schemes in the settings of Gaussian and Eisenstein integers. We will specifically study analogues of the following asymmetric public key encryption schemes: RSA algorithm, Diffie-Hellman Key Exchange and El Gamal.

Acknowledgements

A special thank you goes to Prof. Stephen Buckley, who supervised my master's thesis and always provided me with valuable advice. I am deeply grateful to Stephen for his availability and constant help.

I would like to thank the staff of the Department of Mathematics in Maynooth University for their help and support.

I would also like to thank Bernie for making my time in Ireland unforgettable and for giving me a second home.

To my friends, I am glad to have each and every one of you. Thank you for always being there for me.

The greatest thank you goes to my family. I would not be here without you. Thank you for supporting me and always believing in me.

Contents

Page

1. Introduction 6

2. Ring theoretic background 9

3. Gaussian Integers 14
3.1. Gaussian Sieve 17

4. Eisenstein Integers 26
4.1. Eisenstein Sieve 29

5. Number theory in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ 35
5.1. Exponentiation for Gaussian and Eisenstein integers 53

6. Group theoretic background 56
6.1. The group $\mathbb{Z}_N^*$ 58
6.2. Isomorphisms and the Chinese Remainder Theorem 58
6.3. Cyclic Groups 60

7. Factorization and Primality Testing 63
7.1. Fermat's primality test 64
7.2. The Miller-Rabin Test 66

8. Primality Testing in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ 70

9. Isomorphisms between certain quotient rings 72
9.1. Isomorphisms between quotients of $\mathbb{Z}$ and $\mathbb{Z}[i]$ 72
9.2. Isomorphisms between quotients of $\mathbb{Z}$ and $\mathbb{Z}[\omega]$ 74

10. Public vs Private Key Encryption: overview 77

11. RSA Algorithm in $\mathbb{Z}$ 78
11.1. RSA Digital Signature 80

12. Attacking RSA 82
12.1. Factorization 82
12.1.1. Fermat's Method 82
12.1.2. Pollard Rho Method of Factorization 83
12.2. Encrypting short messages using small $\boldsymbol{e}$ 86
12.3. Encrypting related messages 87
12.4. Sending the same message to multiple receivers 87
12.5. Blinding 88

13. RSA Algorithm in $\mathbb{Z}[i]$ 89
13.1. Case 1: A $\mathbb{Z}$-prime and a non-real Gaussian prime 89
13.2. Case 2: Two $\mathbb{Z}$-primes 90
13.3. Case 3: Two non-real Gaussian primes 91
13.4. RSA Digital Signature in $\mathbb{Z}[i]$ 93
13.5. Attacking RSA in $\mathbb{Z}[i]$ 93
13.6. Conclusion 93
14. RSA Algorithm in $\mathbb{Z}[\omega]$ 94
14.1. Case 1: A $\mathbb{Z}$-prime and a non-real Eisenstein prime 94
14.2. Case 2: Two $\mathbb{Z}$-primes 95
14.3. Case 3: Two non-real Eisenstein primes 95
14.4. Conclusion 97
15. Diffie-Hellman Key Exchange in $\mathbb{Z}$ 98
16. Diffie-Hellman Key Exchange in $\mathbb{Z}[i]$ 99
16.1. Case 1: Using a non-real Gaussian prime 99
16.2. Case 2: Using a $\mathbb{Z}$-prime 100
16.3. Conclusion 100
17. Diffie-Hellman Key Exchange in $\mathbb{Z}[\omega]$ 101
17.1. Case 1: Using a non-real Eisenstein prime 101
17.2. Case 2: Using a $\mathbb{Z}$-prime 101
17.3. Conclusion 102
18. El Gamal in $\mathbb{Z}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ 103
19. Attacking Diffie-Hellman and El Gamal 105
19.1. Brute-force attack 105
19.2. Pohlig-Hellman algorithm 105
19.3. Shank's baby step giant step algorithm 107
Bibliography 109

1. Introduction

Cryptography is the study of methods of sending messages in disguised form so that only the intended recipients can remove the disguise and read the message; a good reference is [**9**]. As stated in [**10**], a principal goal of public key cryptography is to allow people to exchange confidential information, even if they have never met and can communicate only via a channel that is begin monitored by an adversary.

Cryptography is a well-established field in mathematics and computer science. It is commonly used in everyday life and is a vital part of today's society and technology, without most people being aware of where it is used. Cryptography allows us to securely send emails, exchange messages on social media platforms and securely access websites. It is also an essential part of online banking and transactions. Digital currencies such as Bitcoin and Ethereum rely on the ideas developed in cryptography. These are just a few examples to illustrate how important cryptography is for our society.

Before the 1970s, most cryptographic systems used symmetric key algorithms, where the same key is used for both encryption and decryption. In the 1970s, however, asymmetric public key encryption had its breakthrough. Here, two different keys are used: a public key for encryption and a secret private key for decryption. Important cryptographic algorithms developed during this period include the RSA algorithm, the Diffie-Hellman Key Exchange and the El Gamal encryption scheme. Each of these algorithms is discussed in this thesis.

Various cryptographic methods using the ring of integers have been investigated. As the properties of encryption schemes in this ring are well studied, a question that arises is whether we can create an analogue in other rings. The rings of Gaussian and Eisenstein integers are natural choices for the next rings to be be considered in this regard.

In the first chapter, we present basic results in ring theory before going into detail about the rings of Gaussian integers $\mathbb{Z}[i]$ and Eisenstein integers $\mathbb{Z}[\omega]$. Prime numbers play an important role in cryptography. Therefore, we spend some time discussing the different types of primes we obtain in each ring. The Sieve of Eratosthenes allows us to find primes in $\mathbb{Z}$, and we discuss analogues of it for $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$.

Cryptography relies on results from number theory and group theory, so we devote Chapters 5 and 6 to these areas. In particular, we investigate modular arithmetic. In $\mathbb{Z}$, a set of residue classes representatives modulo a given $N > 1$ is simply the set $0, 1, \ldots, N-1$. By contrast, finding a set of residue classes in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ is not as straightforward: in fact, even counting the residue classes takes some effort. In $\mathbb{Z}[i]$, we count the residue classes using a version of Pick's theorem for polygons (see Theorem 5.21) while in $\mathbb{Z}[\omega]$, we use a rather different approach using limits (see Theorem 5.26). For us, the most useful complete set of residue classes in both rings consists of all points in a fundamental rhombic region, although we also give a rather

different set of representatives when N is a Gaussian or Eisenstein prime but not a $\mathbb{Z}$-integer. Moreover, we have to find an analogue for Fermat's Little Theorem in the Gaussian and Eisenstein integers.

Prime numbers and methods of factorizing large numbers into their prime factors play an important role in the security of cryptographic schemes. In fact, the RSA algorithm relies on the fact that it is difficult to find the prime decomposition of large numbers, while the Diffie-Hellman Key Exchange and the El Gamal encryption scheme rely on the difficulty of solving the discrete logarithm problem for rings of prime order. One question that might be asked is how we can efficiently determine whether or not a large integer is prime. Therefore, we introduce the concept of probabilistic primality testing. If a number passes this test we conclude that it is probably a prime (and we can take this probability as close to 1, limited only by the amount of time we are willing to allocate to the test). However, if it fails the test, we can conclude that the number is definitely composite.

In Chapter 9, we establish an isomorphism between quotient rings of $\mathbb{Z}[i]$ and quotient rings of $\mathbb{Z}$ as well as an isomorphism between quotient rings of $\mathbb{Z}[\omega]$ and quotient rings of $\mathbb{Z}$. These isomorphisms will later allow us to translate some of our work in the rings of $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ to equivalent work in $\mathbb{Z}$.

After a quick overview of public and private key encryption in Chapter 10, we will begin to look at the various encryption schemes in detail.

In Chapters 11 and 12, we explain how the RSA algorithm works in $\mathbb{Z}$ and present various methods that can be used to attack the encryption scheme. We then move on to examine the analogue to the RSA algorithm in $Z[i]$. The interesting aspect here is that we have to deal with three different cases. The first case is not secure at all, as it can easily be attacked by efficient factorization. The second case at first seems new, but the isomorphisms we established in Chapter 9 will reduce the problem to working in $\mathbb{Z}$. Thus, this case does not give any new insights. The third case is, however, different from any RSA scheme in $\mathbb{Z}$.

In a similar fashion, we also get three cases when analysing the RSA algorithm for the Eisenstein integers. Again, one of them is not secure, one gives no new insight, and one is new.

After discussing the RSA algorithm, we examine the Diffie-Hellman Key Exchange in Chapters 15–17. We will first introduce the idea in $\mathbb{Z}$ before examining the analogues in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$. In the ring of Gaussian integers, we have two different types of primes: $\mathbb{Z}$-primes and non-real Gaussian primes. Therefore, we obtain two different cases for the Diffie-Hellman Key Exchange in $\mathbb{Z}[i]$. Similarly, we also obtain two different cases for the Eisenstein integers. Here, we also need to explore the number of generators in our multiplicative groups and the form the generators can take.

The Diffie-Hellman Key Exchange only allows us to share a secret key, but it does not allow messages to be exchanged. However, the closely related El Gamal encryption scheme does allow for message to be exchanged. We discuss analogues of El Gamal encryption for the rings $\mathbb{Z}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ in Chapter 18.

In the last chapter, we briefly discuss different methods to attack Diffie-Hellman Key Exchange and El Gamal encryption.

2. Ring theoretic background

The Gaussian integers and the Eisenstein integers both form commutative rings. In fact, both are Euclidean domains. In this chapter, we first discuss unique factorization domains and principal ideal domains before moving on to Euclidean domains. For more details see **[6]**.

Throughout this thesis, $\mathbb{Z}$ is the set of all integers, while $\mathbb{N}$ and $\mathbb{Z}_{\geq 0}$ are the sets of positive and non-negative integers, respectively.

First, we recall the definition of a ring.

Definition 2.1. A ring R is a set together with two binary operations $+$ and $\times$ (called addition and multiplication) satisfying the following conditions for all $r, s, t \in R$:

(1) $(r + s) + t = r + (s + t)$ (*associativity* of $+$).
(2) $(r \times s) \times t = r \times (s \times t)$ (*associativity* of $\times$).
(3) There exists a zero element, denoted 0, such that $r + 0 = r = 0 + r$
(4) $r + s = s + r$ (*commutativity* of $+$).
(5) There exists an *additive inverse* $-r \in R$ such that $r + (-r) = 0 = (-r) + r$.
(6) There exists an element $1 \in R$ such that $r \times 1 = r = 1 \times r$.
(7) The following *distributive laws* hold:
    (a) $(r + s) \times t = r \times t + s \times t$.
    (b) $r \times (s + t) = r \times s + r \times t$.

From now on, we mainly use juxtaposition to indicate multiplication, e.g. we write xy in place of $x \times y$.

Rings that also satisfy the *commutative law* for multiplication, i.e. $rs = sr$ for all $s, t \in R$, are called *commutative rings*. If a, b are elements in a commutative ring R, we say that a divides b, written $a \mid b$, if $b = ca$ for some $c \in R$.

Definition 2.2. An *ideal* of a ring R is a nonempty subset I of R such that $(I, +)$ is a subgroup of $(R, +)$, and both xr and rx lie in R whenever $x \in I$ and $r \in R$. In a commutative ring R, a good way to define an ideal is to take the following set of finite sums for some fixed $x_1, \ldots, x_n \in R$:

$$\left\{ \sum_{i=1}^{n} r_i x_i \;\middle|\; r_1, \ldots, r_n \in R \right\},$$

This set, which is clearly an ideal, is denoted $(x_1, \ldots, x_n)$. A *principal ideal* is an ideal of the form (x) for some $x \in R$.

An element z in a ring R is:

- a *left zero divisor* if $z \neq 0$ and $zx = 0$ for some $x \in R$, $x \neq 0$.
- a *right zero divisor* if $z \neq 0$ and $yz = 0$ for some $y \in R$, $y \neq 0$.

An element $z \in R$ is a *zero divisor* if it is both a left and right zero divisor.

A ring R satisfies the *cancellation law* if

- $ab = ac \implies a = 0$ or $b = c$, and
- $ba = ca \implies a = 0$ or $b = c$

We omit the easy proof of the following lemma.

Lemma 2.3. *A ring R satisfies the cancellation law if and only if it has no left or right zero divisors.*

Definition 2.4. A ring R is called a *domain* if it has no left or right zero divisors. An *integral domain* is a commutative domain.

Definition 2.5. An element $u \in R$ is a *unit* if there exists $v \in R$ such that $uv = 1$ and $vu = 1$, i.e. u is a unit if it is invertible.

Definition 2.6. Elements a, b in a commutative ring R are *associates* if $a = bu$ for some unit u.

Definition 2.7. We say a nonzero element π in a domain R is *irreducible* if it is neither a unit nor the product of two non-units.

Definition 2.8. Let R be a commutative ring and let $p \in R$ be neither zero nor a unit. We say that p is *prime* if the condition $p \mid ab$ for some $a, b \in R$ implies that either $p \mid a$ or $p \mid b$.

Proposition 2.9. *In an integral domain, every prime element is irreducible.*

PROOF. Suppose p is a prime. We know that p is nonzero and not a unit. To prove irreducibility, it remains to prove that if $p = ab$, then a and b cannot both be non-units.

Suppose $p = ab$. By primality, p must divide either a or b. Without loss of generality, $a = pr$ for some r, and so $p \cdot 1 = ab = p \cdot rb$. Thus, $rb = 1$, so b is a unit and p is irreducible. □

Definition 2.10. A *unique factorization domain* (UFD) is an integral domain R in which each nonzero element can be written as a product of a unit and zero or more irreducibles, and this factorization is *unique up to units (and rearrangement)* in the sense that if $z \in R$ and

$$z = u \prod_{i=1}^{m} p_i = v \prod_{j=1}^{n} q_j \,,$$

where u, v are units, and every p_i and q_j is irreducible, then $n = m$ and there is a permutation ϕ on $\{1, \ldots, m\}$ such that p_i and $q_{\phi(i)}$ are associates for all $1 \leq i \leq m$.

In the above definition and in what follows, it is convenient to define an empty product to mean 1 in all cases. If $m = 0$, the existence of the permutation ϕ above can be dropped (or we can define a permutation on an empty set to be the unique empty relation on the empty set). If z is a non-unit, and so $m > 0$, we can always absorb the unit into one of the irreducibles (to get an associated irreducible) and so z is a product of irreducibles.

Lemma 2.11. *In a unique factorization domain, a nonzero element is prime if and only if it is irreducible.*

PROOF. Let R be a unique factorization domain. By Proposition 2.9, primes of R are irreducible, so it remains to show only that each irreducible element is a prime.

Let p be an irreducible in R and assume $p \mid ab$ for some $a, b \in R$. We must show that p divides either a or b. To say that p divides ab is to say $ab = pc$ for some c in R. Writing a and b as a product of irreducibles, we see from this last equation and from the uniqueness of the decomposition into irreducibles of ab that the irreducible element p must be an associate of one of the irreducibles occurring either in the factorization of a or in the factorization of b. We may assume that p is an associate of one of the irreducibles in the factorization of a, i.e. that a can be written as a product $a = (up)p_2 \dots p_n$ for u a unit and some (possibly empty set of) irreducibles $p_2, \dots, p_n$. But then, p divides a, since $a = pd$ with $d = up_2 \dots p_n$, completing the proof. □

Definition 2.12. A *principal ideal domain* (PID) is an integral domain in which every ideal is principal (as defined in Definition 2.2).

Theorem 2.13. *Every principal ideal domain is a unique factorization domain.*

PROOF. Let R be a principal ideal domain and let r be a nonzero element of R. Suppose first that r is a unit. Since $rs = 1$ for some $s \in R$, we deduce that $(r) = R$. Suppose that $r = upq$, where u is a unit, p is irreducible, and q is a product of zero or more irreducibles. Since p is irreducible, it is easy to prove that $(p) \neq R$. Since $(upq) \subset (p)$, this contradicts the equation $(r) = R$. This proves unique factorization for units, so we assume from now on that r is a non-unit.

First, we must show that r can be written as a finite product of irreducible elements of R. Then, we must verify that this decomposition is unique up to units.

If r is irreducible, then we are done, as r is nonzero and not a unit. If not, then r can be written as a product $r = r_1 r_2$, where neither r_1 nor r_2 is a unit. If both of these elements are irreducibles, then again we are done. Otherwise, at least one of the two elements, say r_1, is reducible, and so it can be written as a product of two non-unit elements $r_1 = r_{11} r_{12}$, and so forth.

We now must verify that this process terminates. Suppose this is not the case. From the factorization $r = r_1 r_2$, we obtain the proper inclusion of ideals: $(r) \subset (r_1) \subset R$. Note that the first inclusion is proper since r_2 is not a unit. From the factorization of r_1, we similarly obtain $(r) \subset (r_1) \subset (r_{11}) \subset R$. If this process of factorization did not terminate after a finite number of steps, then we would obtain an infinite ascending chain of ideals

$$(r) \subset (r_1) \subset (r_{11}) \subset \cdots \subset R\,,$$

where all containments are proper, and the Axiom of Choice ensures that an infinite chain exists.

We now show that any ascending chain $I_1 \subseteq I_2 \subseteq \cdots \subseteq R$ becomes stationary, i.e. there is some positive integer n such that $I_k = I_n$ for all $k \geq n$. Equivalently, it is not possible to have an infinite ascending chain of ideals where all containments are proper. Let $I = \bigcup_{i=1}^{\infty} I_i$. It follows easily that I is an ideal. Since R is a PID, it is principally generated, say $I = (a)$. Since I is the union of ideals above, a must be an element of

one of the ideals in the chain, say $a \in I_n$. But then we have $I_n \subseteq I = (a) \subseteq I_n$ and so $I = I_n$ and the chain becomes stationary at I_n This proves that every nonzero element of R which is not a unit has some factorization into irreducibles in R.

We next show that every irreducible is prime. Suppose that p is irreducible and that $p \mid ab$ for some $a, b \in R$. Let $I = (a, p)$ and so there exists d such that $(a, p) = (d)$. Since $p \in I$, we have $p = cd$ for some $c \in R$. By irreducibility, either c or d is a unit. If d is a unit, then $I = R$ and there exist $r, s \in R$ such that $ra + sp = 1$. Thus, $b = rab + spb$ and,'since $p \mid ab$, we have $p \mid b$. On the other hand, if c is a unit, then p and d are associates and so $(p) = (d)$. Since $a \in I$, we deduce that $p \mid a$.

It remains to prove that the above decomposition is unique up to units. We proceed by induction on n, the number of irreducible factors in some factorization of the element r. If $n = 0$, then r is a unit and we already considered this case. Inductively, we prove unique factorization for a product of $n = k > 0$ irreducibles, assuming that it holds for a product of n irreducibles for all $0 \leq n < k$.

Assume therefore that some $r \in R$ has the two factorizations

$$r = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_m \,,$$

where every p_i and q_i is irreducible (and so also prime, by the above). We may also assume, without loss of generality, that $m \geq k$. Then, p_1 divides the product on the right, so p_1 must divide one of the factors. Renumbering if necessary, we may assume that p_1 divides q_1, i.e. $q_1 = p_1 u$ for some element u of R. Now u must be a unit because q_1 is irreducible. Thus, p_1 and q_1 are associates. Cancelling p_1, we obtain the equation

$$p_2 \cdots p_k = u q_2 q_3 \cdots q_m = q_2' q_3 \cdots q_m \,.$$

where $q_2' = u q_2$ is again an irreducible (as associate to q_2). By induction, we conclude that each of the factors on the left matches bijectively (up to associates) with the factors on the far right, and hence (up to associates) with the factors in the middle. Since p_1 and q_1 have already been shown to be associates, this completes the induction step and the proof of the theorem. □

Definition 2.14. Any function $N : R \to \mathbb{Z}_{\geq 0}$ with $N(0) = 0$ is called a *norm* on the integral domain R. If $N(a) > 0$ for $a \neq 0$, we call N a *positive norm*.

Definition 2.15. A ring R is a *Euclidean domain* (ED) if R is an integral domain and there is a *Euclidean function* $\delta : R \setminus \{0\} \to \mathbb{N}$ such that

$$\delta(r) \leq \delta(rs) \qquad \text{for } r, s \in R \setminus \{0\}$$

and such that for all nonzero $a, b \in R$, there exists $q, r \in R$ such that $a = qb + r$ where either $r = 0$ or $r \neq 0$ and $\delta(r) < \delta(b)$, i.e. we have a division algorithm for R.

We will see that the Gaussian integers and the Eisenstein integers both form Euclidean domains with the very nice norm $N(z) = z\bar{z}$, $z \in R$ and $\delta = N|_{R \setminus \{0\}}$.

An ED always has many different Euclidean functions δ; in particular, we can always replace δ by $m\delta$ for any desired $m \in \mathbb{N}$. We typically assume that an ED R is equipped with some fixed Euclidean function δ. Then δ gives rise to a positive norm N, simply

as the extension of δ to R with $N(0) = 0$. We call this N the *standard norm* of the ED (with respect to δ).

Theorem 2.16. *Every Euclidean domain is a principal ideal domain.*

PROOF. Let N be the standard norm of an ED R. If I is the zero ideal, it is certainly principal. Otherwise, let d be any nonzero element of I with $N(d)$ minimal, where N is the standard norm; such a d exists since the set $\{N(a) \mid a \in I \setminus \{0\}\}$ has a minimal element by the Well Ordering of $\mathbb{N}$. Clearly, $(d) \subseteq I$ since d is an element of I. To show the reverse inclusion, let a be any element of I and use the Division Algorithm to write $a = qd + r$ with $N(r) < N(d)$. Then $r = a - qd$ and both a and qd are in I, so r is also an element of I. By the minimality of $N(d)$, we see that r must be 0. Thus, $a = qd \in (d)$, showing that $I = (d)$. □

Combining Theorems 2.16 and 2.13, we deduce:

Corollary 2.17. *Every Euclidean domain is a unique factorization domain.*

3. Gaussian Integers

As usual, we write complex numbers $z \in \mathbb{C}$ in the form $z = a + bi$, where $\mathrm{Re}(z) := a$ and $\mathrm{Im}(z) := b$ are real numbers, and $i^2 = -1$.

Throughout this chapter we are mostly following the ideas of **[15]**.

Definition 3.1. The Gaussian Integers $\mathbb{Z}[i]$ are the collection of all $z \in \mathbb{C}$ such that both $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$ are integers.

For $z = a + bi \in \mathbb{Z}[i]$, the *conjugate of* z is $\bar{z} = a - bi$. Addition and multiplication in the ring of Gaussian integers $\mathbb{Z}[i]$ are inherited from $\mathbb{C}$.

The *Gaussian integer norm*, $N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$, is defined by

$$N(a + bi) = (a + bi)\left(\overline{a + bi}\right) = a^2 + b^2.$$

Thus, $N(z)$ is the square of the absolute value of z.

Since $N(z) = z\bar{z}$, the following proposition follows readily.

Proposition 3.2. *$N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$ is completely multiplicative, i.e. we have $N(zw) = N(z)N(w)$ for all $z, w \in \mathbb{Z}[i]$.*

Our first application of the norm is a characterization of $\mathbb{Z}[i]$-units.

Proposition 3.3. *The following are equivalent for $u \in \mathbb{Z}[i]$.*

(a) u is a unit.
(b) $N(u) = 1$.
(c) $u \in \{\pm 1, \pm i\}$.

Proof. We first prove the equivalence of (a) and (b). Suppose u is a unit, and so $uv = 1$ for some $v \in \mathbb{Z}[i]$. Therefore, $N(u)N(v) = N(uv) = N(1) = 1$, and so both $N(u)$ and $N(v)$ must be equal to 1.

For the converse, assume that $N(u) = 1$. As $N(u) = u\bar{u}$, we get $u\bar{u} = 1$. Hence, u is a unit in $\mathbb{Z}[i]$.

The fact that (c) implies (b) is trivial. Finally, we prove that (b) implies (c). Suppose $u = a + bi$ satisfies $N(u) = 1$, and so $a^2 + b^2 = 1$. The only solutions to this equation are:

- $a = \pm 1$ and $b = 0$, or
- $a = 0$ and $b = \pm 1$.

Throughout this chapter, a *prime* (or *Gaussian prime* for emphasis) refers to a prime in $\mathbb{Z}[i]$. We will refer to primes in $\mathbb{Z}$ as *$\mathbb{Z}$-primes*.

Corollary 3.4. *If $z \in \mathbb{Z}[i]$ is such that $N(z)$ is a $\mathbb{Z}$-prime, then z is a Gaussian prime.*

PROOF. Let $z \in \mathbb{Z}[i]$ have prime norm, say $N(z) = p$. Consider any factorization of z in $\mathbb{Z}[i]$, say $z = xy$. Taking norms, we get $N(x)N(y) = p$. $N(x), N(y) \in \mathbb{Z}$, and as p is prime, either $N(x)$ or $N(y)$ is 1. Hence, either x or y is a unit and thus z is prime. □

As an example of the above corollary, we note that $3+2i$ is prime because $N(3+2i) = 13$ is prime. However, the converse implication is not true. For example, $N(3) = 9$ is not a $\mathbb{Z}$-prime but 3 is a Gaussian prime (as follows for instance from Theorem 3.6 below).

We next show that the norm provides us with a Euclidean function for $\mathbb{Z}[i]$.

Proposition 3.5. *The ring of Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain with Euclidean function $\delta : \mathbb{Z}[i] \setminus \{0\} \to \mathbb{N}$ given by $\delta(a + bi) := N(a + bi) = a^2 + b^2$.*

PROOF. To show that there is a division algorithm, we let $a + bi, c + di \in \mathbb{Z}[i]$ and we work inside $\mathbb{Q}[i] = \{\alpha + \beta i \mid \alpha, \beta \in \mathbb{Q}\}$:

$$\begin{aligned}\frac{a+bi}{c+di} &= \frac{a+bi}{c+di}\frac{c-di}{c-di} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} \\ &= \alpha + i\beta, \qquad \alpha, \beta \in \mathbb{Q}\,.\end{aligned}$$

We can choose $n, m \in \mathbb{Z}$ such that $|\alpha - n| \leq \frac{1}{2}$ and $|\beta - m| \leq \frac{1}{2}$. Now we get

$$\frac{a+bi}{c+di} = n + im + ((\alpha - n) + i(\beta - m))\,.$$

Let $n + im = q$ and $(\alpha - n) + i(\beta - m) = \gamma$, then

$$\frac{a+bi}{c+di} = q + \gamma\,.$$

Multiplying across by $(c + di)$ gives us $a + bi = q(c + di) + \gamma(c + di)$.

Now, $\gamma(c + di) = (a + bi) - q(c + di)$. The right-hand side $(a + bi) - q(c + di)$ lies in $\mathbb{Z}[i]$, and therefore the left-hand side $\gamma(c + di)$ also lies in $\mathbb{Z}[i]$.

Let $\gamma(c + di) = r$. We get

$$\delta(r) = \delta(\gamma(c + di)) = \delta(\gamma)\delta(c + di)\,,$$

and as

$$\delta(\gamma) = (\alpha - n)^2 + (\beta - m)^2 \leq \frac{1}{2} < 1\,,$$

we have $\delta(r) < \delta(c + di)$.

So, given $a + bi, c + di \in \mathbb{Z}[i]$, there exist $q, r \in \mathbb{Z}[i]$ such that $a + bi = q(c + di) + r$ and either $r = 0$ or $r \neq 0$, $\delta(r) < \delta(c + di)$. □

As $\mathbb{Z}[i]$ is a Euclidean domain, a Gaussian integer is irreducible if and only if it is prime. An associate of a Gaussian prime is also a Gaussian prime, as is the conjugate of a Gaussian prime.

Theorem 3.6. *Every Gaussian prime is a factor of a $\mathbb{Z}$-prime. Moreover, every $\mathbb{Z}$-prime p has one of the following $\mathbb{Z}[i]$-factorizations.*

- *If $p = 2$, then $2 = (1 + i)(1 - i)$ is a product of two associate Gaussian primes.*

- *If $p \equiv 3 \pmod 4$, then p is a Gaussian prime.*
- *If $p \equiv 1 \pmod 4$, then p is a product $\pi\bar{\pi}$ of two conjugate non-associate Gaussian primes.*

PROOF. Since $\mathbb{Z} \subset \mathbb{Z}[i]$, the first statement of the theorem is obvious. Thus, to find all Gaussian primes, it suffices to find the $\mathbb{Z}[i]$-prime factorization of all $\mathbb{Z}$-primes.

Let p be a $\mathbb{Z}$-prime. Then $N(p) = p^2$, and so either p is a Gaussian prime or it splits into two Gaussian primes of norm p. If p splits, then $p = \pi\sigma$, where $\mathrm{Im}(\pi), \mathrm{Im}(\sigma) \neq 0$. Since

$$\pi\sigma = p = N(\pi) = \pi\bar{\pi}\,,$$

we have $\sigma = \bar{\pi}$. Thus, if p splits, then $p = \pi\bar{\pi}$. Note that not only do we have $\mathrm{Im}(\pi) \neq 0$, but also $\mathrm{Re}(\pi) \neq 0$ (since π cannot be an associate of a $\mathbb{Z}$-prime).

Suppose $p = \pi\bar{\pi}$, where $\pi = a + bi$. Then $N(\pi) = a^2 + b^2 = p$. Squares of integers are equivalent to either 0 or 1 modulo 4, so $a^2 + b^2$ must be equivalent to 0, 1, or 2 modulo 4. In particular, if $p \equiv 3 \bmod 4$, then p does not split and all such $\mathbb{Z}$-primes are $\mathbb{Z}[i]$-primes.

Since p is prime, we can rule out the possibility that $a^2 + b^2 \equiv 0 \bmod 4$. The case $a^2 + b^2 \equiv 2 \bmod 4$ can occur only for $p = 2$, in which case we have $2 = (1 + i)(1 - i)$, so 2 does indeed split. Note that the split primes are associates: $1 - i = -i(1 + i)$.

The only remaining case in which p might possibly split is when $p \equiv 1 \bmod 4$, so let us consider this case. Euler's criterion gives the following equation for the Legendre symbol of -1 with respect to p:

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = 1\,,$$

and so -1 is a quadratic residue modulo p.

Hence, there exists an integer c such that $c^2 + 1 = kp$ and $0 < c < p$. Thus,

$$N(c + i) = (c + i)(c - i) = c^2 + 1 = kp\,.$$

If p does not split, then either $p \mid (c+i)$ or $p \mid (c-i)$. So, $p \mid 1$, which is a contradiction. We conclude that p is not a $\mathbb{Z}[i]$-prime and that it factors as a product of two $\mathbb{Z}[i]$-primes, i.e. $p = \pi\bar{\pi}$.

Lastly, we must show that $\bar{\pi}$ is not an associate of π. Suppose for the sake of contradiction that $\bar{\pi} = u\pi$ for some unit u. Writing $\pi = a + bi$ with $a, b \neq 0$, the equation $\pi\bar{\pi} = u\pi^2$ becomes

$$A := a^2 + b^2 = u(a^2 - b^2 + 2abi) \tag{3.7}$$

and so $\mathrm{Im}(A) = 0$. Either $u = \pm 1$, in which (3.7) implies that $2ab = 0$ (contradicting the fact that $a, b \neq 0$) or $u = \pm i$, in which (3.7) implies that $a^2 = b^2$ (contradicting the fact that $a^2 + b^2 \equiv 1 \bmod 4$). Since we get a contradiction in either case, we see that $\bar{\pi}$ is not an associate of π. □

Every Gaussian integer can be factorized into a product of (Gaussian) primes. This factorization is unique up to the order of the factors, multiplication by units and the replacement of a Gaussian prime by any of its associates.

3.1. Gaussian Sieve.

Similarly to the Sieve of Eratosthenes for the natural numbers, we can generate a sieve for the Gaussian integers which we will call the *Gaussian Sieve*. We first define the *first octant* O_1 to consist of all nonzero $z \in \mathbb{Z}[i]$ with $\mathrm{Re}(z) > 0$ and $0 \leq \mathrm{Im}(z) \leq \mathrm{Re}(z)$. (Equivalently, $\arg(z) \in [0, \pi/4]$, but we prefer to reserve the symbol π to denote a Gaussian prime.)

For the Gaussian sieve, the primary search region will be O_1. It suffices to find all primes in O_1 since all other primes are generated by conjugation and/or multiplication by units in $\mathbb{Z}[i]$. There are typically eight primes that we can immediately write down once we have any given prime $\pi = a + bi \in O_1$ (although in some cases, these eight reduce to four distinct primes): its associates $-a - bi$, $b - ai$, and $-b + ai$ are also prime, as are the conjugates $a - bi$, $-a + bi$, $b + ai$, and $-b - ai$.

Note though, that there is a big difference between a prime's associates and their conjugates. All associates are "essentially the same prime" for divisibility purposes, but their conjugates (outside of the special cases $a = b = 1$ and $ab = 0$) are distinct from the original prime. In particular, in order to sieve out all composite numbers, we must not only discard the multiples of the Gaussian primes in O_1 but also the multiples of their conjugates. For instance, $i(2-i)(5-2i) = 9+8i \in O_1$ is composite but it does not have prime factors in O_1.

Thus, although we search for primes only in O_1, we also need to deal with the larger set $\widetilde{O}_1 := \{z, \bar{z} \mid z \in O_1\}$. It is useful also to define an equivalence relation $\sim$ on $\widetilde{O}_1$ where $z \sim w$ if and only if $w \in \{z, \bar{z}\}$. We define the associated equivalence classes $[z] = \{z, \bar{z}\}$, $z \in \widetilde{O}_1$ and the set $\widetilde{O}_1/\sim$ of these equivalence classes.

To carry out the sieving process, the first thing that we need to do is to place the elements of O_1 in a sequence by defining a useful total order $\prec$ on O_1. Specifically, we define $\prec$ to be lexicographic order on the coordinates (a, b) of $a+bi \in O_1$, i.e. $a+bi \prec c+di$ if either $a < c$ or $a = c$ and $b < d$. We also denote by $\prec$ the induced total order on $\widetilde{O}_1/\sim$: for $z, w \in O_1$, we write $[z] \prec [w]$ iff $z \prec w$. As usual, we write $x \preceq y$ to mean that either $x \prec y$ or $x = y$, and we write $x \succ y$ as a synonym or $y \prec x$. We similarly define $\preceq$ and $\succ$ on $\widetilde{O}_1/\sim$.

It is clear that the two maps we call $\prec$ are total orders (on O_1 and on $\widetilde{O}_1/\sim$). In order for $\prec$ to be a *useful* total order for our sieving process, it should interact well with factorization in a sense that we now prove.

Lemma 3.8. *Suppose $x, y \in \widetilde{O}_1$ and $[x] \neq [y]$. Then,*

(a) y and x are not associates.
(b) If y is divisible by x, then $[x] \prec [y]$.

Consequently, the Gaussian prime factors $\pi \in \widetilde{O}_1$ of any $y \in \widetilde{O}_1$ all satisfy $[\pi] \preceq [y]$, with equality if and only if y is a Gaussian prime.

Proof. The arguments of any two distinct points in $S_w := \{w, iw, -w, -iw\}$, for any $w \in \mathbb{C} \setminus \{0\}$, differ by at least $\pi/2$. Since the angle of the cone $\widetilde{O}_1$ is $\pi/2$, the only

way that it can contain two distinct associates x, y of the one number is if one has argument $\pi/4$ and the other has argument $-\pi/4$, i.e. they both have the form $a \pm ai$ for some $a \in \mathbb{N}$. This is inconsistent with $[x] \neq [y]$, so we have proven (a).

We next prove (b). Note first that if $z = a + bi \in O_1$, then $a^2 \leq N(z) = a^2 + b^2 \leq 2a^2$, and so

$$(\mathrm{Re}(z))^2 \leq N(z) \leq 2\,(\mathrm{Re}(z))^2\,, \qquad z \in O_1\,.$$

If $u, v \in O_1$, $u \prec v$, then either

(i) $0 < \mathrm{Re}(u) < \mathrm{Re}(v)$ or
(ii) $0 < \mathrm{Re}(u) = \mathrm{Re}(v)$ and $0 \leq \mathrm{Im}(u) < \mathrm{Im}(v)$.

In case (i), we have $N(u) \leq 2(\mathrm{Re}(u))^2 < 2(\mathrm{Re}(v))^2 \leq 2N(v)$. In case (ii), it is clear that $N(u) < N(v)$. So, in both cases, $N(u) < 2N(v)$ whenever $u \prec v$. In view of the identity $N(w) = N(\bar{w})$, $w \in \mathbb{Z}[i]$, we conclude that

$$N(u) < 2N(v), \qquad \text{for all } u, v \in \widetilde{O}_1,\ [u] \prec [v]\,. \tag{3.9}$$

Suppose now that $x, y \in \widetilde{O}_1$, $[x] \neq [y]$, and that $y = zx$ for some $z \in \mathbb{Z}[x]$. By (a), x, y are not associates. Thus, $N(y) = N(z)N(x) \geq 2N(x)$. In view of (3.9), we cannot have $[y] \prec [x]$. Since $[y] \neq [x]$, we have proved (b).

The final statement of the lemma follows easily from (b). □

With Lemma 3.8 in hand, we can find the Gaussian primes by sieving as follows. We first discard 1, the only unit in O_1. We then inductively declare the next remaining element $x \in O_1$ to be a Gaussian prime and discard all multiples $y \in O_1$ of x satisfying $x \prec y$ and repeat the process. (Above and later, terms such as *next* and *previous* are always defined with respect to $\prec$.) To carry this out in a practical setting (typically on a computer), we confine our work to a bounded initial segment of O_1 such as all $z \in O_1$ satisfying $z \preceq M + Mi$ for some $M \in \mathbb{N}$. We illustrate this process by carrying it out for $M = 10$.

Below, whenever we talk of discarding multiples of x, we mean that we discard from consideration all Gaussian integer multiples $y \in O_1$ of x satisfying $x \prec y \preceq 10 + 10i$.

Example 3.10. We use the Gaussian Sieve to find all primes $\pi \in O_1$, $\pi \preceq 10 + 10i$. In the diagrams associated with each step of the sieving process, the Gaussian integers are the intersection points of dotted lines, we mark the primes in O_1 that we have found so far by $\bullet$, we mark the unit 1 by $\times$, and we mark all associated (discarded) multiples of those primes by either $\otimes$ or $\times$, depending on whether this element was discarded at the current step or an earlier step, respectively. Gaussian integers in O_1 whose primality status has yet to be decided after some step of the induction are shown as undecorated intersection points.

After discarding 1, the next element of O_1 is $1 + i$. We declare it to be a prime and discard all multiples y of $1+i$. (Unlike later primes, we do not need to discard multiples of the conjugate prime $1 - i$ because $1 + i$ and $1 - i$ are associates.) This concludes Step 1 and is illustrated in Figure 1.

We now move on to $2 + i$, the next remaining element of O_1 after $1 + i$. We declare $2 + i$ to be a prime and discard all multiples y of $2 \pm i$. This concludes Step 2 and is illustrated in Figure 2.

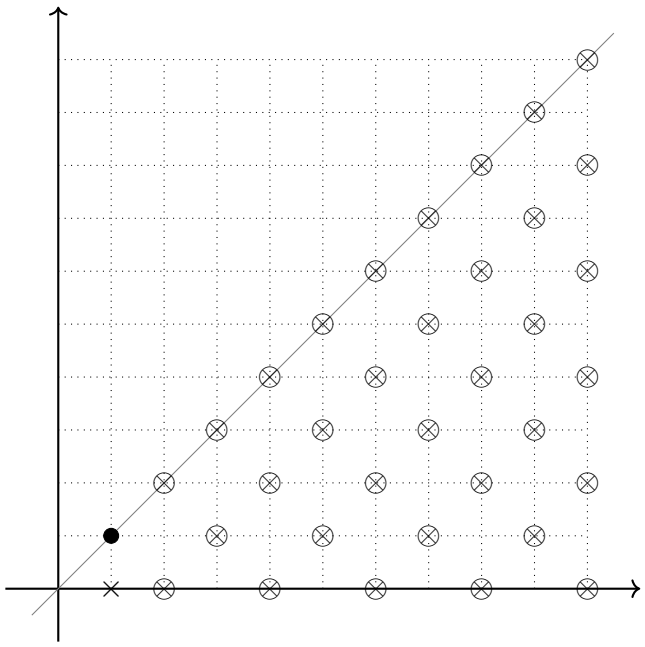

FIGURE 1. Gaussian Sieve after Step 1

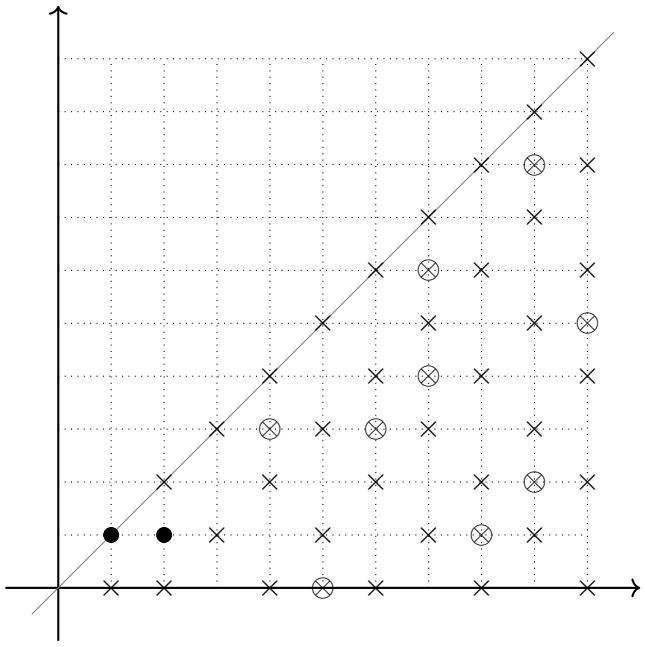

FIGURE 2. Gaussian Sieve after Step 2

The next remaining element of O_1 after $2+i$ is 3. We declare 3 to be prime and discard all its multiples. This concludes Step 3 and is illustrated in Figure 3.

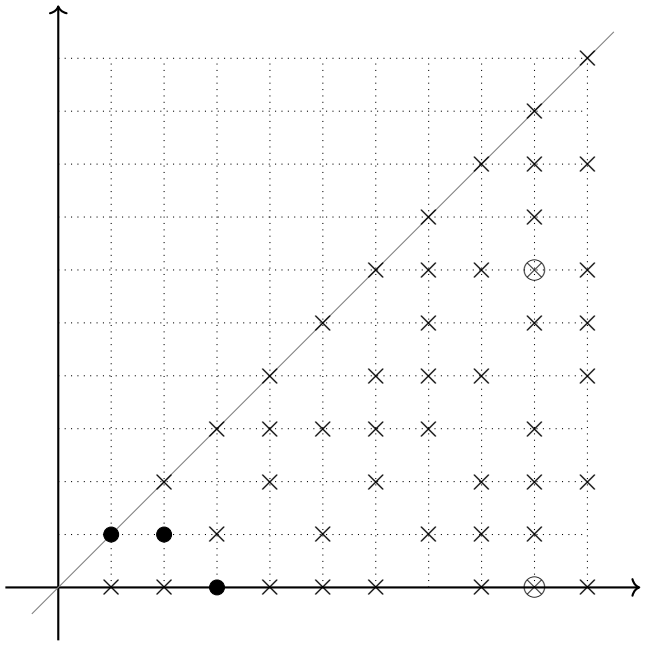

FIGURE 3. Gaussian Sieve after Step 3

The next remaining element of O_1 after 3 is $3+2i$. We declare it to be a prime and discard all multiples of $3 \pm 2i$ to get Figure 4.

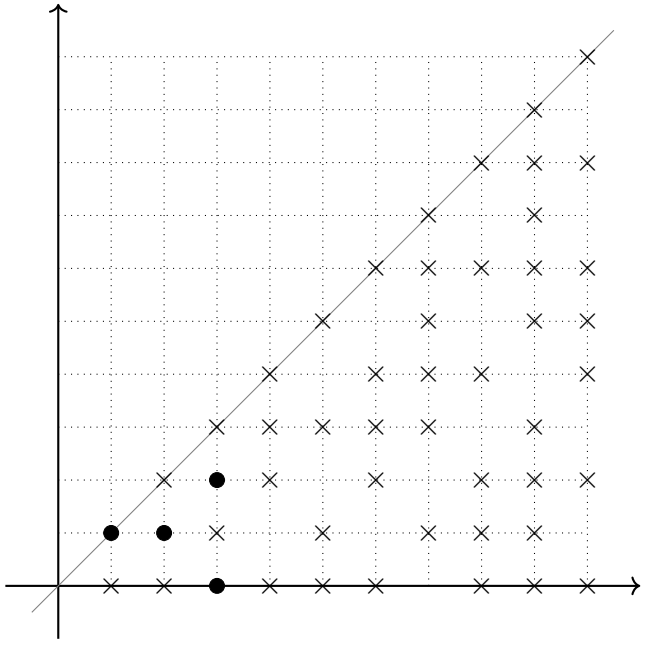

FIGURE 4. Gaussian Sieve after Step 4

The next remaining element of O_1 after $3+2i$ is $4+i$. We claim that if $x, y \in O_1$ are such that $4+i \preceq x \prec y \preceq 10+10i$ and x is a factor of y, then y must also have a factor $z \in \widetilde{O}_1$ such that $[z] \prec [x]$. The prime factors $\pi \in \widetilde{O}_1$ of z satisfy $[\pi] \preceq [z] \prec [x]$ by Lemma 3.8 so if we accept this claim, then all such composite y are already discarded. It follows that, in addition to declaring $4+i$ to be prime, we can also declare to be prime every $x \in O_1$ satisfying $4+i \prec x \preceq 10+10i$ that has not previously been discarded.

Let us prove the claim. First, it is straightforward to show that if $x, y \in O_1$, $y \preceq 10+10i$, with $y = zx$ for some $z \in \mathbb{Z}[i]$, then $z \in \widetilde{O}_1$. Suppose for the sake of contradiction that $y = zx$ where we do not have $[x] \preceq [z]$.

If $4+i \preceq w \in O_1$, then $w = a+bi$, where either $a = 4$ and $b \geq 1$, or $a > 4$. In the former case, $N(w) \geq 4^2+1 = 17$, while in the latter case, $N(w) \geq 5^2 = 25$. Thus, in either case, $N(w) \geq 17$. Taking $w = x$ and $w = z$ in this inequality gives

$$N(y) = N(z) \cdot N(x) \geq 17^2 > 200 = N(10+10i) \geq N(y)\,,$$

yielding a contradiction. Note that the last inequality follows readily from the inequality $y \preceq 10+10i$.

Now that the claim is proved, we see that we have the following complete list of Gaussian primes $\pi \in O_1$ satisfying $\pi \preceq 10+10i$:

$1+i$, $2+i$, 3, $3+2i$, $4+i$, $5+2i$, $5+4i$, $6+i$, $6+5i$, 7,
$7+2i$, $8+3i$, $8+5i$, $8+7i$, $9+4i$, $10+i$, $10+3i$, $10+7i$, $10+9i$.

The situation is thus as in Figure 5; since the process is now finished (for primes $\pi \preceq 10+10i$), we indicate only primes in this final diagram.

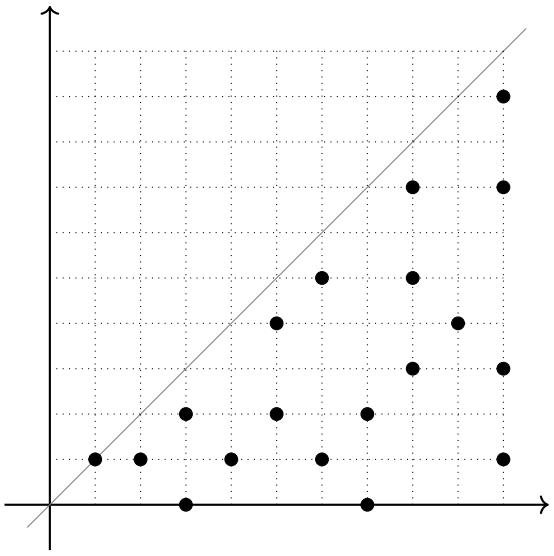

FIGURE 5. Gaussian primes in first octant

If we want to find all Gaussian primes, not just those in O_1, we simply multiply by units and take conjugates of the primes provided by our Gaussian Sieve. For instance, doing this for the primes in Figure 5 gives Figure 6. Axes of symmetry are added to emphasise the symmetry.

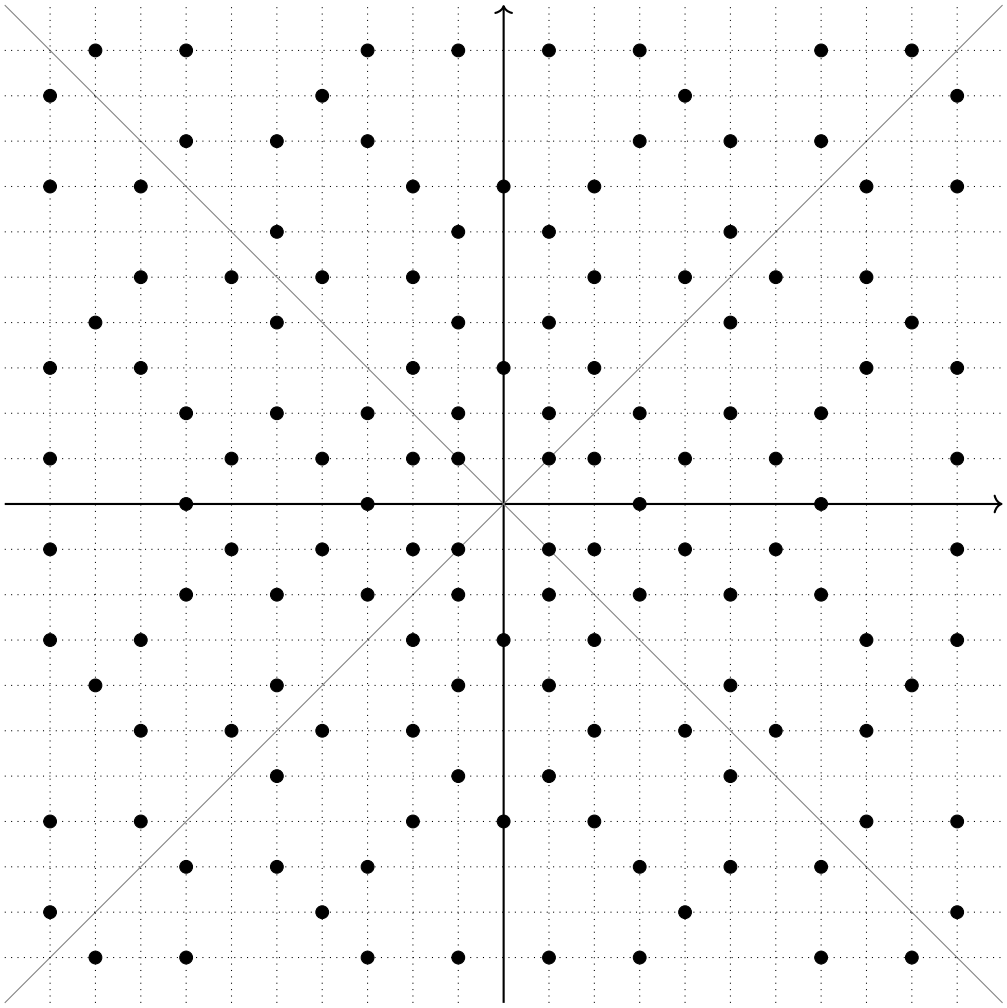

FIGURE 6. Gaussian primes

Having completed the sieving process for $x \in O_1 \prec 10+10i$, let us discuss some details of the process for $x \prec M + Mi$ for general $M \in \mathbb{N}$, $M > 1$, that are not completely clear from the above special case.

First of all, there is the question of stopping conditions for the main set of steps of the sieving process, i.e. when can we declare that we have gone far enough and all remaining $x \in O_1$, $x \preceq M + Mi$ are prime?

As in the special case $M = 10$ examined above, if a number $x \in O_1$, $x \preceq M + Mi$, is composite, it must have a prime factor π satisfying $N(\pi) \leq M\sqrt{2} = \sqrt{N(M + Mi)}$. Thus, we need to discard only multiples of such numbers. This means that we need to consider $x = a + bi \in O_1$ only for $a^2 \leq M\sqrt{2}$ and, for fixed $a > 1$, we need to consider b only for $b \leq \min\{a - 1, \sqrt{M\sqrt{2} - a^2}\}$. (We never need to consider $b = a$ since $a + ai$ is already discarded as a multiple of $1 + i$.) Thus, the required sieving region is roughly as shaded in Figure 7.

Given a fixed prime $\pi = a + bi$ found in the sieving process, we now need to discuss what multiples of $\tau \in \{\pi, \bar{\pi}\}$ need to be discarded. In both cases, we should discard

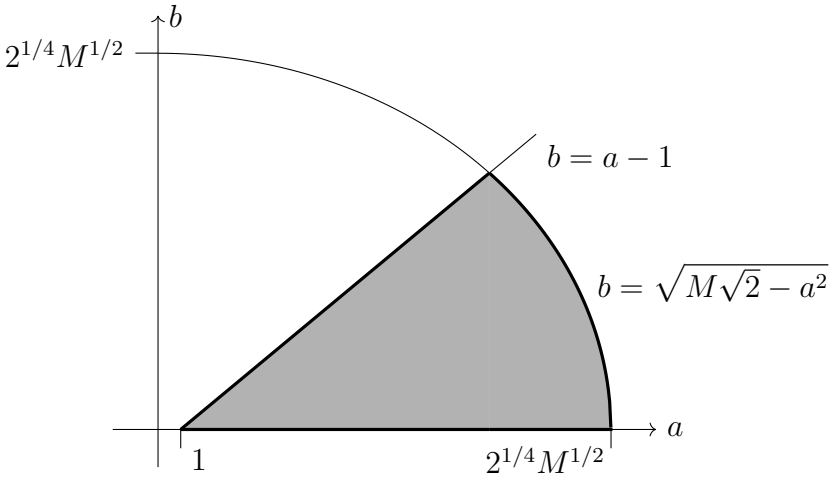

FIGURE 7. Sieving region

$w = z\tau$ for every $z \in \mathbb{Z}[i]$ that produces a product $w \in O_1$, $w \preceq M + Mi$, *unless we know that it has already been discarded.* By considering how the arguments of z, τ, and $w \in O_1$ relate to each other, we see that

- if $\tau = \pi$, then z must lie in $\widetilde{O}_1$,
- if $\tau = \bar{\pi}$ with $b > 0$, then z must lie in the first quadrant with $d > 0$.

Throughout our analysis, we write $z = c+di$ and $w = e+fi$, so that either $e = ac-bd$ and $f = ad + bc$ (if $\tau = \pi$) or $e = ac + bd$ and $f = ad - bc$ (if $\tau = \bar{\pi}$).

Let us first consider the special case $\pi = 1 + i$. In this case, $e = d + c$ and $f = d - c$, so we need to discard every $e + fi$ where $2 \le e \le M$, $0 \le f \le e$, and $e - f$ is even. Note in particular that we discard $e + ei$ for all $2 \le e \le M$. From now on, we assume that $\pi \succ 1 + i$, and so $N(\pi) \ge N(2 + i) = 5$.

We implement the deletion of multiples of τ in computer code using an outer loop over appropriate $c \in \mathbb{Z}$ and an inner loop (for fixed c) over appropriate $d \in \mathbb{Z}$. The problem now is to choose the definitions of *appropriate* in this last sentence so as to minimise the amount of calculations involved. More explicitly, we consider multipliers $z = c + di$ for all $c_{\min} \le c \le c_{\max}$ and $d_{\min} \le d \le d_{\max}$ where $c_{\min}$ and $c_{\max}$ depend only on τ, but $d_{\min}$ and $d_{\max}$ can additionally depend on c.

Below, we consider possible computational efficiencies that involve increasing $c_{\min}$ and $d_{\min}$ and decreasing $c_{\max}$ and $d_{\max}$. There are other efficiencies that we do not discuss because they involve coding optimizations rather than mathematical optimizations. For instance, computers can add and subtract faster than they can multiply so, once we have computed $w = e+fi$ for $\tau = \pi$ and a given $z = c+di$ and we wish to compute the new w for $z = c + (d + 1)i$ in the next pass of the inner loop, it is more efficient to subtract b from e and add a to f instead of carrying out another multiplication in our code.

Consider first $\tau = \pi$. We first consider the outer loop, i.e. the values of c that we need to handle. Since $z \in \widetilde{O}_1$, we certainly have $c \in \mathbb{N}$. Moreover, $N(w) = N(z)N(\pi)$, $N(z) \geq c^2$, and $N(x) \leq 2M^2$ for all $x \in O_1$, $x \preceq M + Mi$, so it would certainly suffice to take $c_{\min} = 1$ and $c_{\max} = \lfloor M\sqrt{2/N(\pi)} \rfloor$. However, we can do better than this, at least for $c_{\min}$. If $[z] \prec [\pi]$, then Lemma 3.8 ensures that w is a multiple of some prime or its conjugate that was considered at an earlier step. Thus, we can take $c_{\min} = a$ rather than $c_{\min} = 1$.

We now consider the inner loop. As already mentioned, we must have $z \in \widetilde{O}_1$, and so we could take $d_{\min} = -c$ and $d_{\max} = c$. However, we can do better than this for both $d_{\min}$ and $d_{\max}$.

Since $w \in O_1$, we must certainly have $-\arg(\pi) \leq \arg(z)$. Thus, instead of taking $d_{\min} = -c$, we can take $d_{\min} = -\lfloor cb/a \rfloor$. We then loop through successive values of d and calculate the associated products w until we reach a point where $f \geq e$, at which point we can stop because any further increase in b decreases e and increases f, taking w outside O_1.

We have not explicitly written down a new $d_{\max}$, but our stopping condition defines it implicitly. By considering the equation $\arg(w) = \arg(z) + \arg(\pi)$, we see that this stopping condition is reached before d reaches c, i.e. the new (implicitly defined) $d_{\max}$ is strictly less than the initial inefficient choice $d_{\max} = c$.

Other minor improvements could be made. For instance, since we only need to consider z with $[\pi] \preceq [z]$, when $c = a$, it suffices to consider $d = -b$ and $d \geq b$. To see this, note that in our discussion of the inner loop, we said that we could start with $d = -\lfloor cb/a \rfloor$, which here means starting with $d = -b$. We do not need to consider values $-b < d < b$ because then $[z] \prec [\pi]$. (The number of multipliers that each of the previous improvements removed from consideration grows quadratically in M, but the savings here grow only linearly in M, which is why we described it as a minor improvement.)

The minor improvement above is part of the reason why we did not discard any point in Step 4 of the sieving process for $M = 10$. In this step, we were considering multiples of $\pi = 3 + 2i$. Since $\pi\bar{\pi} = 13 \succ 10$, we only need to consider $z = 3 + di$ for $d \geq 2$ and $z = c + di$ for $c \geq 4$. In the former case, we get a product $e + fi$ with $f > e$, so we have already reached the stopping condition in the inner loop. We do not need to consider the latter case because of the stopping condition for the outer loop, i.e. because $M\sqrt{2/N(\pi)} = 10\sqrt{2/13} < 4$. (This is only a partial explanation of the absence of any points discarded in this step because it relates only to multiples of π. We have yet to consider multiples of $\bar{\pi}$, but a very similar analysis applies there, as we will see.)

Let us now discuss how to efficiently deal with multiples of $\bar{\pi} = a - bi$ where, without loss of generality, $b > 0$. For the outer loop, we can, as for $\tau = \pi$, take $c_{\min} = a$ and $c_{\max} = \lfloor M\sqrt{2/N(\pi)} \rfloor$.

As for the inner loop, recall that the multiplier $z = c + di$ must lie in the first quadrant with $d > 0$. Since we need $f = ad - bc \geq 0$, we can choose $d_{\min} = \lceil cb/a \rceil$. As we

increment d, each of e, f, and $\arg(z)$ increases, and $\arg(w)$ also increases. Thus, we can stop when either $f \geq e$ or $\min\{e, f\} > M$.

4. Eisenstein Integers

Definition 4.1. The set of Eisenstein Integers is $\mathbb{Z}[\omega] = \{a + b\omega \mid a, b \in \mathbb{Z}\}$, where

$$\omega = \frac{-1 + i\sqrt{3}}{2}$$

is a primitive cube root of unity.

It is easily verified that the (complex) conjugate of ω is ω^2. It follows that the conjugate of $z = a + b\omega \in \mathbb{Z}[\omega]$ is $\bar{z} = a + b\omega^2$, $a, b \in \mathbb{Z}$.

When writing $z \in \mathbb{Z}[\omega]$ in "expanded" form, we can choose between the $a + b\omega$ form and the $a + b\omega^2$ form. These are equivalent since $\omega^2 = -\omega - 1$. For either form, we employ the implicit convention that $a, b \in \mathbb{Z}$. We mostly use the $a + b\omega$ form, but the equation $a + b\omega^2 = \overline{a + b\omega}$ makes the $a + b\omega^2$ form useful for sieving in Section 4.1.

Addition and multiplication in the ring of Eisenstein integers $\mathbb{Z}[\omega]$ are inherited from $\mathbb{C}$. It follows that if $a + b\omega, c + d\omega \in \mathbb{Z}[\omega]$, then

- $(a + b\omega) + (c + d\omega) = (a + c) + (b + d)\omega$, and
- $(a + b\omega)(c + d\omega) = (ac - bd) + (bc + ad - bd)\omega$

A *norm of the Eisenstein integers*, $N : \mathbb{Z}[\omega] \to \mathbb{Z}_{\geq 0}$, is defined by

$$N(z) = z\bar{z} = a^2 - ab + b^2 = \frac{3a^2}{4} + \left(b - \frac{a}{2}\right)^2 .$$

whenever $z = a+b\omega$. Throughout this chapter, we will use without comment whichever of these forms of the norm is most convenient.

Using the form $N(z) = z\bar{z}$, the following proposition follows readily.

Proposition 4.2. *The norm $N : \mathbb{Z}[\omega] \to \mathbb{Z}_{\geq 0}$ is completely multiplicative (as implicitly defined in Proposition 3.2).*

Proposition 4.3. *The following are equivalent for $u \in \mathbb{Z}[\omega]$.*

(a) u is a unit.
(b) $N(u) = 1$.
(c) $u \in \{\pm 1, \pm\omega, \pm\omega^2\}$.

Proof. The fact that (c) implies (b) is trivial, and we also omit the easy proof of equivalence of (a) and (b) (which in any case is very similar to the corresponding equivalence in Proposition 3.3).

Finally, we prove that (b) implies (c). Let $u = a + b\omega \in \mathbb{Z}[\omega]$ be such that $N(u) = 1$, i.e.

$$\frac{3a^2}{4} + \left(b - \frac{a}{2}\right)^2 = 1 .$$

Multiplying both sides by 4 gives us:

$$3a^2 + (2b - a)^2 = 4 .$$

The solutions to this equation are:

(i) $a^2 = 1$ and $(2b - a)^2 = 1$, or
(ii) $a^2 = 0$ and $(2b - a)^2 = 4$.

Case (i) implies that either

- $a = 1$ and $b \in \{0, 1\}$, or
- $a = -1$ and $b \in \{-1, 0\}$.

Case (ii) implies that $a = 0$ and $b = \pm 1$.
Hence, the units in $\mathbb{Z}[\omega]$ are $\pm 1, \pm\omega$ and $\pm\omega^2$. □

Throughout this chapter, a *prime* (or *Eisenstein prime* for emphasis) refers to a prime in $\mathbb{Z}[\omega]$. We will refer to primes in $\mathbb{Z}$ as $\mathbb{Z}$*-primes*.

Corollary 4.4. *If $z \in \mathbb{Z}[\omega]$ is such that $N(z)$ is a $\mathbb{Z}$-prime, then z is an Eisenstein prime.*

We omit the proof of this corollary, as it is analogous to the proof of Corollary 3.4.

As an example of the above corollary, note that $3+2\omega$ is prime because $N(3+2\omega) = 7$ is prime. However, the converse is not true. For example, $N(11) = 121$ is not a $\mathbb{Z}$-prime but 11 is an Eisenstein prime (as follows for instance from Theorem 4.6 below).

We next show that the norm provides us with a Euclidean function for $\mathbb{Z}[\omega]$.

Proposition 4.5. *The ring of Eisenstein integers $\mathbb{Z}[\omega]$ is a Euclidean domain with Euclidean function $\delta : \mathbb{Z}[\omega] \setminus \{0\} \to \mathbb{N}$ given by $\delta(a + b\omega) = a^2 - ab + b^2$.*

Proof. To show that there is a division algorithm, we let $a + b\omega, c + d\omega \in \mathbb{Z}[\omega]$ and we work inside $\mathbb{Q}[\omega] = \{\alpha + \beta\omega \mid \alpha, \beta \in \mathbb{Q}\}$.

$$\begin{aligned}\frac{a + b\omega}{c + d\omega} &= \frac{a + b\omega}{c + d\omega} \cdot \frac{c + d\omega^2}{c + d\omega^2} = \frac{(ac - ad + bd) + (bc - ad)\omega}{c^2 - cd + d^2} \\ &= \alpha + \beta\omega\,, \quad \text{where } \alpha, \beta \in \mathbb{Q}\,.\end{aligned}$$

We can choose $n, m \in \mathbb{Z}$ such that $|\alpha - n| \leq \frac{1}{2}$ and $|\beta - m| \leq \frac{1}{2}$. Now, we get

$$\frac{a + b\omega}{c + d\omega} = n + \omega m + ((\alpha - n) + (\beta - m)\omega)\,.$$

Let $n + m\omega = q$ and $(\alpha - n) + (\beta - m)\omega = \gamma$. Then

$$\frac{a + b\omega}{c + d\omega} = q + \gamma\,.$$

Multiplying across by $(c + d\omega)$ gives

$$a + b\omega = q(c + d\omega) + \gamma(c + d\omega)\,.$$

Now, $\gamma(c + d\omega) = (a + b\omega) - q(c + d\omega)$. The right-hand side, $(a + b\omega) - q(c + d\omega)$, lies in $\mathbb{Z}[\omega]$, and therefore the left-hand side, $\gamma(c + d\omega)$ also lies in $\mathbb{Z}[\omega]$.

Let $\gamma(c + d\omega) = r$. We get

$$\delta(r) = \delta(\gamma(c + d\omega)) = \delta(\gamma)\delta(c + d\omega)\,,$$

and as

$$\delta(\gamma) = (\alpha - n)^2 - (\alpha - n)(\beta - m) + (\beta - m)^2 \leq \frac{3}{4} < 1\,,$$

we get $\delta(r) < \delta(c + d\omega)$. So, given $a + b\omega, c + d\omega \in \mathbb{Z}[\omega]$, there exist $q, r \in \mathbb{Z}[\omega]$ such that $a + b\omega = q(c + d\omega) + r$ and either $r = 0$ or $r \neq 0$, $\delta(r) < \delta(c + d\omega)$. □

As $\mathbb{Z}[\omega]$ is a Euclidean domain, an Eisenstein integer is irreducible if and only if it is prime. An associate of an Eisenstein prime is also an Eisenstein prime, as is the conjugate of an Eisenstein prime.

Theorem 4.6. *Every Eisenstein prime is a factor of a $\mathbb{Z}$-prime. Moreover, every $\mathbb{Z}$-prime p has one of the following $\mathbb{Z}[\omega]$-factorizations.*

- *If $p = 3$, $3 = (2+\omega)(2+\omega^2)$ is the product of two associate Eisenstein primes.*
- *If $p \equiv 2 \pmod 3$, then p is an Eisenstein prime.*
- *If $p \equiv 1 \pmod 3$, then p is a product of two conjugate non-associate Eisenstein primes $\pi\bar{\pi}$.*

Proof. Since $\mathbb{Z} \subset \mathbb{Z}[\omega]$, the first statement of the theorem is obvious. Thus, to find all Eisenstein primes, it suffices to find the $\mathbb{Z}[\omega]$-prime factorization of all $\mathbb{Z}$-primes.

Let p be a $\mathbb{Z}$-prime. Then $N(p) = p^2$, and so either p is an Eisenstein prime or it splits into two Eisenstein primes of norm p. If p splits, then $p = \pi\sigma$, where $\mathrm{Im}(\pi), \mathrm{Im}(\sigma) \neq 0$. Since

$$\pi\sigma = p = N(\pi) = \pi\bar{\pi}\,,$$

we have $\sigma = \bar{\pi}$. Thus, if p splits, then $p = \pi\bar{\pi}$.

Suppose $p = \pi\bar{\pi}$, where $\pi = a + b\omega$, and so $N(\pi) = a^2 - ab + b^2$. If either a or b is a multiple of 3, it is clear that $N(\pi)$ is equivalent to either 0 or 1 modulo 3. If $a = b = \pm 1 \bmod 3$, then $N(\pi) = 1 \pmod 3$. Finally, if one of a and b is equivalent to 1 modulo 3 and the other equivalent to -1 modulo 3, then $N(\pi) \equiv 0 \bmod 3$. Note that we never get $N(\pi) \equiv 2 \bmod 3$, and so we conclude that if $p \equiv 2 \bmod 3$, then p does not split and all such $\mathbb{Z}$-primes are Eisenstein primes.

Given that $N(\pi) = p$ is prime, the case $N(\pi) \equiv 0 \bmod 3$ can occur only for $p = 3$, in which case we have $3 = (1 + 2\omega)(1 + 2\omega^2)$. These two Eisenstein primes are associates since $-(1 + 2\omega^2) = 1 + 2\omega$.

Finally, it remains to consider the case $p \equiv 1 \bmod 3$. Suppose we can show that $(c + \omega)(c + \omega^2) = kp$, for $0 < c < p$, and some $k \in \mathbb{Z}$. If p does not split, then either $p \mid (c + \omega)$ or $p \mid (c + \omega^2)$, so $p \mid 1$, which is a contradiction. So we want to prove that there exists such a c.

Now, $N(c+\omega) = c^2 - c + 1 \equiv 0 \pmod p$. A solution to this in the field $\mathbb{Z}_p$ is formally given by the usual equation

$$c = \frac{1 \pm \sqrt{-3}}{2}\,,$$

assuming that this makes sense. The only reason that this equation might not make sense is if -3 does not have a square root.

By Euler's criterion and the Law of Quadratic Reciprocity,

$$\left(\frac{-3}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{3}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{3}\right)(-1)^{\frac{(p-1)(3-1)}{4}} = \left(\frac{p}{3}\right).$$

Clearly, $\left(\frac{p}{3}\right) = 1$ if $p \equiv 1 \pmod 3$. So there exists c such that $0 < c < p$ and $c^2 - c + 1 = kp$. Hence, p must split, i.e. $p = (a + b\omega)(a + b\omega^2)$. □

Every Eisenstein integer can be factorized into a product of Eisenstein primes. This factorization is unique up to the order of the factors, multiplication by units and the replacement of an Eisenstein prime by any of its associates.

4.1. Eisenstein Sieve.

We can generate an *Eisenstein Sieve* in a similar fashion to the Gaussian Sieve. We first define the *fundamental cone* $C_1 \subset \mathbb{Z}[\omega]$ to consist of all nonzero $z = a+b\omega \in \mathbb{Z}[\omega]$ such that $b \le a/2$. (Equivalently, $\arg(z) \in [0, \pi/6]$, but we prefer to reserve the symbol π to denote an Eisenstein prime.)

For the Eisenstein Sieve, the primary search region will be C_1, since all other Eisenstein integers can be generated by conjugation and/or multiplication by units in $\mathbb{Z}[\omega]$. If an Eisenstein integer in C_1 is prime, then its associates and their conjugates are also prime. (As in the Gaussian case, a *prime* will always refer to an Eisenstein prime, and we use $\mathbb{Z}$-prime to refer to a prime in $\mathbb{Z}$.)

Since there are six units, every Eisenstein prime has six associates (including itself), and these six each have their conjugates, giving twelve related primes in general (although, as in the Gaussian case, the total number halves in certain cases). Explicitly, given a prime $a+b\omega$, its associates are itself, $b+(b-a)\omega$, $b-a-a\omega$, and the negatives of these, and the conjugates are $a-b-b\omega$, $b+a\omega$, $a+(a-b)\omega$, and the negatives of these.

As for the Gaussian primes, we need to deal not only with the search region C_1, but also with the larger region $\widetilde{C}_1 := \{z, \bar{z} \mid z \in C_1\}$. We define an equivalence relation $\sim$ on $\widetilde{C}_1$ where $z \sim w$ if and only if $w \in \{z, \bar{z}\}$. We define the associated equivalence classes $[z] = \{z, \bar{z}\}$, $z \in \widetilde{C}_1$ and the set $\widetilde{C}_1/\sim$ of these equivalence classes.

As with the Gaussian integers, the first thing that we need to do is to place the elements of C_1 in a sequence by defining a useful total order $\prec$ on C_1. Specifically, we define $\prec$ to be lexicographic order on the "coordinates" (a, b) of $a+b\omega \in C_1$, i.e. $a+b\omega \prec c+d\omega$ if either $a < c$ or $a = c$ and $b < d$. We also denote by $\prec$ the inherited total order on $\widetilde{C}_1/\sim$: for $z, w \in C_1$, we write $[z] \prec [w]$ iff $z \prec w$. We define the relations $\preceq$ and $\succ$ on C_1 and on $\widetilde{C}_1/\sim$ in the obvious way.

It is clear that the two maps we call $\prec$ are total orders (on C_1 and on $\widetilde{C}_1/\sim$). In order for $\prec$ to be a *useful* total order, we need the following Eisenstein analogue of Lemma 3.8.

Lemma 4.7. *Suppose $x, y \in \widetilde{C}_1$ and $[x] \neq [y]$.*

(a) y and x are not associates.

(b) If y is divisible by x, then $[x] \prec [y]$.

Consequently, the Eisenstein prime factors $\pi \in \widetilde{C}_1$ of any $y \in \widetilde{C}_1$ all satisfy $[\pi] \preceq [y]$, with equality if and only if y is an Eisenstein prime.

PROOF. We first claim that

$$N(y) \leq \frac{4N(x)}{3}, \qquad \text{for all } x, y \in \widetilde{C}_1,\ [y] \preceq [x] \tag{4.8}$$

Since $N(y) = N(\bar{y})$, it suffices to prove this claim for $x, y \in C_1$. This follows rather easily from the fact that the minimum value of $a^2 - ab + b^2$ for fixed $a \in \mathbb{N}$ and $b \in \mathbb{R}$ satisfying $0 \leq b \leq a/2$ occurs at $b = a/2$ (which in turn follows readily by a variety of means, including calculus or plane geometry, so we omit the details).

With (4.8) in hand, the rest of the proof is as for Lemma 3.8. □

To carry out the Eisenstein sieve, we first discard the units 1 and $1 + \omega$. We declare the $\prec$-minimal remaining element of C_1, namely 2, to be prime, and we discard all its (Eisenstein) multiples $y \in C_1$. Then, we move on to the next Eisenstein integer and we repeat the process of deleting its multiples. (Above and later, terms such as *next* and *previous* are always defined with respect to $\prec$.) To carry this out in a practical setting (typically on a computer), we confine our work to a bounded initial segment of C_1 such as all $z \in C_1$ satisfying $z \preceq 2M + Mi$ for some $M \in \mathbb{N}$. We illustrate this process by carrying it out for $M = 5$.

Below, whenever we talk of discarding multiples of x, we mean that we discard from consideration all Eisenstein integer multiples $y \in C_1$ of x satisfying $x \prec y \preceq 10 + 5\omega$.

Example 4.9. We use the Eisenstein Sieve to find all primes $\pi \in C_1$, $\pi \preceq 10 + 5\omega$. In the diagrams associated with each step of the sieving process, the Eisenstein integers are the intersection points of dotted line segments. Each such segment consists of all points in C_1 of the form $n + t\omega$ or $m + t\omega^2$, as $t \in \mathbb{R}$ ranges over $\mathbb{R}$, and n or m is a fixed integer. Since an Eisenstein integer can be written in both forms, any Eisenstein integer in C_1 is obtained by intersecting two such line segments.

At each step, we mark the primes in C_1 that we have found so far by $\bullet$, we mark the units by $\times$, and we mark all associated (discarded) multiples of those primes by either $\otimes$ or $\times$, depending on whether this element was discarded at the current step or an earlier step, respectively. Eisenstein integers in C_1 whose primality status has yet to be decided after some step of the induction are shown as undecorated intersection points.

After discarding the unit 1, the first Eisenstein integer we find is 2. We discard all multiples multiples of 2 up to $10 + 10i$, as illustrated in Figure 8.

Now, we move on to $2+\omega$, the next Eisenstein integer in C_1 that has not been discarded. We declare $2 + \omega$ to be an (Eisenstein) prime and discard all multiples y of $2 + \omega$, as illustrated in Figure 9.

Let us note two points here. First, note that for this step, there is no need to discard multiples of the conjugate prime $2 + \omega^2$, because it is an associate of $2 + \omega$.

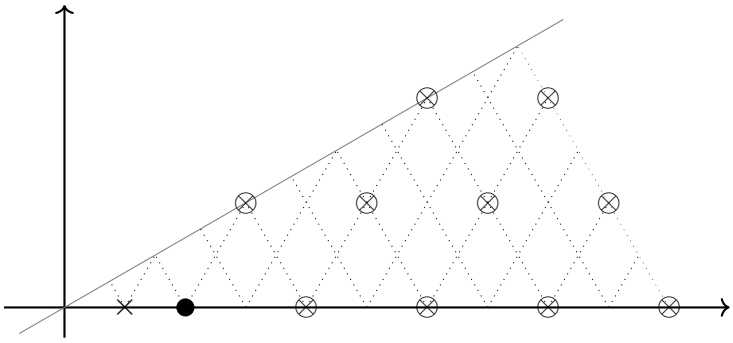

FIGURE 8. Eisenstein Sieve after Step 1

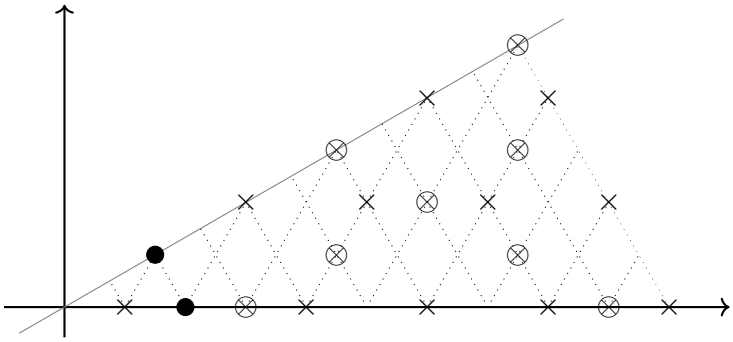

FIGURE 9. Eisenstein Sieve after Step 2

Second, as in the Gaussian case, it suffices to consider multipliers $a + b\omega$ such that $[2 + \omega] \preceq [a + b\omega]$. To do this with multipliers in $\widetilde{C}_1 \setminus C_1$, it is convenient to use the format $a + b\omega^2$ instead of $a + b\omega$ since it interacts better with $\prec$. For instance, if using the format $a + b\omega^2$, we can take $2 + \omega^2$ as the first multiplier. If using the other format, it would be a mistake to look only at multipliers of the form $a + b\omega$ for $a \geq 2$ because we would miss the multiple $(2 + \omega)(1 - \omega) = 3$.

We next declare $3+\omega$ to be prime and discard all multiples of $3+\omega$ and of its conjugate $3 + \omega^2$, as illustrated in Figure 10.

The next remaining number is $4+\omega$. As before, we declare this to be prime. However, we claim that in fact all remaining $x \in C_1$, $4 + \omega \prec x \prec 10 + 5\omega$ are prime.

To prove our claim, suppose that $y \in C_1$, $y \preceq 10 + 5\omega$ is composite. We first use (4.8) to deduce that $N(y) \leq 4N(10 + 5\omega)/3 = 100$. Suppose y has a prime factor $\pi = a + b\omega \in \widetilde{C}_1$, with $[4 + \omega] \preceq [\pi]$. Then, we also have $[4] \prec [\pi]$, and so we similarly deduce that $N(\pi) \geq 3N(4)/4 = 12$. Since 12 exceeds the square root of 100, y cannot have two such prime factors. Thus, y must have a prime factor τ such that $[\tau] \prec [4+\omega]$ and so it is already discarded at an earlier step. This proves the claim.

Thus, at the end of the sieving process, we have the following complete list of Eisenstein primes $\pi \in C_1$, $\pi \preceq 10 + 10\omega$:
$2, \quad 2 + \omega, \quad 3 + \omega, \quad 4 + \omega, \quad 5, \quad 5 + 2\omega, \; 6 + \omega,$
$7 + \omega, \quad 7 + 3\omega, \quad 9 + \omega, \quad 9 + 2\omega, \quad 9 + 4\omega, \quad 10 + 3\omega.$

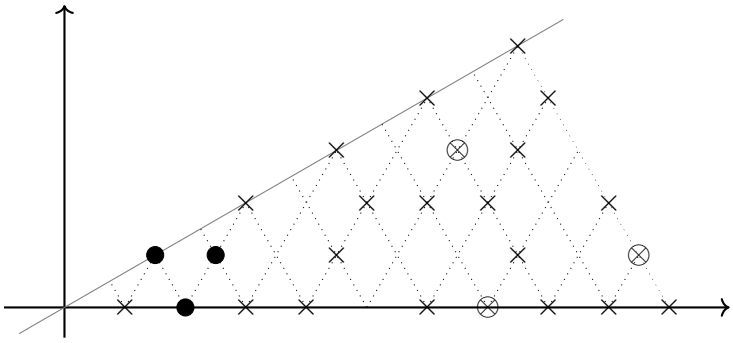

FIGURE 10. Eisenstein Sieve after Step 3

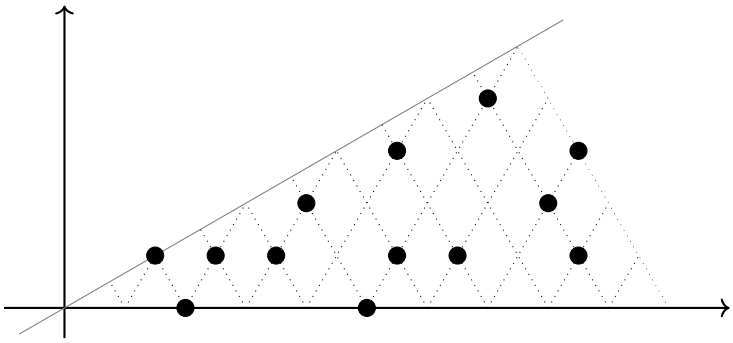

FIGURE 11. Eisenstein primes in the first octant

The situation is thus as in Figure 11; since the process is now finished (for primes $\pi \preceq 10 + 5\omega$), we indicate only primes in this final diagram.

If we want to find all Eisenstein primes, not just those in C_1, we simply multiply by units and take conjugates of the primes provided by our Eisenstein Sieve. For instance, doing this for the primes in Figure 11 gives Figure 12. Axes of symmetry are added to emphasise the symmetry.

Having completed the sieving process for $x \prec 10 + 5\omega$, let us discuss some details of the process for $x \prec 2M + M\omega$ for general $M \in \mathbb{N}$, $M > 1$.

First of all, there is the question of stopping conditions for the main steps of the sieving process, i.e. when can we declare that we have gone far enough and all remaining $x \in C_1$, $x \preceq 2M + M\omega$ are prime?

If a number $x \in C_1$, $x \preceq 2M + M\omega$, is composite, it must have a prime factor π satisfying

$$N(\pi) \leq \sqrt{N(x)} \leq \sqrt{\frac{4 \cdot N(2M + M\omega)}{3}} = 2M\,,$$

and so we need to discard only multiples of such numbers; note that the second inequality above follows from (4.8). Again by (4.8), if $\pi = a+b\omega \in C_1$ (and so $a \preceq \pi$), we have $N(\pi) \geq 3N(a)/4 = 3a^2/4$. This means that we need to examine only $x = a + b\omega \in C_1$

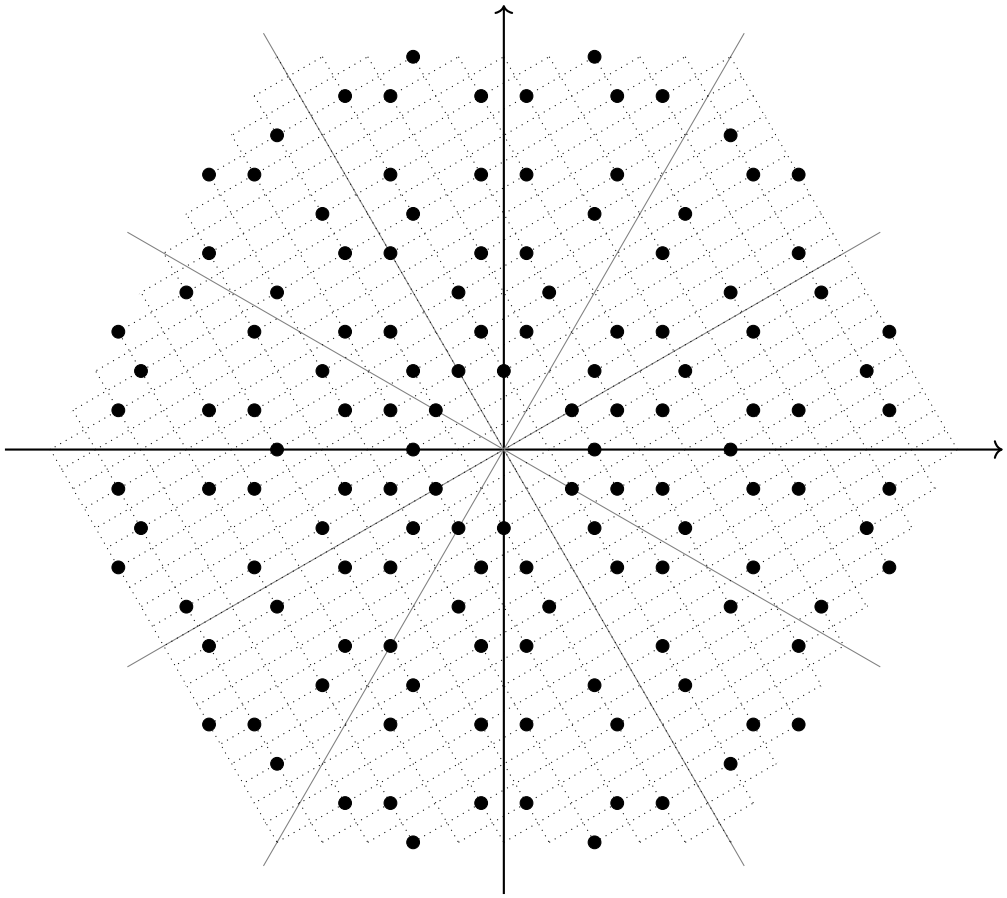

FIGURE 12. Eisenstein primes

where $3a^2/4 \leq 2M$, and so $a \leq 2\sqrt{2M/3}$. Also, for fixed $a > 0$, we need to consider b only for $\max\{a/2 - \sqrt{2M - 3a^2/4}, 0\} \leq b \leq a/2$.

Given a fixed prime $\pi = a + b\omega$ found in the sieving process, we now need to discuss what multiples of $\tau \in \{\pi, \bar{\pi}\}$ need to be discarded. In both cases, we should discard $w = z\tau$ for every $z \in \mathbb{Z}[\omega]$ that produces a product $w \in C_1$, $w \preceq 2M + M\omega$, *unless we know that it has already been discarded.* By considering how the arguments of z, τ, and $w \in C_1$ relate to each other, we see that

- if $\tau = \pi$, then z must lie in $\widetilde{C}_1$,
- if $\tau = \bar{\pi}$ with $b > 0$, then z must lie in the region consisting of all $c + d\omega$, $0 < d \leq c$.

Throughout our analysis, we write $z = c + d\omega$ (if $d > 0$) or $z = c - d\omega^2$ (if $d < 0$) and $w = e + f\omega$.

We implement the deletion of multiples of τ in computer code using an outer loop over appropriate $c \in \mathbb{Z}$ and an inner loop (for fixed c) over appropriate $d \in \mathbb{Z}$, and we would like to minimise the amount of calculations involved. More explicitly, we consider multipliers $z = c+d\omega$ (if $d \geq 0$) or $z = c-d\omega^2$ (if $d < 0$) for all $c_{\min} \leq c \leq c_{\max}$

and $d_{\min} \leq d \leq d_{\max}$ where $c_{\min}$ and $c_{\max}$ depend only on τ, but $d_{\min}$ and $d_{\max}$ can additionally depend on c.

Consider first $\tau = \pi$. Note that either $e = ac - bd$ and $f = bc + ad - bd$ if $d \geq 0$ or $e = ac - bd + ad$ and $f = bc + ad$ if $d < 0$. We first consider the outer loop, i.e. the values of c that we need to handle. Since $z \in \widetilde{C}_1$, we certainly have $c \in \mathbb{N}$. Moreover, $N(w) = N(z)N(\pi)$, $N(z) \geq 3c^2/4$, and $N(x) \leq 4M^2$ for all $x \in C_1$, $x \preceq 2M + M\omega$, so it would certainly suffice to take $c_{\min} = 1$ and $c_{\max} = \left\lfloor 2M/\sqrt{N(\pi)} \right\rfloor$. However, we can do better than this, at least for $c_{\min}$. If $[z] \prec [\pi]$, then Lemma 4.7 ensures that w is a multiple of some prime or its conjugate that was considered at an earlier step. Thus, we can take $c_{\min} = a$ rather than $c_{\min} = 1$.

We now consider the inner loop. As mentioned already, we must have $z \in \widetilde{C}_1$, and so we could take $d_{\min} = -\lfloor c/2 \rfloor$ and $d_{\max} = \lfloor c/2 \rfloor$. However, we can do better than this for both $d_{\min}$ and $d_{\max}$.

Since $w \in C_1$, we must certainly have $\arg(z) \geq -\arg(\pi)$. Thus, instead of taking $d_{\min} = -\lfloor c/2 \rfloor$, we can take $d_{\min} = -\lfloor cb/a \rfloor$. We then loop through successive values of d and calculate the associated products w until we reach a point where $f > e/2$. By considering the arguments of π, z, and w, it is clear that this point is reached for some d satisfying $0 \leq d \leq c/2$. We can stop because any further increase in d increases the arguments of both z and w, taking w further outside C_1.

Let us now discuss how to efficiently deal with multiples of $\bar{\pi} = a + b\omega^2$ where, without loss of generality, $b > 0$. Note that we now have $e = ac + bd - bc$ and $f = ad - bc$ (since we must have $d > 0$, or equivalently $\arg(z) > 0$, to ensure that $\arg(w) > 0$).

For the outer loop, we can, as for $\tau = \pi$, take $c_{\min} = a$ and $c_{\max} = \left\lfloor 2M/\sqrt{N(\pi)} \right\rfloor$.

As for the inner loop, recall that the multiplier $z = c + di$ must satisfy $0 < d < c$. Since we need $f = ad - bc \geq 0$, we can choose $d_{\min} = \lceil cb/a \rceil$. As we increment d, each of e, f, and $\arg(z)$ increases, and $\arg(w)$ also increases. Thus, we can stop when $f \geq e/2$, $e > 2M$, or $f > M$.

5. Number theory in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$

In this chapter, we discuss various number theoretic results for $\mathbb{Z}$ that are also true in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$. We often take R to be any one of these three rings. In fact, our presentation is valid in any ring R that has a norm $N : R \to \mathbb{Z}_{\geq 0}$ which provides a Euclidean function (by restricting to nonzero elements) and which is completely multiplicative (as implicitly defined in Proposition 3.2), but we prefer to keep the emphasis on these special rings because they are the ones that are important for us. As a standing assumption, $N(z) := z\bar{z}$ in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ and $N(z) := |z|$ in $\mathbb{Z}$.

We begin with a result that depends only on the fact that N provides us with a Euclidean function.

Theorem 5.1 (Division Theorem)**.** *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. For $a, b \in R$ with $b \neq 0$, there exist $q, r \in R$ such that $a = qb + r$ and $N(q) < N(b)$.*

The number q is the quotient, while the number r is the remainder. The remainder is bounded in size by the size of the divisor b, where the size of the numbers is determined by their norm.

We omit the obvious proof of the following proposition.

Proposition 5.2. *Suppose $a, b, c \in \mathbb{Z}$, with $c \neq 0$. The Gaussian integer $z = a + bi$ and the Eisenstein integer $z = a + b\omega$ are each divisible by c if and only if $c \mid a$ and $c \mid b$ in $\mathbb{Z}$.*

Letting $b = 0$ in Proposition 5.2 tells us that divisibility between ordinary integers does not change when working in $\mathbb{Z}[i]$ or $\mathbb{Z}[\omega]$, i.e.

$$c \mid a \text{ in } \mathbb{Z} \;\Leftrightarrow\; c \mid a \text{ in } \mathbb{Z}[i] \;\Leftrightarrow\; c \mid a \text{ in } \mathbb{Z}[\omega]\,, \qquad a, c \in \mathbb{Z}\,. \tag{5.3}$$

Because the norm N is completely multiplicative on each of our special rings R, and because $N(z) > 0$ for $z \neq 0$ (a consequence of the fact that $N|_{R\setminus\{0\}}$ is a Euclidean function), it is easy to prove the following result.

Proposition 5.4. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Then*

(a) if $a \in R$, then a is a unit if and only if $N(a) = 1$;
(b) if $a, b \in R$ and $b \mid a$ in R, then $N(b) \mid N(a)$ in $\mathbb{Z}$.

Definition 5.5. Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. For nonzero a and b in R, a greatest common divisor (GCD) of a and b is a common divisor with maximal norm.

For instance, in all three rings R, z is always a GCD of z and 0: certainly, it is a common divisor and, by Proposition 5.4, no common divisor can have a larger norm than z.

Note that if we have an equation $a = qb + r$ in R as above, then every GCD of a and b is also a GCD of b and r, and vice versa. This follows from the easily proven fact that any common divisor of a and b is also a common divisor of q and r, and vice

versa. (This paragraph and the previous one are key ingredients in the proof of the Euclidean algorithm later.)

Definition 5.6. Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let $a, b \in R$. We say a and b are coprime if their only common factors are units.

We now state the Euclidean algorithm. The proof for $\mathbb{Z}$ is well-known and that for $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ is very similar, so we do not include it.

Theorem 5.7 (Euclidean algorithm)**.** *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let $a, b \in R$ be nonzero, where we also write $r_{-1} = a$ and $r_0 = b$. Recursively apply the division theorem, starting with this pair, and make the divisor r_{i-1} and remainder r_i in the ith equation the new dividend and divisor, respectively, in the $(i+1)$st equation for as long as r_i is non-zero:*

$$\begin{aligned} a &= q_1 b + r_1, & N(r_1) < N(b) \\ b &= q_2 r_1 + r_2, & N(r_2) < N(r_1) \\ r_1 &= q_3 r_2 + r_3, & N(r_3) < N(r_2) \\ &\vdots \end{aligned}$$

We stop the process at the nth equation if the remainder r_n equals 0. The last nonzero remainder r_{n-1} is a greatest common divisor of a and b.

Example 5.8. In this example, we illustrate how to use the division algorithm on two Gaussian integers a and b to calculate their greatest common divisor. Let $a = 41 + 24i$ and $b = 11 - 2i$, then

$$\frac{41 + 24i}{11 - 2i} = \frac{(41 + 24i)(11 + 2i)}{11^2 + 2^2} = \frac{403 + 346i}{125},$$

where $3 < 403/125 < 7/2$ and $5/2 < 346/125 < 3$. Rounding the real and the imaginary parts to the closest integer, we get $3+3i$. $N(3+3i) = 18 < N(11-2i) = 125$. Hence, we get

$$41 + 24i = (3 + 3i)(11 - 2i) + (2 - 3i).$$

Now, using the division algorithm on $11 - 2i$ and $2 - 3i$ we get

$$\frac{11 - 2i}{2 - 3i} = \frac{(11 - 2i)(2 + 3i)}{2^2 + 3^2} = \frac{28 + 29i}{13}$$

where $2 < 28/13 < 5/2$ and $2 < 29/13 < 5/2$. So, rounding both fractions to the closest integer gives us $2 + 2i$, where $N(2 + 2i) = 8 < N(2 - 3i) = 13$. Hence

$$11 - 2i = (2 + 2i)(2 - 3i) + 1$$

and $2 - 3i = (2 - 3i) \cdot 1 + 0$.

We conclude that 1 is a GCD of $41 + 24i$ and $11 - 2i$ (as are -1 and $\pm i$). We have also shown that $41 + 24i$ and $11 - 2i$ are coprime.

Example 5.9. In this example, we illustrate how to use the division algorithm on two Eisenstein integers a and b to calculate their greatest common divisor. Let $a = 15+12\omega$ and $b = 5 + 3\omega$, then

$$\frac{15 + 12\omega}{5 + 3\omega} = \frac{(15 + 12\omega)(5 + 3\omega^2)}{5^2 - 5 \cdot 3 + 3^2} = \frac{66 + 15\omega}{19},$$

where $3 < 66/19 < 7/2$ and $1/2 < 15/19 < 1$. Rounding the real and the imaginary parts to the closest integer, we get $3 + \omega$. $N(3 + \omega) = 7 < N(5 + 3\omega) = 19$. Hence, we get

$$15 + 12\omega = (3 + \omega)(5 + 3\omega) + (3 + \omega).$$

Now, using the division algorithm on $5 + 3\omega$ and $3 + \omega$ we get

$$\frac{5 + 3\omega}{3 + \omega} = \frac{(5 + 3\omega)(3 + \omega^2)}{3^2 - 3 \cdot 1 + 1^2} = \frac{13 + 4\omega}{7}$$

where $3/2 < 13/7 < 2$ and $1/2 < 4/7 < 1$. So, rounding both fractions to the closest integer gives us $2 + \omega$, where $N(2 + \omega) = 3 < N(3 + \omega) = 7$. Hence

$$5 + 3\omega = (2 + \omega)(3 + \omega) + (-\omega)$$

and $3 + \omega = (2 + 3\omega) \cdot (-\omega)$.

We conclude that $-\omega$ is a GCD of $15 + 12\omega$ and $5 + 3\omega$ (as are ω, $\pm\omega^2$, and ± 1). We have also shown that $15 + 12\omega$ and $5 + 3\omega$ are coprime.

Note that above, we talk about "a" GCD rather than "the" GCD. This is because associates of any GCD are also GCDs, as we will show next. This is true also in $\mathbb{Z}$ if we defined GCD in a similar fashion. More explicitly, we could define $N(z) = |z|$ for all $z \in \mathbb{Z}$ to give a completely multiplicative norm whose restriction to $\mathbb{Z} \setminus \{0\}$ is a Euclidean function. With the analogous definition of GCD in $\mathbb{Z}$, $-d$ is a GCD of two elements of $\mathbb{Z}$ if and only if d is a GCD of these same elements. However, in the case of $\mathbb{Z}$, we normally insist that the GCD is positive (as is implicit in this original use of the terminology), so this makes the GCD unique in $\mathbb{Z}$.

Proposition 5.10. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. For nonzero a and b in R, let d be a greatest common divisor produced by the Euclidean algorithm. The set of greatest common divisors of a and b is precisely the set of unit multiples of d.*

Proof. Let d' be a greatest common divisor of a and b in R. From the proof of the Euclidean algorithm, $d' \mid d$, as d' is a common divisor. Hence, we can write $d = d'c$, so

$$N(d) = N(d')N(c) \geq N(d').$$

Since d' is a greatest common divisor, its norm is maximal among the norms of common divisors, so the inequality $N(d) \geq N(d')$ has to be an equality. That implies $N(c) = 1$ which is equivalent to c being a unit . Thus, d and d' are unit multiples of each other.

Conversely, if d' is a unit multiple of d in R, then the completely multiplicative nature of N plus the fact that $N(u) = 1$ for every unit u implies that d' is also a greatest common divisor. □

Our next theorem is a form of *Bezout's identity* for each of our three rings.

Theorem 5.11. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let d be any greatest common divisor of two nonzero elements of R. Then d is a linear combination of a and b, i.e. it can be written in the form $xa + yb$ for some $x, y \in R$.*

PROOF. Writing d as a linear combination of a and b is unaffected by replacing d with a unit multiple. Thus, by Proposition 5.10, we need to prove the statement only for $d := r_{n-1}$, the greatest common divisor given by the Euclidean algorithm in Theorem 5.7.

If $b \mid a$ in R, then b is a GCD of a and b and it clearly can be written as a linear combination of a and b. This fact is also clear if $n = 2$, i.e. if $d = r_1$. From now on, we assume that $n > 2$ in Theorem 5.7).

We claim that d is a linear combination of a and b. To prove this, we proceed by "backward induction". We first have $r_{n-3} = q_{n-1}r_{n-2} + r_{n-1}$, so certainly $d = r_{n-1}$ is a linear combination of r_{n-3} and r_{n-2}. Suppose $d = x_k r_{k-1} + y_k r_k$ is a linear combination of r_{k-1} and r_k for some $1 \leq k \leq n-2$. Since $r_{k-2} = q_k r_{k-1} + r_k$, we deduce that

$$d = x_k r_{k-1} + y_k(r_{k-2} - q_k r_{k-1}) = y_k r_{k-2} + (x_k - y_k q_k) r_{k-1}\,,$$

and so d is also a linear combination of r_{k-2} and r_{k-1}. Continuing in this matter until we reach $k = 0$, our claim follows. □

Corollary 5.12. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Nonzero elements a and b of R are coprime if and only if we can write*

$$1 = xa + yb$$

for some $x, y \in R$.

PROOF. If $a, b \in R$ are coprime, then 1 is a greatest common divisor of a and b, thus $1 = ax + by$ for some $x, y \in R$ by Theorem 5.11. Conversely, if $1 = ax + by$ for some $x, y \in R$, then any common divisor of a and b is a divisor of 1, and thus a unit. Hence, this implies that a and b are coprime. □

In the next two examples, we extend the euclidean algorithm to write a GCD of specific a and b in $\mathbb{Z}[i]$ or $\mathbb{Z}[\omega]$ as a linear combination of a and b.

Example 5.13. In Example 5.8, we applied the division algorithm to $a = 41 + 24i$ and $b = 11 - 2i$ and got:

$$\begin{aligned} 41 + 24i &= (3 + 3i)(11 - 2i) + (2 - 3i) \\ 11 - 2i &= (2 + 2i)(2 - 3i) + 1 \\ 2 - 3i &= (2 - 3i) \cdot 1 + 0 \end{aligned}$$

Running the division algorithm backwards we get:

$$\begin{aligned} 1 &= (11 - 2i) - (2 + 2i)(2 - 3i) \\ &= (11 - 2i) - (2 + 2i)((41 + 24i) - (3 + 3i)(11 - 2i)) \\ &= (1 + 12i)(11 - 2i) - (2 + 2i)(41 + 24i) \end{aligned}$$

Example 5.14. In Example 5.9, we applied the division algorithm to $a = 15 + 12\omega$ and $b = 5 + 3\omega$ and got:

$$\begin{aligned} 15 + 12\omega &= (3 + \omega)(5 + 3\omega) + (3 + \omega) \\ 5 + 3\omega &= (2 + \omega)(3 + \omega) + (-\omega) \\ 3 + \omega &= (2 + 3\omega)(-\omega) + 0 \end{aligned}$$

Running the division algorithm backwards we get:

$$\begin{aligned} -\omega &= (5 + 3\omega) - (2 + \omega)(3 + \omega) \\ &= (5 + 3\omega) - (2 + \omega)((15 + 12\omega) - (3 + \omega)(5 + 3\omega)) \\ &= (6 + 4\omega)(5 + 3\omega) - (2 + \omega)(15 + 12\omega) \end{aligned}$$

Corollary 5.15. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. If $a \mid c$ and $b \mid c$ in R, where a and b are coprime, then $ab \mid c$.*

PROOF. Let $c = ka$ and $c = lb$ for some $k, l \in R$. Since a and b are coprime, we can solve the equation

$$1 = xa + yb$$

for some $x, y \in R$. Multiplying both sides of the equation by c, we get:

$$\begin{aligned} c &= xac + ybc \\ &= xalb + ybka \\ &= (xl + yk)ab\,. \end{aligned}$$

Thus, $ab \mid c$. □

Proposition 5.16. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. If $c \mid ab$ in R, and a and c are coprime, then $c \mid b$. Thus, if p is prime and $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

PROOF. As $c \mid ab$ we have $nc = ab$ for some $n \in R$. If a and c are coprime, then we know that there exist $x, y \in R$ such that $1 = xa + yc$. Multiplying both sides by b, we obtain

$$b = xab + ycb = xnc + ycb = c(xn + yb).$$

Since $(xn + yb)$ is an element of R, it follows that $c \mid b$.

The second part of the proposition follows from the fact that if $p \nmid a$ then a and p are coprime. □

We now state a version of the Chinese remainder theorem. Note that if m is a nonzero element of $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$, then x lies in (m), the ideal generated by m, if and only if $m \mid x$.

Theorem 5.17 (Chinese remainder theorem)**.** *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Suppose $m, n \in R$ are nonzero and coprime. The map $f : R/(mn) \to R/(m) \times R/(n)$ defined on cosets by $f(x + (mn)) = (x + (m), x + (n))$ is a ring isomorphism.*

PROOF. If $x+(mn) = y+(mn)$, then $x-y \in (mn)$, which means that mn divides $x-y$, and so both m and n divide $x-y$. It follows that $x+(m) = y+(m)$ and $x+(n) = y+(n)$, and so the map f is well-defined. The ring homomorphism properties are similarly easily established, so we omit the proof.

Next, we show that f is injective. Suppose $x+(m) = y+(m)$ and $x+(n) = y+(n)$ for some $x, y \in \mathbb{Z}$. Then $m \mid x-y$ and $n \mid x-y$. Since m and n are coprime, it follows from Corollary 5.15 that $mn \mid x-y$, and so $x+(mn) = y+(mn)$. This establishes injectivity.

Finally, we want to prove that f is surjective. Suppose we want to find $z \in R$ such that $z \in a+(m)$ and $z \in b+(n)$. Since m and n are coprime, Corollary 5.12 implies that there exist $x, y \in R$ such that $xm+yn = 1$. Let $z = yna + xmb$. Then

$$z = (1-xm)a + xmb = a + m(-xa+xb) \in a+(m)$$

and

$$z = yna + (1-yn)b = b + n(ya-yb) \in b+(n)\,,$$

so $f(z+(mn)) = (a+(m), b+(n))$, as required. □

We now define congruences in $R \in \{\mathbb{Z}, \mathbb{Z}[i], \mathbb{Z}[\omega]\}$. Note that our definition is consistent with the usual one on $\mathbb{Z}$ because $x \in (c)$ is equivalent to x being divisible by c.

Definition 5.18. Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let $a, b, c \in R$, $c \neq 0$. We say that a is congruent to b modulo c, denoted $a \equiv b \pmod{c}$, if $a-b \in (c)$. The congruence class of a mod c is $\{b \mid b \in R \text{ and } b \equiv a \pmod{c}\}$.

Note that if c and c' are associates, then $(c) = (c')$, so congruence modulo c and modulo c' give the same equivalence classes.

We next need to discuss polygons in preparation for a theorem that we will need. For us, a *polygonal curve* consists of a finite set of positive length line segments L_k, $1 \leq k \leq n$, in $\mathbb{R}^2$ such that the endpoints of each L_k are v_{k-1} and v_k, where $v_0 = v_n$. A *polygon* is the closed bounded subset of $\mathbb{R}^2$ consisting of a polygonal curve and the planar region bounded by this curve; we denote it by $\text{Poly}(v_1, \ldots, v_n)$ since specifying the vertices specifies the polygon. We call a polygon *simple* if the boundary line segments L_i and L_j meet only at v_k and then only if $i = j+1$ or $i = j-1$. We call a simple polygon P *strictly convex* if the line segment L between two distinct boundary points x, y of P is a subset of P and meets the boundary of P only at x and y, unless x, y lie on the same L_k. As is well known, this is equivalent to the interior angle at v_k—meaning the angle between L_{k-1} and L_k on the side of the interior—being strictly less than 180 degrees for all $1 \leq k \leq n$.

The following is a restricted version of a well-known theorem of Pick.

Theorem 5.19 (Pick's theorem: special case)**.** *Suppose S is a strictly convex polygon in $\mathbb{R}^2$, all of whose vertices lie in $\mathbb{Z}^2$. Then the area of S is $m + n/2 - 1$, where m is the number of elements of $\mathbb{Z}^2$ lying strictly inside P and n is the number of elements of $\mathbb{Z}^2$ lying on the boundary of P.*

PROOF. For any simple polygon P, we write m_P for the number of elements of $\mathbb{Z}^2$ lying strictly inside P and n_P for the number of elements of $\mathbb{Z}^2$ lying on the boundary of P. The *Pick function* is the function f with domain the set of all simple polygons P defined by $f(P) = \text{area}(P) - (m_P + n_P/2 - 1)$. We call a simple polygon P a *Pick polygon* if $f(P) = 0$. We are required to prove that all strictly convex polygons are Pick polygons.

The basic idea of the proof is that we first prove that certain building blocks are Pick polygons, and then we prove that if we take some Pick polygons and cut or glue them in certain ways to create new polygons, then the new polygons are also Pick polygons. Finally, we show that such cutting and gluing allows us to create a general strictly convex polynomial from our building blocks.

Our building blocks are rectangles R with sides parallel to the coordinate directions. Suppose R is a rectangle with vertices (x, y), $(x + a, y)$, $(x + a, y + b)$, and $(x, y + b)$ for some $x, y \in \mathbb{Z}$ and $a, b \in \mathbb{N}$. The number of interior points m is $(a - 1)(b - 1)$ and the number n of boundary points is $2(a - 1) + 2(b - 1) + 4 = 2a + 2b$, so

$$m + \frac{n}{2} - 1 = (a - 1)(b - 1) + a + b - 1 = ab = \text{area}(R)\,,$$

showing that R is a Pick polygon.

The gluing procedure involves two simple polygons P_1 and P_2 that have a common side S but that have no other points in common. Then $P_0 := P_1 \cup P_2$ is another simple polygon. Suppose the number of interior and boundary points in P_i, $i \in \{0, 1, 2\}$, are m_i and n_i, respectively, and suppose that there are k points of $\mathbb{Z}^2$ on S. The endpoints of S are vertices of P_i, $i \in \{0, 1, 2\}$. All other points of S are boundary points of P_1 and P_2 but become interior points of P_0. Thus, $m_0 = m_1 + m_2 + k - 2$ and $n_0 = n_1 + n_2 - 2k + 2$ and so

$$m_0 + n_0/2 - 1 = (m_1 + m_2) + (n_1 + n_2)/2 - 2 = (m_1 + n_1/2 - 1) + (m_2 + n_2/2 - 1)\,.$$

Since also $\text{area}(P_0) = \text{area}(P_1) + \text{area}(P_2)$, we see that $f(P_0) = f(P_1) + f(P_2)$; we call this last equation the *gluing equation*. The gluing procedure can be iterated to get an equation of type $f(P_0) = \sum_{i=1}^{k} f(P_k)$ whenever repeated gluing of polygons P_i, $1 \le i \le k$, in the manner described above, eventually creates a new polygon P_0. We call this last equation the *extended gluing equation*. The extended gluing equation implies in particular that if a polygon P_0 is formed by gluing a set of Pick polygons together, it is also a Pick polygon.

Next, we prove that T is a Pick polygon if T is a right-angled triangle whose non-hypotenuse sides are in the coordinate directions. Note that any such T forms half of a rectangle R with sides parallel to the coordinate directions. The hypotenuse of T is a diagonal of R and the image of T under a central symmetry through the midpoint of the hypotenuse of T is a congruent triangle T' with the same hypotenuse such that $R := T \cup T'$. Furthermore, the same central symmetry sets up a 1-1 correspondence between the elements of $\mathbb{Z}^2$ in T and those in T', with interior points corresponding to interior points and boundary points to boundary points. Thus, we have $m_{T'} = m_T$ and $n_{T'} = n_T$, and of course $\text{area}(T') = \text{area}(T)$, so $f(T) = f(T')$. By the gluing

equation for $(P_0, P_1, P_2) = (R, T, T')$, we see that $0 = f(R) = f(T) + f(T') = 2f(T)$, so T is also a Pick polygon.

Next, suppose T is a general triangle with vertices in $\mathbb{Z}^2$. The minimum and maximum values of the x-coordinates and of the y-coordinates provide us with four numbers and, since there are only three vertices, two of these numbers must be associated with a single vertex, in the sense that this vertex takes on an extreme value in both coordinates among the values given by the three vertices.

If we reflect through the first or second coordinate axis, we swap the notions of maximum and minimum for the other coordinate while keeping the numbers of interior and boundary points and the area fixed (and so the reflected triangle is Pick if and only if the original triangle is Pick). Thus, we can assume without loss of generality that a fixed vertex u of T minimises both coordinate values among the set of vertices.

There are now two cases. Case 1 is where one of the other vertices, v, maximizes the first coordinate while the final vertex w maximizes the second coordinate. As illustrated in Figure 13, we can adjoin triangles T_1, T_2, and T_3 to T to get a rectangle R. (One or more of these adjoined triangles might be missing if one or more sides of T are in coordinate directions.) In any case, since the adjoined triangles and the rectangle are all Pick polygons, it follows from the extended gluing equation that T is also a Pick polygon.

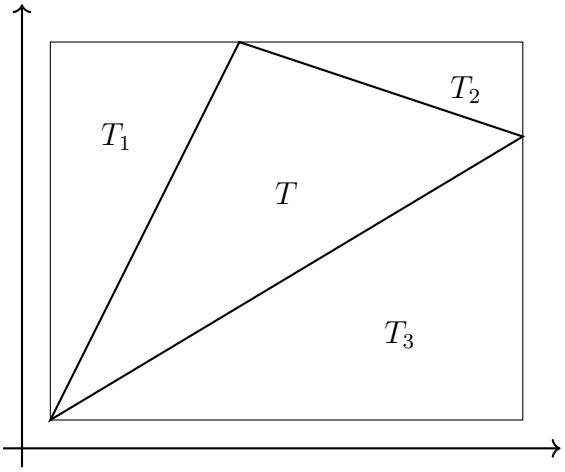

FIGURE 13. General triangle T: Case 1

Case 2 is where one of the other coordinates, v, maximizes both of the other coordinates. As illustrated in Figure 14, we can again adjoin right-angled triangles T_i, $1 \leq i \leq 5$, with non-hypotenuse sides in the coordinate directions to get a rectangle R. We then deduce as in Case 1 that T is a Pick polygon.

Finally, we prove by induction on the number n of vertices that a strictly convex polygon is Pick. We have shown it to be true for a triangle, so suppose $P := \text{Poly}(v_1, \ldots, v_n)$ is a strictly convex polygon for some $n > 3$, and that every strictly convex polygon with strictly fewer than n vertices is Pick. By strict convexity, drawing a line segment from v_1 to v_3 allows us to cut P into two subpolygons: $P_1 := \text{Poly}(v_1, v_2, v_3)$ and

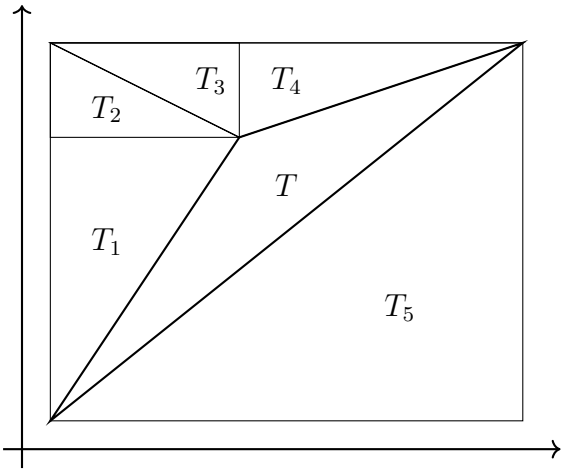

FIGURE 14. General triangle T: Case 2

$P_2 := \text{Poly}(v_1, v_3, v_4, \ldots, v_n)$. Then P_1 and P_2 are strictly convex polygons (since their interior angles at v_1 and v_3 are strictly less than the angles at the same points for P, and all other interior angles are the same as they are in P). Since P_1, P_2 have fewer than n vertices, they are Pick, and now P inherits the Pick property from $P1$ and P_2. □

Remark 5.20. In the terminology of the above proof, the full-strength Pick's theorem says that all simple polygons are Pick polygons. The above proof gives most of the proof of this full-strength result. We only need to show that a simple polygon with $n > 3$ vertices can always be cut into two simple polygons, each with strictly fewer than n vertices. This can be shown, but we will not give the details because we do not need this full-strength result.

The next theorem further illustrates the analogy between the norm in $\mathbb{Z}[i]$ and the absolute value in $\mathbb{Z}$, as we have $|n|$ distinct $\mathbb{Z}$-congruence classes modulo n for $n \in \mathbb{Z}$.

Theorem 5.21. *For $z \in \mathbb{Z}[i] \setminus \{0\}$, there are $N(z)$ distinct congruence classes modulo z.*

PROOF. The multiples of $z = a + ib$ are precisely all numbers of the form $(c + di)z = cz + d(iz)$, i.e. the linear combinations of z and iz. If we join neighbouring multiples with line segments, we get a square tiling of the plane with sidelength $\sqrt{a^2 + b^2} = \sqrt{N(a)}$, and so this square has area $N(z)$. (By a neighbouring multiple, we mean that we change $c + di$ to $c' + d'i$, where $(c', d') \in \{(c \pm 1, d), (c, d \pm 1)\}$.)

This tiling allows for a geometric understanding of congruence classes. If we think of these squares as tiles that can be shifted (or translated), then u and v are equivalent modulo z if and only if, when we shift the square containing u directly on top of the square containing v, u lies directly on top of v. In particular, the set of congruence classes corresponds, mostly in a 1-1 fashion, with the subset of $\mathbb{Z}[i]$ lying in one of these closed squares, which we call S. We say "mostly" above because the boundary points are exceptional: it is clear that the four vertices of S all correspond to the same

congruence class and all other $\mathbb{Z}[i]$ points lying on the boundary of S correspond in pairs to residue classes (since any two opposite sides of S give the same set of residue classes).

Thus, the number M of congruence classes is $m + (n-4)/2 + 1 = m + n/2 - 1$, where m is the number of $\mathbb{Z}[i]$ points in the interior of S and n is the number of lattice points on the boundary of S. By Pick's theorem for a square, we have shown that M is the area of S, which we know to be $N(z)$, and so we are done. □

When talking about congruence modulo a specific nonzero Gaussian integer c, we use the term *fundamental region* to mean a subset of $\mathbb{C}$ that contains exactly one representative from each congruence class. (It is the set $R \cap \mathbb{Z}[i]$ of congruence class representatives that really interests us, but we define the region R because it is often more easily specified than $R \cap \mathbb{Z}[i]$.)

Given $c \in \mathbb{Z}[i] \setminus \{0\}$, we now explicitly define an associated fundamental region R (dependent on c) and a method for *reducing* $a \in \mathbb{Z}[i]$ modulo c, meaning a method for finding $b \in \mathbb{Z}[i] \cap R$ such that a is equivalent to b modulo c. The term "reducing" is used because b has a smaller norm than most Gaussian integers b' that are equivalent to a modulo c. However, because we choose R to be simple geometrically, the reduced number b is often not the norm-minimising representative.

We distinguish between two cases:

- Case 1: c is an associate of a $\mathbb{Z}$-integer.
- Case 2: c is not an associate of a $\mathbb{Z}$-integer.

Case 1:
We can assume without loss of generality that $c > 0$. To find a b of small norm satisfying $a \equiv b \pmod{c}$, where a is a Gaussian integer and $c \neq 0$ is a $\mathbb{Z}$-integer, we simply find the remainders of the real and imaginary parts of a upon $\mathbb{Z}$-division by c, where both remainders have values between 0 and $|c| - 1$ inclusive. This gives $|c|^2 = N(c)$ representatives, so we have found a complete set of representatives of the congruence classes. Thus, the fundamental region is the "half-open" square $R = R_c = [0, c)^2 \subset \mathbb{R}^2$ (where we have identified $\mathbb{C}$ with $\mathbb{R}^2$ for notational brevity).

Example 5.22. In this example, we reduce $-50344 + 74730i \mod 437$.

$$-50344 \equiv 348 \quad \text{and} \quad 74730 \equiv 3 \pmod{437}.$$

Hence, $-50344 + 74730i \equiv 348 + 3i \pmod{437}$.

Case 2:
To define a fundamental region R for the set of congruence classes modulo a Gaussian integer c that is not an associate of an integer, it is first convenient to extend equivalence modulo c from $\mathbb{Z}[i]$ to $\mathbb{C}$ in the obvious way: $z, w \in \mathbb{C}$ are equivalent modulo c if $z - w$ is a Gaussian integer that is divisible by c. Below, *equivalent* means this extended notion of equivalence modulo c.

We define our fundamental region $R = R_c$ to be the "half-open" square with vertices at the origin, c, ic and $(1 + i)c$. To define what we mean by "half-open", we need

to specify what parts of the boundary are included in R. Specifically, R contains all points on the line segments from 0 to c and from 0 to ic except for c and ic. We do not include the other two boundary line segments, since any point on one of those line segments is equivalent to some point on a parallel boundary line segment that we have included. Note that 0 is the only included vertex; the other three are equivalent to it.[1]

The square R is a single tile in a regular square tiling of the plane that partitions $\mathbb{C}$ into tiles given by translates of R: the other tiles R' are precisely the translates by $(a+ib)c$ of R for some $a, b \in \mathbb{Z}$: to get from R to R', we need to hop a tiles in the direction given by c, and hop b tiles in the direction given by ic. Note that, in order to ensure that we get a partition, it is crucial that we included exactly two (non-parallel) sides and one vertex in R. It follows that R contains precisely one representative of each equivalence class modulo c, and so it is indeed a fundamental region.

It remains to give an algorithm for finding the element $w \in \mathbb{Z}[i] \cap R$ that is equivalent to a given $z \in \mathbb{Z}[i]$. We first define the *Gaussian floor function* $f : \mathbb{C} \to \mathbb{Z}[i]$ by $f(x+iy) = m+in$ for all $x, y \in \mathbb{R}$, where m and n are the greatest integers that are no larger than x and y, respectively. We then define $g : \mathbb{Z}[i] \to \mathbb{Z}[i]$ by $g(z) := z - c \cdot f(z/c)$.

It is clear that $g(z) - z$ is a multiple of c, and so $g(z)$ is equivalent to z. To show that g gives the required reduction modulo c, it therefore suffices to prove that $g(z) \in R$ for all $z \in \mathbb{Z}[i]$. First, note that $g(z)/c = w - f(w)$, where $w := z/c$. It is clear that for all $u \in \mathbb{C}$, $u - f(u)$ lies in the unit square $U := \{a + ib \in \mathbb{C} \mid 0 \leq a, b < 1\}$. Multiplication by c is a similarity map from U to R, so it follows that $g(z) \in R$, as required.

To emphasise that the square R in Case 2 does not have sides parallel to the real and imaginary axes, we will often refer to it as a *rhombus* or *rhombic region* below.

Observation 5.23. If c is a prime above (Case 1 or Case 2), then 0 is the only Gaussian integer that is an included boundary element of R. To see this, we first note that, if $a \in R$ is a nonzero Gaussian integer on the line segment between 0 and c, then it cannot be c. Now, a and c have a greatest common divisor b lying on the line segment from a to c: in fact, we can define $b = c \cdot \gcd(N(a), N(c))/N(c)$, where N is the norm on $\mathbb{Z}[i]$. Now, b is a divisor of c but is not an associate of it (since it has smaller norm) and b is not a unit (since it does not lie on the real or imaginary axes), so this contradicts the assumption that c is prime. We similarly rule out any nonzero Gaussian integer on the line segment from 0 to ic.

Example 5.24. Let us find congruence class representatives of $3+2i$. Our fundamental region R is the half-open square with vertices $0+0i$, $3+2i$, $-2+3i$ and $1+5i$. As always, $0 \in R$. However, since $3 + 2i$ is prime, all other congruence class representatives are in the interior of R. Visualising this, we get Figure 15, where we indicate all elements of $\mathbb{Z}[i] \cap R$ with solid dots, and some nearby elements of $\mathbb{Z}[i] \setminus R$ with open dots.

As indicated in Figure 15, the set $\mathbb{Z}[i] \cap R$ of congruence class representatives is:

$$\{0,\ i,\ 1+i,\ -1+2i,\ 2i,\ 1+2i,\ 2+2i,\ -1+3i,\ 3i,\ 1+3i,\ 2+3i,\ 4i,\ 1+4i\}\,.$$

[1]The difference in the way we defined $R = R_c$ in Cases 1 and 2 is solely due to the tools and notation available to us for defining R when $c \in \mathbb{Z}$. In actuality, regardless of whether c is handled by Case 1 or Case 2, every R_c is a member of the same similarity class of sets. In fact, if $c_1, c_2 \in \mathbb{Z}[i] \setminus \{0\}$, then multiplication by c_2/c_1 defines a similarity map from R_{c_1} to R_{c_2}.

Example 5.25. In this example, we reduce $-50344 + 74730i \mod 3 + 2i$. We start by computing $(-50344 + 74730i)/(3 + 2i)$:

$$\frac{-50344 + 74730i}{3 + 2i} = \frac{(-50344 + 74730i)(3 - 2i)}{13} = \frac{-1572 + 324878i}{13}$$

Since

$$\left\lfloor \frac{-1572}{13} \right\rfloor = -121 \qquad \text{and} \qquad \left\lfloor \frac{324878}{13} \right\rfloor = 24990\,,$$

we have $f((-50344 + 74730i)/(3 + 2i)) = -121 + 24900i$. We now calculate

$$g(-50344 + 74730i) = (-50344 + 74730i) - (3 + 2i)(-121 + 24990i) = -1 + 2i\,,$$

and so $-50344 + 74730i \equiv -1 + 2i \pmod{3 + 2i}$.

We next state the Eisenstein version of Theorem 5.21.

Theorem 5.26. *For $z \in \mathbb{Z}[\omega] \setminus \{0\}$, there are $N(z)$ distinct residue classes modulo z.*

PROOF. The multiples of $z = a + b\omega$ are precisely all numbers of the form $(c + d\omega)z = cz + d(z\omega)$, i.e. the linear combinations of z and $z\omega$. If, as in the Gaussian case, we join neighbouring multiples with line segments, we get a tiling of the plane by rhombuses with sidelength $\sqrt{a^2 - ab + b^2} = \sqrt{N(z)}$. Each rhombus is defined as a closed set, so neighboring ones intersect along a side.

Let us label as $R_{0,0}$ the tile with vertices 0, z, $z + z\omega$, and $z\omega$, and label as $R_{c,d}$ the tile obtained by shifting R by $cz + d(z\omega)$ for some $c, d \in \mathbb{Z}$. Since all the tiles are copies of $R_{0,0}$ that have been shifted by an element of $\mathbb{Z}[\omega]$, all tiles have the same number N_E of elements of $\mathbb{Z}[\omega]$.

Let N_C be the number of congruence classes modulo z. A set of representatives of the congruence classes corresponds, mostly in a 1-1 fashion, with the subset of $\mathbb{Z}[\omega]$ contained in any one of these tiles R. We say "mostly" because the four vertices of R are all in the same congruence class and opposite pairs of other boundary points give the same congruence class. Thus, $N_E > N_C$.

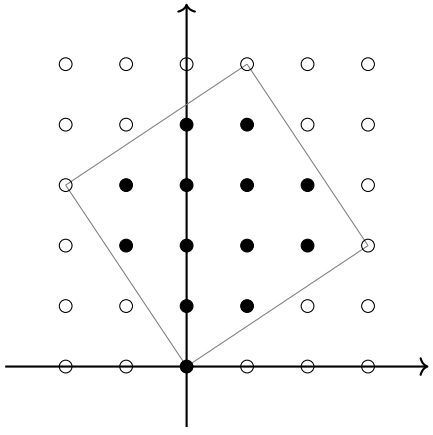

FIGURE 15. Fundamental region R for $3 + 2i$; $\mathbb{Z}[i] \cap R$ highlighted

We next define $R(M)$ for each $M \in \mathbb{N}$ to be the rhombic subset of $\mathbb{C}$ given by the union of the tiles $R_{c,d}$, $0 \leq c < M$, $0 \leq d < M$. Let $V(M)$ denote the number of elements of $\mathbb{Z}(\omega)$ in $R(M)$ divided by M^2. By definition, $V(1) = N_E > N_C$. For $M > 1$, we have $V(M) < N_E$ since points interior to $R(M)$ that are vertices of individual tiles are only counted once across the four tiles containing them, and all other points interior to $R(M)$ that are on the boundary of individual tiles are only counted once across the two tiles containing them. This lowered counting exactly matches how we should count congruence classes, and dividing by M^2 compensates for the repeated counting of the same congruence class on different tiles. Despite this, we do not have $V(M) = N_C$ because the correspondence between points in $R(M)$ and congruence classes goes wrong at the boundary of $R(M)$. However, because the length of the boundary of $R(M)$ is proportional to M rather than M^2, this error tends to 0 as $M \to \infty$. Thus, $V(M) \to N_C$ as $M \to \infty$.

We will next approximately count the number of elements of $\mathbb{Z}(\omega)$ in $R(M)$ using the area of $R(M)$. $R_{0,0}$ is a rhombus of sidelength $\sqrt{N(z)}$ and angle 60 degrees between one pair of neighboring sides, so its area is $\sqrt{3}N(z)/2$. Thus, the area of $R(M)$ is $M^2\sqrt{3}N(z)/2$.

We now define a *microtile* to be any rhombus of sidelength 1 that has one pair of sides parallel to the direction from 0 to 1 and the other pair of sides parallel to the direction from 0 to ω. Note that the area of a microtile is $\sqrt{3}/2$. Let us "roughly tile" $R(M)$ with microtiles. By this, we mean that we tile the whole plane with microtiles and then take only the microtiles that lie fully within $R(M)$. We insist that the tiling is arranged so that 0 is a vertex of one of the microtiles: this ensures that the vertices of the microtiles are precisely the elements of $\mathbb{Z}[\omega]$.

Comparing areas, we see that the number $T(M)$ of microtiles that we use to roughly tile $R(M)$ is at most $(M^2\sqrt{3}N(z)/2)/(\sqrt{3}/2) = M^2N(z)$. Writing $U(M) := T(M)/M^2$, we see that $U(M) \to N(z)$ as $M \to \infty$ because the area of $R(M)$ that is not covered by microtiles is at most proportional to M.

Let us assign each point of $\mathbb{Z}[\omega]$ to the microtile in this tiling for which it is the bottom right vertex: this sets up a 1-1 correspondence between elements of $\mathbb{Z}[\omega]$ and microtiles in the planar tiling. It follows that $V(M)$ as defined above is roughly equal to $U(M)$ where, as always, there is an error that is bounded by a multiple of $1/M$ in this estimate. Since $V(M) \to N_C$ and $U(M) \to N(z)$ as $M \to \infty$, we deduce that $N_C = N(z)$, as required. □

It is straightforward to tweak the above proof to produce a different proof of Theorem 5.21.

We now discuss how to *reduce* an Eisenstein integer a modulo some other Eisenstein integer $c \neq 0$, meaning how to find another Eisenstein integer b that is in the same congruence class (mod c) as a but which has small norm.

As in the case of Gaussian integers, for a given $c \in \mathbb{Z}[\omega]\backslash\{0\}$, we now explicitly define an associated fundamental region R (dependent on c) and a method for *reducing* $a \in \mathbb{Z}[\omega]$

modulo c, meaning a method for finding $b \in \mathbb{Z}[\omega] \cap R$ such that a is equivalent to b modulo c.

We distinguish between two cases:

- Case 1: c is an associate of a $\mathbb{Z}$-integer.
- Case 2: c is not an associate of a $\mathbb{Z}$-integer.

Case 1:
We can assume without loss of generality that $c > 0$. To find a b of small norm satisfying $a \equiv b \pmod{c}$, where a is an Eisenstein integer, we simply write $a = d + e\omega$, and find remainders of d and e upon $\mathbb{Z}$-division by c, where both remainders have values between 0 and $|c| - 1$ inclusive. This gives $|c|^2 = N(c)$ representatives, giving a complete set of representatives of the congruence classes. The fundamental region is the "half-open" square $R = R_c = [0, c)^2 \subset \mathbb{R}^2$ (where we again identify $\mathbb{C}$ with $\mathbb{R}^2$).

Example 5.27. In this example, we reduce $9348 + 3109\omega \pmod{314}$.

$$9348 \equiv 242 \pmod{314} \qquad \text{and} \qquad 3109 \equiv 283 \pmod{314}.$$

Hence, $9348 + 3109\omega \equiv 242 + 283\omega \pmod{314}$.

Case 2:
To find a b satisfying $a \equiv b \pmod{c}$, where a is an Eisenstein integer and c is a nonzero Eisenstein integer that is not an associate of a $\mathbb{Z}$-integer, we first need to define our fundamental region for the set of congruence classes.

We define our fundamental region R for the Eisenstein integers in a similar way to the Gaussian integers. We again can determine our rhombic fundamental region by the vertices at the origin, c, ωc and $(1 + \omega)c$. Similar to the Gaussian integers, all four vertices represent the same congruence class so we include only the origin, and we also include the two open intervals from 0 to c and from 0 to ωc. (Unlike the situation for $\mathbb{Z}[i]$, the fundamental region for $\mathbb{Z}[\omega]$ congruence classes is not a square, although it is still a rhombus.)

The rest of the argument is similar to the Gaussian integers, so we omit the details. We can find the representative of the equivalence class of an Eisenstein integer z, by the same function as presented earlier: $g(z) := z - c \cdot f(z/c)$, where $f : \mathbb{C} \to \mathbb{Z}[\omega]$ is the *Eisenstein floor function* defined by $f(x + \omega y) = m + \omega n$ whenever $x, y \in \mathbb{R}$ and m, n are the greatest integers that are no larger than x, y, respectively.

Let us consider an example with Eisenstein integers.

Example 5.28. Let us find congruence class representatives of $5 + 2\omega$. Our fundamental region R is the half-open rhombus with vertices $0 + 0i$, $5 + 2\omega$, $-2 + 3\omega$, and $3 + 5\omega$. As always, $0 \in R$. However, since $5 + 2\omega$ is prime, all other congruence class representatives are in the interior of R. Visualising this, we get Figure 16, where we indicate all elements of $\mathbb{Z}[\omega] \cap R$ with solid dots, and some nearby elements of $\mathbb{Z}[\omega] \setminus R$ with open dots.

As indicated in Figure 16, the set $\mathbb{Z}[\omega] \cap R$ of congruence class representatives is:

$$\{0,\ \omega,\ 1+\omega,\ 2+\omega,\ -1+2\omega,\ 2\omega,\ 1+2\omega,\ 2+2\omega,\ 3+2\omega,\ 4+2\omega,$$
$$-1+3\omega,\ 3\omega,\ 1+3\omega,\ 2+3\omega,\ 3+3\omega,\ 4+3\omega,\ 1+4\omega,\ 2+4\omega,\ 3+4\omega\}\,.$$

Example 5.29. In this example, we reduce $9348+3109\omega \mod 9+7\omega$. We start by computing $f((9348+3109\omega)/(9+7\omega))$:

$$\frac{9348+3109\omega}{9+7\omega} = \frac{(9348+3109\omega)(9+7\omega^2)}{67} = \frac{40459-37455\omega}{67}$$

Since

$$\left\lfloor \frac{40459}{67} \right\rfloor = 603 \qquad \text{and} \qquad \left\lfloor \frac{-37455}{67} \right\rfloor = -560\,,$$

we have $f((9348+3109\omega)/(9+7\omega)) = 603-560\omega$.

We now calculate

$$g(9348+3109\omega) = (9348+3109\omega) - (9+7\omega)(603-560\omega) = 1+8\omega\,,$$

and so $9348+3109\omega \equiv 1+8\omega \pmod{9+7\omega}$.

We next explore other aspects of equivalence classes modulo z and their representatives.

Definition 5.30. Let $x, c \in R$ with $c \neq 0$, where R stands for $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. We denote by $[x]_c$, or simply $[x]$ if c is understood, the equivalence class of all $z \in R$ that are congruent to x mod c. (Thus $[x]_c = x + (c)$.)

Cross **[3]** found an alternative simple and explicit set of congruence class representatives, at least modulo a $\mathbb{Z}[i]$-prime power. We next present his theorem.

Theorem 5.31. *We denote by p and q positive primes in $\mathbb{Z}$, subject to the congruences $p \equiv 3 \pmod 4$ and $q \equiv 1 \pmod 4$. We write π for one of the $\mathbb{Z}[i]$-prime factors of q and let α be the prime $1+i$. The equivalence classes of $\mathbb{Z}[i]$ modulo a power of a $\mathbb{Z}[i]$-prime are given as follows; all listed classes are distinct.*

(1) $\mathbb{Z}[i]/(\pi^n) = \{[a] : 0 \le a \le q^n - 1\}$.

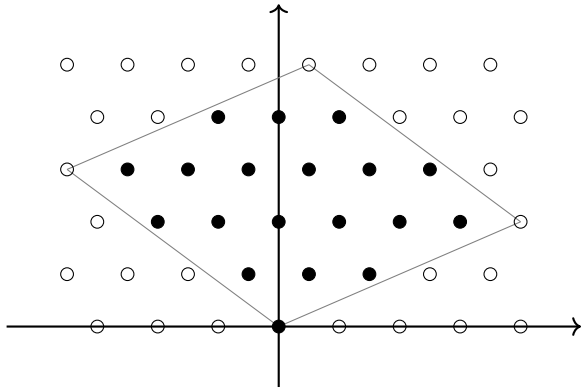

FIGURE 16. Fundamental region R for $5+2\omega$; $\mathbb{Z}[i] \cap R$ highlighted

(2) $\mathbb{Z}[i]/(p^n) = \{[a+bi] : 0 \le a \le p^n - 1 \text{ and } 0 \le b \le p^n - 1\}$.
(3) $\mathbb{Z}[i]/(\alpha^{2m}) = \{[a+bi] : 0 \le a \le 2^m - 1 \text{ and } 0 \le b \le 2^m - 1\}$.
(4) $\mathbb{Z}[i]/(\alpha^{2m+1}) = \{[a+bi] : 0 \le a \le 2^{m+1} - 1 \text{ and } 0 \le b \le 2^m - 1\}$.

PROOF. First, we observe that $\mathbb{Z}[i]/(\alpha^{2m}) = \mathbb{Z}[i]/(2^m)$ and $\mathbb{Z}[i]/(\alpha^{2m+1}) = \mathbb{Z}[i]/(2^m\alpha)$ because $\alpha^{2m} \sim 2^m$, where $x \sim y$ means that x and y are associates (and so $\sim$ is an equivalence relation).

Now, if $a + bi \equiv c + di \pmod{\alpha^{2m}}$, then 2^m divides both $a - c$ and $b - d$, so that the classes of (3) are distinct. A similar argument applies to the classes in (2). If $[a] = [b]$ in $\mathbb{Z}[i]/(\pi^n)$, then π^n divides $a - b$. Let $\pi^n\gamma = a - b$ for some $\gamma \in \mathbb{Z}[i]$. Taking complex conjugates, we get $\bar{\pi}^n = \overline{a-b} = a - b$, so that $\bar{\pi}^n$ also divides $a - b$. Since π and $\bar{\pi}$ are not associates, $\pi^n\bar{\pi}^n = q^n$ divides $a - b$ implying that the classes in (1) are distinct.

If $[a + bi] = [c + di]$ in $\mathbb{Z}[i]/(\alpha^{2m+1}) = \mathbb{Z}[i]/(2^m\alpha)$, then $2^m\alpha$ divides $a - c + (b - d)i$. A fortiori, 2^m divides $b - d$ and, if $0 \le b, d \le 2^m - 1$ as in (3), then $b = d$. With $b = d$, we now see that $2^m\alpha$ divides $a - c$. Let $a - c = 2^m k$, where $k \in \mathbb{Z}$ since $a - c \in \mathbb{Z}$. Then α divides k so that $N(\alpha) = 2$ divides $N(k) = k^2$. It follows that k is even, i.e. 2^{m+1} divides $a - c$ and the classes in (4) are distinct.

Finally, since the listed elements are pairwise noncongruent in each case, and since the number of elements in each case matches the norm of the prime power, we see that we indeed have complete sets of representatives. □

We next adapt the above theorem to the Eisenstein integers, based on the ideas in **[3]** for the Gaussian integers. Here, we denote by p and q positive primes in $\mathbb{Z}$, subject to the congruences $p \equiv 2 \pmod 3$ and $q \equiv 1 \pmod 3$. We write π for one of the $\mathbb{Z}[\omega]$-prime factors of q and let α denote the prime $1 + 2\omega$.

Theorem 5.32. *The equivalence classes of $\mathbb{Z}[\omega]$ modulo a power of a prime are given as follows; all listed classes are distinct.*

(1) $\mathbb{Z}[\omega]/(\pi^n) = \{[a] : 0 \le a \le q^n - 1\}$.
(2) $\mathbb{Z}[\omega]/(p^n) = \{[a+bi] : 0 \le a \le p^n - 1 \text{ and } 0 \le b \le p^n - 1\}$.
(3) $\mathbb{Z}[\omega]/(\alpha^{2m}) = \{[a+bi] : 0 \le a \le 3^m - 1 \text{ and } 0 \le b \le 3^m - 1\}$.
(4) $\mathbb{Z}[\omega]/(\alpha^{2m+1}) = \{[a+bi] : 0 \le a \le 3^{m+1} - 1 \text{ and } 0 \le b \le 3^m - 1\}$.

The proof is very similar to that for the Gaussian integers, so we omit it.

The divisibility equivalence (5.3) immediately implies the following congruence equivalence.

Theorem 5.33. *Let R denote $\mathbb{Z}[i]$ or $\mathbb{Z}[\omega]$. For $a, b, c \in \mathbb{Z}$, we have $a \equiv b \pmod c$ in $\mathbb{Z}$ if and only if $a \equiv b \pmod c$ in R.*

Theorem 5.34. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. If π is prime in R, and $a, b \in R$, then $ab \equiv 0 \pmod \pi$ if and only if $a \equiv 0 \pmod \pi$ or $b \equiv 0 \pmod \pi$.*

PROOF. If one of a and b is congruent to 0 mod π, it follows that $ab \equiv 0 \pmod \pi$.

Conversely, suppose $ab \equiv 0 \pmod \pi$, and so $\pi \mid (ab - 0)$. Since π is prime, it follows that either $\pi \mid a$ or $\pi \mid b$, i.e. either $a \equiv 0 \pmod \pi$ or $b \equiv 0 \pmod \pi$.

The proof for $\mathbb{Z}[i]$ and for $\mathbb{Z}[\omega]$ is similar. □

We have seen two different sets of congruence class representatives for both Gaussian and Eisenstein integers: one is the approach of Cross **[3]**, while the other involves the (Gaussian/Eisenstein) integers contained in a rhombic fundamental region R. From now on, when we are reducing modulo a Gaussian or an Eisenstein integer, we always use the representatives contained in the rhombic fundamental region.

Theorem 5.35. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. For a and b in R with $b \neq 0$, the congruence $ax \equiv 1 \pmod{b}$ is solvable if and only if a and b are coprime in R. If a and b are coprime, then any linear congruence $ax \equiv c \bmod b$ has a unique solution.*

PROOF. To find $x \in R$ such that $ax \equiv 1 \pmod{b}$, we need to solve $ax + by = 1$ with $x, y \in R$. By Corollary 5.12, this can be done if and only if a and b are coprime.

Once we can invert $a \pmod{b}$, we can solve $ax \equiv c \pmod{b}$ by multiplying both sides by the inverse of $a \pmod{b}$. This gives us the unique solution. □

Proposition 5.36. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let $b, n \in R$, where n is not a unit. Then b is invertible modulo n if and only if 1 is a GCD of b and n.*

PROOF. Assume b is invertible modulo n, and let c denote its inverse. Since $bc = 1 \pmod{n}$, this implies $bc - 1 = kn$ for some $k \in \mathbb{Z}$. Equivalently, $bc - kn = 1$. It follows from Corollary 5.12 that 1 is a GCD of b and n.

For the converse, suppose that 1 is a GCD of b and n. By Theorem 5.11, there exist integers x, y such that $xb + yn = 1$. It follows that x is a multiplicative inverse of b. □

Definition 5.37 (Euler ϕ-Function)**.** For $n \in \mathbb{N}$, $n > 1$, we define the Euler ϕ-function, $\phi(n)$, as the number of positive integers less than n that are coprime with n. We define $\phi(1) = 1$. Alternatively, in view of Corollary 5.12, $\phi(n)$ is the number of units in the ring $\mathbb{Z}_m$.

For every prime p, it is clear that $\phi(p) = p - 1$.

We now show that ϕ is *multiplicative*, meaning that $\phi(mn) = \phi(m)\phi(n)$ whenever m and n are coprime.

Theorem 5.38. *Suppose $m, n \in \mathbb{N}$ are coprime. Then $\phi(mn) = \phi(m)\phi(n)$.*

PROOF. By our alternative definition of ϕ, $\phi(mn)$ is the number of units in $\mathbb{Z}_{mn}$, while $\phi(m)\phi(n)$ is the number of units in $\mathbb{Z}_m \times \mathbb{Z}_n$. When m and n are coprime, the Chinese remainder theorem Theorem 5.17 establishes an isomorphism between $\mathbb{Z}_{mn}$ and $\mathbb{Z}_m \times \mathbb{Z}_n$, so we are done. □

We now give a general formula for $\phi(n)$.

Theorem 5.39. *Suppose that $n = \prod_{i=1}^{t} p_i^{\alpha_i}$, where the p_i are distinct $\mathbb{Z}$-primes and $\alpha_i \in \mathbb{N}$, $1 \leq i \leq t$. Then*

$$\phi(n) = n \prod_{i=1}^{t} \left(\frac{p_i - 1}{p_i} \right) .$$

PROOF. As ϕ is multiplicative, we know that

$$\phi(n) = \phi \left(\prod_{i=1}^{t} p_i^{\alpha_i} \right) = \prod_{i=1}^{t} \phi(p_i^{\alpha_i}).$$

Clearly, every $n \in \mathbb{N}$ that is not a multiple of p_i is coprime to p_i, and hence to $p_i^{\alpha_i}$. Thus,

$$\phi(n) = \prod_{i=1}^{t} \left(p_i^{\alpha_i} - p_i^{\alpha_i - 1} \right) = \prod_{i=1}^{t} p_i^{\alpha_i - 1}(p_i - 1) .$$

It follows that

$$\phi(n) = n \prod_{i=1}^{t} (1 - 1/p_i) ,$$

as required. □

Lemma 5.40. *For $n > 2$, $\phi(n)$ is even.*

PROOF. We split this proof up into two cases.

Case 1: Assume n has (at least one) odd prime factor, say p. It follows from Theorem 5.39 that $p - 1 \mid \phi(n)$. As p is an odd prime, $p - 1$ is even, and so $2 \mid p - 1 \mid \phi(n)$. Thus, $\phi(n)$ is even.

Case 2: Assume that n has no odd prime factors. Thus, the only prime factor of n is 2 and so $n = 2^k$, where $k > 1$. By Theorem 5.39, $\phi(n) = 2^k(1 - \frac{1}{2}) = 2^{k-1}$, where $k - 1 > 0$. Hence $\phi(n)$ is even. □

We model the definition of ϕ-function for $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ on the alternative definition of the Euler ϕ-function, see **[15]**:

Definition 5.41. Let R denote $\mathbb{Z}[i]$ or $\mathbb{Z}[\omega]$. For nonzero a in R, let $\phi(a)$ be the number of units in $R/(a)$. To distinguish between the ϕ functions on different rings, we often write $\phi_{\mathbb{Z}}$, $\phi_{\mathbb{Z}[i]}$, etc.

We saw that if p is a $\mathbb{Z}$-prime then $\phi_{\mathbb{Z}}(p) = p - 1$. It is similarly clear that if π is a Gaussian prime, then $\phi_{\mathbb{Z}[i]}(\pi) = N(\pi) - 1$, while if σ is an Eisenstein prime, then $\phi_{\mathbb{Z}[\omega]}(\sigma) = N(\sigma) - 1$.

The following theorem gives the $\mathbb{Z}[i]$- and $\mathbb{Z}[\omega]$-analogues of Theorems 5.38 and 5.39.

Theorem 5.42. *Let R denote $\mathbb{Z}[i]$ or $\mathbb{Z}[\omega]$.*

(a) If m and n are coprime in R, then $\phi_R(mn) = \phi_R(m)\phi_R(n)$.
(b) If π is a prime in R, and $k \in \mathbb{N}$, then $\phi(\pi^k) = N(\pi^{k-1})(N(\pi) - 1)$.

Let us simply sketch the proof of the above theorem. The proof of (a) is formally the same as Theorem 5.38 since our version of the Chinese remainder theorem holds in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ as well as in $\mathbb{Z}$. As for (b), we simply need to note that there is a unique congruence class mod π corresponding to non-units in R, and this corresponds to $N(\pi^{k-1})$ congruence classes mod π^k.

We now state a version of Fermat's Little Theorem for all three of our rings.

Theorem 5.43. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Suppose $a, \pi \in R$ are coprime and π is a prime. Then $a^{N(\pi)-1} \equiv 1 \pmod{\pi}$.*

PROOF. In any ring, the units form a multiplicative group. Here, π is prime, so every nonzero element in the factor ring $R/(\pi)$ is a unit. By Lagrange's theorem, the order of the cyclic subgroup generated by any unit in $R/(\pi)$ divides the order $N(\pi) - 1$ of the group of units. □

We now state Euler's extension of Fermat's Little Theorem, and a simple but useful corollary of it.

Theorem 5.44. *Let $a, n \in \mathbb{Z}$, $n \neq 0$. If $\gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$.*

Corollary 5.45. *Let p and q be two distinct primes and let $m \in \mathbb{N}$. For every $k \in \mathbb{N}$:*

$$m^{k(p-1)(q-1)+1} \equiv m \pmod{pq}.$$

We now restate Euler's extension in a way that makes it true in all three of the rings that interest us. The proof remains essentially the same.

Theorem 5.46. *Let R denote $\mathbb{Z}$, $\mathbb{Z}[i]$, or $\mathbb{Z}[\omega]$. Let $\phi_R(\nu)$ denote the number of units in the ring $R/(\nu)$. Then, for any $a \in R$, we have $a^{\phi_R(\nu)} \equiv 1 \pmod{\nu}$.*

5.1. Exponentiation for Gaussian and Eisenstein integers.

Raising a Gaussian or Eisenstein integer to an integer power can be viewed as special cases of exponentiating a complex number.

For instance, suppose $z = a + bi$, $a, b \in \mathbb{Z}$, and suppose we want to compute z^n for some $n \in \mathbb{N}$. We can first write z in exponential form, facilitating the computation of z^n, and then rewrite the solution in the usual $\mathbb{Z}[i]$-format. Explicitly,

$$z = a + bi = r \cdot e^{i\theta},$$

where $r = \sqrt{a^2 + b^2}$ and $\theta = \arg(a + bi)$. Then, $z^n = r^n \cdot e^{i\theta n}$. Thus, $z^n = c + di$, where

$$c = r^n \cdot \cos\left(n \cdot \arg\left(\frac{b}{a}\right)\right) \qquad \text{and} \qquad d = r^n \cdot \sin\left(n \cdot \arg\left(\frac{b}{a}\right)\right).$$

Example 5.47. In this example, we want to compute $(3 + 2i)^5 \bmod (4 + i)$.

First, we calculate $r = \sqrt{3^2 + 2^2} = \sqrt{13}$ and $\theta = \arctan(2/3)$, so

$$z := 3 + 2i = \sqrt{13} \cdot \exp(i \cdot \arctan(2/3)).$$

Now,

$$z^5 = 13^{5/2} \cdot \exp(5i \cdot \arctan(2/3)).$$

We can rewrite this as $c + di$, $c, d \in \mathbb{Z}$, where

$$c = 13^{5/2} \cdot \cos(5 \arctan(2/3)) = -597\,,$$
$$d = 13^{5/2} \cdot \sin(5 \arctan(2/3)) = 122\,.$$

Hence, $(3 + 2i)^5 = -597 + 122i$. Finally, we reduce this: $-597 + 122i \equiv -1 + i \pmod{4 + i}$.

However, the above method only works efficiently for small numbers. For large $|z|$, the numbers c and d computed above prior to reduction are too big to be calculated efficiently. To speed up the calculations and to use a more efficient way of exponentiating large numbers, we use the following algorithm:

Fast Exponentiation Algorithm

The fast exponentiation algorithm is an efficient method to calculate powers using modular arithmetic. For small exponents, exponentiation is easily carried out. For big exponents, however, exponentiation gets increasingly difficult. The fast exponentiation algorithm enables us to efficiently calculate the solutions using large exponents. This algorithm is based on calculating squares and multiplication, hence why it is often also referred to as the Square-and-Multiply Algorithm. A key feature is that we continually reduce mod n as we proceed, stopping the calculations from getting cumbersome.

To illustrate this algorithm, we first demonstrate it using an example in $\mathbb{Z}$, and then discuss an example in $\mathbb{Z}[i]$.

Example 5.48. Let us calculate $3^{412} \pmod{101}$.

$$\begin{aligned}
3^1 = 3 &\equiv 3 \pmod{101}\\
(3^1)^2 = 3^2 &\equiv 9 \pmod{101}\\
(3^2)^2 = 3^4 &\equiv 81 \pmod{101}\\
(3^4)^2 = 3^8 &\equiv 97 \pmod{101}\\
(3^8)^2 = 3^{16} &\equiv 16 \pmod{101}\\
(3^{16})^2 = 3^{32} &\equiv 54 \pmod{101}\\
(3^{32})^2 = 3^{64} &\equiv 88 \pmod{101}\\
(3^{64})^2 = 3^{128} &\equiv 68 \pmod{101}\\
(3^{128})^2 = 3^{256} &\equiv 79 \pmod{101}
\end{aligned}$$

Now, $3^{412} = 3^{256+128+16+8+4}$, and so

$$3^{412} \equiv 79 \cdot 68 \cdot 16 \cdot 97 \cdot 81 \equiv 80 \pmod{101}\,.$$

Example 5.49. As in Example 5.47, we calculate $(3 + 2i)^5 \pmod{4 + i}$. However, this time, we use the fast exponentiation method.

$(3 + 2i)^2 = 5 + 12i$ and $5 + 12i \equiv -2i \pmod{4 + i}$, and so

$$(3 + 2i)^4 \equiv (-2i)^2 \equiv -4 \pmod{4 + i}\,.$$

Now,

$$(3+2i)^5 \equiv (3+2i)^4 \cdot (3+2i) \equiv (3+2i) \cdot (-4) \equiv -12-8i \equiv -1-i \pmod{4+i}.$$

The fast exponentiation algorithm can similarly be used to efficiently raise Eisenstein integers to large exponents modulo any given number.

6. Group theoretic background

Definition 6.1. A group is a set G equipped with a binary operation $\circ : G \times G \to G$ satisfying the following conditions:

- (Closure:) For all $g, h \in G$, $g \circ h \in G$.
- (Existence of identity:) There exists an identity $e \in G$ such that for all $g \in G$, $e \circ g = g = g \circ e$.
- (Existence of inverses:) For all $g \in G$, there exists an element $h \in G$ such that $g \circ h = e = h \circ g$. Such an h is called an inverse of g.
- (Associativity:) For all $g_1, g_2, g_3 \in G$, $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$.

If G has a finite number of elements, we say G is a *finite group*. We let $|G|$ denote the order of the group, i.e. the cardinality of the set G.

If the group operation $\circ$ is *commutative*, i.e. if for all $g, h \in G$, $g \circ h = h \circ g$, then the group G is an *abelian group*.

It is readily verified that the identity element in the group G is unique and each element $g \in G$ has a unique inverse. From now on, 1 denotes the identity element and g^{-1} denotes the inverse of g.

A set $H \subseteq G$ is a subgroup of G if H itself forms a group under the same operation associated with G. Every group G always has the trivial subgroups G and $\{1\}$. We call H a *strict subgroup* of G if $H \neq G$.

Lemma 6.2. *Let G be a group and $a, b, c \in G$. If $ac = bc$, then $a = b$. In particular, if $ac = c$, then a is the identity in G.*

Proof. We know that $ac = bc$. Multiplying both sides by the inverse c^{-1} of c, we get

$$ab = bc \Rightarrow (ac)c^{-1} = (bc)c^{-1} \Rightarrow a(cc^{-1}) = b(cc^{-1}) \Rightarrow a \cdot 1 = b \cdot 1 \Rightarrow a = b\,. \qquad \square$$

Theorem 6.3. *Let G be a finite group, and we write $m = |G|$. Then $g^m = 1$ for every $g \in G$.*

We present a proof only for the abelian case, which suffices for us. (The well-known proof of the full-strength result involves a consideration of cosets of the subgroup generated by g. However, we do not even formally define cosets in this thesis.)

Proof. Fix arbitrary $g \in G$, and let $g_1, \ldots, g_m$ be the elements of the finite abelian group G. We claim that

$$g_1 \cdot g_2 \cdots g_m = (gg_1) \cdot (gg_2) \cdots (gg_m).$$

To see this, note that $gg_i = gg_j$ implies $g_i = g_j$ by Lemma 6.2. So each of the m elements in parentheses on the right-hand side is distinct. Because there are exactly m elements in G, the m elements being multiplied together on the right-hand side are simply all elements of G in some permuted order. Since G is abelian, the order in which elements are multiplied does not matter, and so the right-hand side is equal to the

left-hand side. Again using the fact that G is abelian, we can pull out all occurrences of g and obtain

$$g_1 \cdot g_2 \cdots g_m = (gg_1) \cdot (gg_2) \cdots (gg_m) = g^m \cdot (g_1 \cdot g_2 \cdots g_m).$$

By Lemma 6.2, this implies that $g^m = 1$. □

Corollary 6.4. *Let G be a finite group with $|m| = G > 1$. If $x, y \in \mathbb{Z}$ are such that $x \equiv y \pmod{m}$, then for every $g \in G$, we have $g^x = g^y$.*

PROOF. Suppose without loss of generality that $x > y$ and so $x = km + y$ for some $k \in \mathbb{N}$. Using Theorem 6.3, we get

$$g^x = g^{km+y} = g^{km} \cdot g^y = (g^m)^k \cdot g^y = 1^k \cdot g^y = g^y \,,$$

as claimed. □

Corollary 6.5. *Let G be a finite group with $m = |G| > 1$. Let $e > 0$ be an integer, and define $f_e : G \to G$ by $f_e(g) = g^e$. If $\gcd(e, m) = 1$, then f_e is a permutation (i.e. it is a bijection). Moreover, if we choose $d \in \mathbb{N}$ such that $d \equiv e^{-1} \pmod{m}$ then f_d is the inverse of f_e.*

PROOF. Since G is finite, the result follows once we show that f_d is the inverse of f_e. The assumption on d says that $de = 1 + km$ for some $m \in \mathbb{N}$. Thus,

$$f_d(f_e(g)) = f_d(g^e) = (g^e)^d = g^{ed} = g^1 \cdot g^{km} = g^1 = g \,,$$

where the fifth equality follows from Corollary 6.4. □

In the following proposition, finiteness is essential. For instance, $\mathbb{N}$ is a nonempty subset of $(\mathbb{Z}, +)$ which is closed under addition, but it is not a subgroup because it contains neither an identity nor inverses.

Proposition 6.6. *Let G be a finite group, and let H be a nonempty subset of G. If $ab \in H$ for all $a, b \in H$, then H is a subgroup of G.*

PROOF. We need to verify that H satisfies all the conditions of Definition 6.1. By assumption, H is closed under the group operation. Associativity in H is inherited from G. Let $m = |G|$ (here we use the fact that G is finite), and consider an arbitrary element $a \in H$. Closure of H means that H contains $a^m = 1$ as well as $a^{m-1} = a^m \cdot a^{-1} = a^{-1}$. Thus, H contains the identity as well as the inverse of each of its elements, making it a group. □

Lemma 6.7. *Let H be a strict subgroup of a finite group G (i.e. $H \neq G$). Then, $|H| \leq |G|/2$.*

PROOF. Let $\bar{h}$ be an element of G that is not in H; since $H \neq G$, we know that $\bar{h}$ exists. Consider the set $\bar{H} := \{\bar{h}h \mid h \in H\}$. We show that (1) $|\bar{H}| = |H|$, and (2) every element of $\bar{H}$ lies outside of H, i.e. the intersection of H and $\bar{H}$ is empty. Since both H and $\bar{H}$ are subsets of G, these imply $|G| \geq |H| + |\bar{H}| = 2|H|$, proving the lemma.

Suppose $h_1, h_2 \in H$. If $\bar{h}h_1 = \bar{h}h_2$ then by Lemma 6.2, we have $h_1 = h_2$. This shows that every distinct element $h \in H$ corresponds to a distinct element $\bar{h}h \in \bar{H}$, proving (1).

Assume for the sake of contradiction that $h \in H$ is such that $h' := \bar{h}h \in H$. Then $\bar{h} = h'h^{-1} \in H$, contradicting the assumption that $\bar{h} \in G \setminus H$. This proves (2) and completes the proof of the lemma. □

6.1. The group $\mathbb{Z}_N^*$.

Let us define the modulo function $\mathrm{Mod}_N : \mathbb{Z} \to \mathbb{Z}_N := \{0, \ldots, N-1\}$ by $\mathrm{Mod}_N(n) = k$, where $k \in \mathbb{Z}_N$ is such that $n \equiv k \pmod{N}$.

The set $\mathbb{Z}_N$ is a group under addition modulo $N \in \mathbb{N}$, i.e. addition $+_N$ on $\mathbb{Z}_\mathbb{N}$ is defined by $x +_N y = \mathrm{Mod}_N(x + y)$. We simply write $+$ in place of $+_N$ from now on. More interesting to us, however, is a related group under multiplication, namely

$$\mathbb{Z}_N^* := \{b \in \{1, \ldots, N-1\} \mid \gcd(b, N) = 1\}$$

i.e. $\mathbb{Z}_N^*$ consists of integers in the set $\{1, \ldots, N-1\}$ that are coprime to N. The multiplication operation, $\cdot_N$, is defined by $x \cdot_N y = \mathrm{Mod}_N(xy)$. Below, we more briefly write $x \cdot y$, or simply xy, in place of $x \cdot_N y$.

Corollary 6.8. *For $N > 1$ and for integer $e > 0$, define $f_e : \mathbb{Z}_N^* \to \mathbb{Z}_N^*$ by $f_e(x) = x^e \pmod{N}$. If e is coprime to N, then f_e is permutation. Moreover, if $d = e^{-1} \pmod{\phi(N)}$, then f_d is the inverse of f_e.*

6.2. Isomorphisms and the Chinese Remainder Theorem.

Definition 6.9. Let G, H be groups with group operations $\circ_G$ and $\circ_H$ respectively. A function $f : G \to H$ is an *isomorphism* from G to H if:

1. f is a bijection, and
2. for all $g_1, g_2 \in G$ we have $f(g_1 \circ_G g_2) = f(g_1) \circ_H f(g_2)$.

We say that two groups G and H are *isomorphic* and write $G \cong H$ if there exists an isomorphism from G to H, .

An isomorphism from G to H is just a renaming of elements of G as elements of H. If G is finite and $G \cong H$, then H must be finite and of the same cardinality as G. Moreover, if there exists an isomorphism f from G to H, then f^{-1} is an isomorphism from H to G.

Definition 6.10. Given groups G, H with group operations $\circ_G, \circ_H$, respectively, we define a new group $G \times H$, the *direct product* of G and H as follows: The elements of $G \times H$ are ordered pairs (g, h) with $g \in G$ and $h \in H$. Thus, if G has n elements and H has n' elements, then $G \times H$ has $n \cdot n'$ elements. The group operation $\circ$ on $G \times H$ is applied componentwise; that is

$$(g, h) \circ (g', h') := (g \circ_G g', h \circ_H h').$$

We now state a group theoretic version of the Chinese remainder theorem. We omit the proof since it follows from Theorem 5.17.

Theorem 6.11 (Chinese remainder theorem)**.** *Let $N = pq$ where $p, q > 1$ are coprime. Then*

$$\mathbb{Z}_N \cong \mathbb{Z}_p \times \mathbb{Z}_q \quad \textit{and} \quad \mathbb{Z}_N^* \cong \mathbb{Z}_p^* \times \mathbb{Z}_q^*.$$

Moreover, let f be the function mapping elements $x \in \{0, \ldots, N-1\}$ to pairs (x_p, x_q) with $x_p \in \{0, \ldots, p-1\}$ and $x_q \in \{0, \ldots, q-1\}$ defined by

$$f(x) := (\mathrm{Mod}_p(x), \mathrm{Mod}_q(x)).$$

Then f is an isomorphism from $\mathbb{Z}_N$ to $\mathbb{Z}_p \times \mathbb{Z}_q$, and the restriction of f to $\mathbb{Z}_N^*$ is an isomorphism from $\mathbb{Z}_N^*$ to $\mathbb{Z}_p^* \times \mathbb{Z}_q^*$.

Let G, H be two groups with group operations $\circ_G, \circ_H$, respectively, and let f be an isomorphism from G to H, where both f and f^{-1} can be computed efficiently. Then, for $g_1, g_2 \in G$, we can compute $g = g_1 \circ_G g_2$ in two ways: either by directly computing it in G, or via the following steps:

1. Compute $h_1 = f(g_1)$ and $h_2 = f(g_2)$.
2. Compute $h = h_1 \circ_H h_2$ using the group operation in H.
3. Compute $g = f^{-1}(h)$.

Now, we look at specific examples working modulo n. We write $a \leftrightarrow (b, c)$ or $(b, c) \leftrightarrow a$ whenever $f(a) = (b, c)$ in Theorem 6.11.

Example 6.12. In this example, we let $p = 5$ and $q = 13$. Then $\mathbb{Z}_{65} \cong \mathbb{Z}_5 \times \mathbb{Z}_{13}$. Suppose that we want to compute $16 \cdot 37 \pmod{65}$. Since $16 \equiv 1 \pmod 5$, $16 \equiv 3 \pmod{13}$, $37 \equiv 2 \pmod 5$, and $37 \equiv 11 \pmod{13}$, we have

$$16 \leftrightarrow (1, 3) \quad \text{and} \quad 37 \leftrightarrow (2, 11).$$

In $\mathbb{Z}_5 \times \mathbb{Z}_{13}$, we have

$$(1, 3) \cdot (2, 11) \equiv (\mathrm{Mod}_5(1 \cdot 2), \mathrm{Mod}_{13}(3 \cdot 11)) \equiv (2, 7).$$

Finally, $(2, 7) \leftrightarrow 7$, and so $16 \cdot 37 \equiv 7 \pmod{65}$.

Example 6.13. Again, we let $p = 5$ and $q = 13$, so $\mathbb{Z}_{65} \cong \mathbb{Z}_5 \times \mathbb{Z}_{13}$. Suppose that we want to compute $16^{37} \pmod{65}$. As before, $16 \leftrightarrow (1, 3)$, so $16^{37} \leftrightarrow (1, 3)^{37}$.

$$(1, 3)^{37} \equiv (\mathrm{Mod}_5(1^{37}), \mathrm{Mod}_{13}(3^{37})) \equiv (1, 3).$$

Thus, $16^{37} \equiv 16 \pmod{65}$. (Note that the quickest way to compute $\mathrm{Mod}_{13}(3^{37})$ above is to use Theorem 6.3: since $|\mathbb{Z}_{13}^*| = 12$, we have $3^{37} = 3^{1+3\times 12} \equiv 3 \pmod{13}$.)

We also need to discuss how to convert back and forth between the elements in $\mathbb{Z}_n$ and the elements in $\mathbb{Z}_p$ and $\mathbb{Z}_q$, where $n = pq$ and p, q are distinct primes. Assuming p and q are known, we can as above map an element x modulo n to the corresponding element of $\mathbb{Z}_p \times \mathbb{Z}_q$ by mapping x to $(\mathrm{Mod}_p(x), \mathrm{Mod}_q(x))$.

To discuss the other direction, we take an element $(x_p, x_q) \in \mathbb{Z}_p \times \mathbb{Z}_q$ and we want to map it to the corresponding element of $\mathbb{Z}_n$. We can write (x_p, x_q) as

$$(x_p, x_q) = x_p \cdot (1, 0) + x_q \cdot (0, 1).$$

So, if we can find elements $1_p, 1_q \in \{0, \ldots, N-1\}$ such that $1_p \leftrightarrow (1,0)$ and $1_q \leftrightarrow (0,1)$, then, using the Chinese remainder theorem, we get

$$(x_p, x_q) \leftrightarrow \mathrm{Mod}_n(x_p \cdot 1_p + x_q \cdot 1_q).$$

Since p, q are distinct primes, $\gcd(p,q) = 1$. Therefore, using the Euclidean algorithm and running it backwards, we can find integers a, b such that

$$ap + bq = 1\,.$$

Note that $bq \equiv 0 \pmod{q}$ and $bq \equiv 1 \pmod{p}$. Hence, $bq \pmod{n} \leftrightarrow (1,0)$, i.e. $bq \pmod{n} \equiv 1_q$. Similarly, $ap \pmod{n} \equiv 1_q$.

In summary, we can convert an element represented as (x_p, x_q) to its representation modulo n via the following steps (assuming p and q are known):

1. Compute a, b such that $ap + bq = 1$.
2. Set $1_p := bq$ and $1_q := ap$.
3. Compute $x := \mathrm{Mod}_n(x_p \cdot 1_p + x_q \cdot 1_q)$.

Note that we do not need to reduce 1_p and 1_q mod n because it suffices to reduce them in the final step.

Example 6.14. Again, we let $p = 5$ and $q = 13$, and so $\mathbb{Z}_{65} \cong \mathbb{Z}_5 \times \mathbb{Z}_{13}$. Suppose that we are given $(4,3) \in \mathbb{Z}_5 \times \mathbb{Z}_{13}$ and we want to convert this to the corresponding element of $\mathbb{Z}_{65}$. Using backwards substitution for the Euclidean algorithm, we get

$$2 \cdot 13 - 5 \cdot 5 = 1.$$

Thus, $1_5 \equiv 2 \cdot 13 \equiv 26 \pmod{65}$ and $1_{13} \equiv -5 \cdot 5 \equiv 40 \pmod{65}$. (To verify this, we check that $1_5 \equiv 26 \equiv 1 \pmod{5}$, $1_5 \equiv 26 \equiv 0 \pmod{13}$, $1_{13} \equiv 40 \equiv 0 \pmod{5}$ and $1_{13} \equiv 40 \equiv 1 \pmod{13}$.)

Using these values, we compute

$$\begin{aligned}(4,3) &= 4 \cdot (1,0) + 3 \cdot (1,0) \\ &\leftrightarrow 4 \cdot 1_p + 3 \cdot 1_q \pmod{65} \\ &= 4 \cdot 26 + 3 \cdot 40 \pmod{65} \\ &= 29 \pmod{65}\end{aligned}$$

Since 29 gives $29 \equiv 4 \pmod{5}$ and $29 \equiv 3 \pmod{13}$, this is indeed the correct result.

6.3. Cyclic Groups.

Definition 6.15. Let G be a finite group and $g \in G$. The *order* of g, written $\mathrm{ord}(g)$, is the smallest positive integer i with $g^i = 1$.

Below, g^x for negative integers x is to be interpreted as h^{-x} where $h = g^{-1}$.

Proposition 6.16. *Let G be a finite group, and $g \in G$ an element of order i. If $x, y \in \mathbb{Z}$, then $g^x = g^y$ if and only if $x = y \pmod{i}$.*

PROOF. Suppose $x = y \pmod{i}$, and so $x - y = ki$ for some $k \in \mathbb{Z}$. If $k \geq 0$, then

$$g^x = g^y g^{ki} = g^y \cdot (g^i)^k = g^y \cdot 1 = g^y \,,$$

as required. If instead $k < 0$, reverse the roles of x and y above to reach the same conclusion.

To prove the other direction, suppose $g^x = g^y$ where, without loss of generality, $y \leq x$. We write $x - y$ in the form $qi + r$ for some non-negative $q, r \in \mathbb{Z}$, $0 \leq r < i$. Then,

$$1 = g^{x-y} = g^{qi+r} = (g^i)^q g^r = g^r \,.$$

Thus, $g^r = 1$. This contradicts the definition of the order of g unless $r = 0$, in which case it is clear that $x \equiv y \pmod{i}$. □

In any group G, the identity element is the only element of order 1, and generates the group $(1) = \{1\}$. If there is an element $g \in G$ that has order m, where m is the order of the group G, then (g), the subgroup generated by g, equals G. In this case, we call G a *cyclic group* and say that g is a generator of G.

Note that every cyclic group is abelian.

Definition 6.17. A group G is cyclic if G can be generated by a single element, i.e., there exists some element $g \in G$ such that $G = \{g^n \mid n \in \mathbb{Z}\}$.

If g is a generator of G then, by definition, every element $h \in G$ is equal to g^x for some $x \in \{0, \ldots, m-1\}$, where m is the order of G.

Proposition 6.18. *Let G be a finite group of order m, and say $g \in G$ has order i. Then $i \mid m$.*

PROOF. By Theorem 6.3, we know that $g^m = 1 = g^0$. Proposition 6.16 implies that $m = 0 \pmod{i}$. □

Corollary 6.19. *If G is a group of prime order p, then G is cyclic. Furthermore, all elements of G except the identity are generators of G.*

PROOF. By Proposition 6.18, the only possible orders of elements in G are 1 and p. Only the identity element has order 1, and so all other elements have order p and generate G. □

Lemma 6.20. *Let $g \in G$, with $\operatorname{ord}(g) = n$. Then, for any $k \mid n$, there exists $h \in G$ with $\operatorname{ord}(h) = k$.*

PROOF. We simply let $h = g^{n/k}$. □

Lemma 6.21. *Suppose G is an abelian group and $a, b \in G$ are such that $\operatorname{ord}(a) = n$ and $\operatorname{ord}(b) = m$, where $\gcd(n, m) = 1$. Then, there exists $c \in G$, with $\operatorname{ord}(c) = nm$.*

PROOF. We claim that ab has order nm. As $(ab)^{nm} = (a^n)^m (b^m)^n = 1^m 1^n = 1$, we see that $\operatorname{ord}(ab) = k$, for some $k \mid nm$.

$$(ab)^k = 1 \implies a^k = b^{-k} \,.$$

Raising both sides to the m^{th} power, we get $a^{mk} = 1$. Thus, $n \mid mk$. But as $\gcd(n, m) = 1$, this implies that $n \mid k$. As G is abelian, we can switch the roles of a and b and get $m \mid k$. Thus, $nm \mid k$ and hence, $k = nm$. □

Lemma 6.22. *Suppose G is an abelian group and $a, b \in G$ are such that $ord(a) = n$ and $ord(b) = m$. Then, there exists $c \in G$ such that $ord(c) = [n, m]$, where $[n, m]$ is the lowest common multiple of n and m.*

Proof. By Lemma 6.20, there exists $c_1, c_2, c_3 \in G$ with

$$\text{ord}(c_1) = \gcd(n, m), \quad \text{ord}(c_2) = \frac{n}{\gcd(n, m)} \quad \text{and} \quad \text{ord}(c_3) = \frac{m}{\gcd(n, m)}.$$

Each of the orders are pairwise coprime, hence, by Lemma 6.21, there exists $c \in G$ such that

$$\text{ord}(c) = \gcd(n, m) \cdot \frac{n}{\gcd(n, m)} \cdot \frac{m}{\gcd(n, m)} = \frac{nm}{\gcd(n, m)} = [n, m] \ . \qquad \square$$

Theorem 6.23. *If p is prime, then $\mathbb{Z}_p^*$ is a cyclic group of order $p - 1$.*

Proof. Since p is a prime, all nonzero elements of $\mathbb{Z}_p$ are coprime to p. This means that $(\mathbb{Z}_p, +, *)$ is actually a field, a fact that we will need later. For now, though, we note that $\mathbb{Z}_p^*$ contains all the nonzero numbers and so the order of $\mathbb{Z}_p^*$ is $p - 1$, as required. Hence, we only need to show that $\mathbb{Z}_p^*$ is cyclic, and we prove this by contradiction.

Assume $\mathbb{Z}_p^*$ is not cyclic. Let $m_i := \text{ord}(i)$. By Lemma 6.22, there exists $c \in \mathbb{Z}_p^*$ with $\text{ord}(c) = d := [m_1, \ldots, m_{p-1}]$. Since $\mathbb{Z}_p^*$ is not cyclic, d must be a strict divisor of $p-1$, lest c be a generator of the group.

Since d is a multiple of every m_i, we have $i^d - 1 = 0 \pmod{p}$, of course) for all $i \in \mathbb{Z}_p^*$. Thus, the polynomial $x^d - 1$ has $p - 1$ roots in the field $\mathbb{Z}_p$. Since a polynomial of degree d can have at most d roots in a field, and since d is strictly less than $p - 1$, we get a contradiction. □

7. Factorization and Primality Testing

Given a positive integer n, the idea of factorization is to find integers $p, q > 1$ such that $n = pq$. Theoretically, the *trial division factorization algorithm* (TDFA, below) is straightforward: we simply test if n is divisible by p for all possible divisors p, beginning with $p = 2$, and in turn for increasingly larger p until we find a factor.

TDFA always succeeds. In fact, the smallest prime factor of composite n is no larger than $\lfloor\sqrt{n}\rfloor$, so we do not need to go any further than this: if no factor has been found once we have tested for divisibility by all numbers smaller than this, we can conclude that n is prime.

However, TDFA is slow if n *has no small factors*, in the sense that its smallest factor $d > 1$ has digit length comparable with the digit length of $\sqrt{n}$ or n. Below, we discuss more precisely how slow this method is and what we mean by "digit length".

There are some optimizations that could be performed on TDFA. The simplest of these involves trying division only by $p = 2$ and all odd $p \leq \lfloor\sqrt{n}\rfloor$. Even with this and other optimizations, the time required to carry out TDFA grows exponentially as a function of the number of digits in n when n is written in some fixed base, making this an *exponential time algorithm*, at least when n has no small factors. Exponential time algorithms are notorious for being computationally infeasible for modestly large n (e.g. for a 1000-digit number).

To be more precise, suppose we use base 2, which is often done. We say that n is of *digit length* k if its standard base-2 representation has k digits. The number C_k of positive integers that are of digit length at most k is $2^k - 1$ (assuming that we "waste" only one such digit string to represent 0). Even if we only count odd numbers, C_k is comparable with 2^k for all large k, meaning that $C_k \geq c \cdot 2^k$, where $c > 0$ is independent of k. It follows that if we apply TDFA to a number n of digit length k, then the number of trial divisions is comparable with $2^{k/2}$.

There are more advanced factorization methods that are faster than TDFA, but all known ones are superpolynomial in the digit length, meaning that in the worst case scenario, the amount of calculation required to factorize a number of digit length k grows at a rate faster than k^m for any fixed m. Any growth rate that is subexponential is significantly better than an exponential growth rate but algorithms that require superpolynomial time are still considered slow and the underlying problem (e.g. factorization in this case) is still computationally infeasible once n is modestly large (as above).

The RSA cryptographic system, which we discuss later, depends on the difficulty of factorizing large numbers. If we used an integer in RSA that could be easily factorized, the system could be "cracked" once we factorize this integer, i.e. we could efficiently decrypt any encrypted message.

Although *factorizing* a large integer n into its prime factors may not be computationally feasible, it is easier to test n for *primality*. The aim of primality testing is

to determine whether a given integer is prime or not without having to factorize the integer into its prime decomposition.

There are both deterministic and probabilistic primality testing algorithms. A deterministic method gives a definitive conclusion in all cases, while a probabilistic method does not. The advantage of probabilistic methods is that they can give a near-certain conclusion much faster than the time required to give a definitive answer.

The idea of probabilistic primality testing is that we choose a random parameter a from a sample space that depends on the number n that we wish to test. There are two possible outcomes for the test: either n is definitely composite if it "fails" the test or its status remains undetermined if it "passes" the test. However, if it passes the test, the likelihood that n is prime increases and the likelihood that it is composite decreases, typically by at least some fixed factor (such as 2).

The strategy therefore is carry out such tests repeatedly for various different choices of a. If n fails even one of these tests, we stop testing because we have determined that n is composite. However, once n passes enough of these tests, we can confidently declare it to be prime. This declaration may be erroneous, but it is very unlikely to be so. Thus, we do not seek to formally prove that n is prime but rather we seek very strong evidence that it is prime. By contrast, if n is not prime, these methods will usually reveal this with certainty within a reasonable amount of time.

In this chapter, we discuss two commonly used probabilistic primality tests, namely the Fermat test and the Miller-Rabin test, which can be used to test if a given odd $n \in \mathbb{N}$ is prime.

7.1. Fermat's primality test.

Fermat's primality test is based on Fermat's little theorem. We first pick an integer b such that $\gcd(b, n) = 1$; we use the Euclidean algorithm to find $\gcd(b, n)$. Then, we reduce $b^{n-1} \pmod{n}$; because modern computer languages typically have a `mod` operator that performs such a reduction, the reduction is listed in the pseudocode below simply as computing $b^{n-1} \mod n$. If $b^{n-1} \not\equiv 1 \pmod{n}$, Fermat's little theorem says that n cannot possibly be prime.

The following pseudocode shows how to implement this algorithm to carry out a single test of this type.

```
Choose 1 < b < n − 1 at random
x := b^{n−1} mod n
if x = 1 then
    return "n is possibly prime"
else
    return "n is composite"
end if
```

We could add a preliminary check that $\gcd(b, n) = 1$. This can be verified quickly by the Euclidean algorithm. If this equation fails, then n is certainly not prime.

Note that we do not allow the choices $b = 1$ and $b = n - 1$ above (since n would pass the test regardless, in the latter case because -1 raised to an even power is equivalent to 1 modulo n).

If $b^{n-1} \not\equiv 1 \pmod{n}$, where $\gcd(b, n) = 1$, then n is not prime and we call b a *witness* (to the fact that n is not a prime). We now show that if there is any witness, then there are many.

Theorem 7.1. *Fix $N \in \mathbb{N}$. Suppose there exists a witness that N is composite. Then at least half the elements of Z_N^* are witnesses that N is composite.*

PROOF. Let B be the set of elements in $\mathbb{Z}_N^*$ that are not witnesses, that is, $b \in B$ means $b^{N-1} \equiv 1 \pmod{N}$. Clearly, $1 \in B$. If $a, b \in B$, then $(ab)^{N-1} = a^{N-1}b^{N-1} = 1 \cdot 1 \equiv 1 \pmod{N}$, and hence $ab \in B$. By Proposition 6.6, we conclude that B is a subgroup of $\mathbb{Z}_N^*$. Since by assumption there is at least one witness, B is a strict subgroup of $\mathbb{Z}_N^*$. Lemma 6.7 then shows that $|B| \leq |\mathbb{Z}_N^*|/2$, showing that at least half the elements of $\mathbb{Z}_N^*$ are not in B and are therefore witnesses. □

It can be shown that there always exist composite integers n such that the condition $b^{n-1} \equiv 1 \pmod{n}$ holds for any given a. Such numbers n are called *pseudoprimes*. Therefore, we run the test s times (for some suitably large s), with a different value of b each time, to increase our confidence in n being prime.

Example 7.2. We use Fermat's primality test to test whether 117 is prime or composite. Consider 3 as the base, noting that $\gcd(3, 117) = 1$. We reduce $3^{117-1} \pmod{117}$. Applying the fast exponentiation algorithm, we get $3^{116} \equiv 9 \pmod{117}$. Hence, 117 is composite.

Example 7.3. We use Fermat's primality test to test whether 113 is prime or composite. We pick 2 as the base, noting that $\gcd(2, 113) = 1$. We reduce $2^{113-1} \pmod{113}$. Applying the fast exponentiation algorithm, we get $2^{112} \equiv 1 \pmod{113}$. Hence, 113 is possibly prime (and indeed it is prime).

Definition 7.4. Let n be a composite integer and let $b \in \mathbb{N}$. If $\gcd(b, n) = 1$ and $b^{n-1} \equiv 1 \pmod{n}$, then n is called a pseudoprime to the base b.

There are composite integers where Fermat's little theorem holds for all choices of b.

Definition 7.5. A composite number n is a *Carmicheal number* if it is a pseudoprime to every base b such that $\gcd(b, n) = 1$.

Carmichael numbers are a particular problem for Fermat's primality test. If we are unlucky enough to pick a Carmichael number as our n, and we use this method to test primality, then we will probably eventually declare n to be prime even if we carry out many tests. Fortunately, it can be shown that Carmichael numbers do not have positive density in the sense that the proportion of numbers less than or equal to N that are Carmichael numbers tends to 0 as $N \to \infty$; see for instance [**7**].

7.2. The Miller-Rabin Test.

The Miller-Rabin test [**13**] is another probabilistic primality test. We will see that it effectively involves organising the calculations of Fermat's primality test in such a way that we get more benefit from our work. Crucially, there are no Miller-Rabin equivalents to Carmichael numbers, i.e. if a number passes the Miller-Rabin test to all possible bases, then it is definitely prime. In practice though, we usually only carry out enough tests to ensure high confidence in a number being prime, so this is also effectively a probabilistic primality test.

The test is based on the following theorem:

Theorem 7.6. *Suppose $n > 1$ is an odd integer, and let us write $n-1$ in the form $2^s t$, where $s, t \in \mathbb{N}$ and t is odd. If there exists an integer b, $0 < b < n$, such that*

$$b^t \not\equiv 1 \pmod{n} \quad \text{and} \quad b^{2^i t} \not\equiv -1 \pmod{n}, \qquad i \in \{0, 1, \ldots, s-1\}, \tag{7.7}$$

then n is composite.

Proof. We prove the theorem by contradiction. Suppose $0 < b < n$ is such that (7.7) holds, but that n is prime. In particular, b is coprime to n. Let us write $b \in M_i$ if $b^{2^i t} \equiv 1 \pmod{n}$. By Fermat's little theorem, $b \in M_s$ while, by assumption, $b \notin M_0$. Thus there exists some integer k, $0 < k < s$, such that $b \in M_k$ but $b \notin M_{k-1}$.

Because n is prime, $(\mathbb{Z}_n, +, *)$ is a field and so quadratic polynomials have at most two distinct roots there. Since 1 and -1 are roots of the polynomial $P(x) := x^2 - 1$, there are no other roots. (An alternative number theoretic proof of this fact is given in Lemma 7.10 below.) Since $b \in M_k$ but $b \notin M_{k-1}$, it follows that $b^{2^{k-1}t} \equiv -1 \pmod{n}$, contradicting (7.7), and so we are done. □

Based on the above, the following algorithm implements a single Miller-Rabin primality test.

(1) Write $n - 1 = 2^s t$ for some integers $s, t > 0$.
(2) Choose $1 < b < n - 1$ at random.
(3) Let $x \equiv b^t \pmod{n}$.
(4) If $x \equiv \pm 1 \pmod{n}$, then stop. The number n is possibly prime.
(5) Otherwise, repeatedly square x, at most $s - 1$ times. If you get -1 at any stage, stop. The number n is possibly prime.
(6) If after squaring $s - 1$ times, x is still not -1 mod n, then stop. The number n is composite.

As in the Fermat primality test, we could add a preliminary check that $\gcd(b, n) = 1$. If this equation fails, then n is certainly not prime.

Writing the test in pseudocode, we get the following:

```
Write n − 1 = 2^s t for some integer s and odd t
Choose 1 < b < n − 1 at random
x := b^t mod n
if x = 1 or x = n − 1 then
```

```
    return "n is possibly prime"
else
    for i from 1 to s − 1 do
        x := x^2 mod n
        if x = −1 then
            return "n is possibly prime"
        end if
    end for
    return "n is composite"
end if
```

Example 7.8. In this example, we use the Miller-Rabin test to determine whether 117 is prime or composite. First, we write $n-1 = 116 = 2^s t$, where $s = 2$ and $t = 29$. Now, we pick 3 as the base, and check that $\gcd(3, 117) = 1$ and that $1 < 3 < 116$.

We now calculate $x \equiv 3^{29} \equiv 9 \pmod{117}$. Since $x \not\equiv \pm 1$, we repeatedly square x, at most $s = 2$ times.

$$s = 1: \qquad x^2 \equiv 9^2 \equiv 81 \pmod{117}.$$
$$s = 2: \quad (x^2)^2 \equiv 81^2 \equiv 9 \pmod{117}.$$

Now, x has been squared 2 times without getting ± 1, so we conclude that 117 is composite.

Example 7.9. In this example, we use the Miller-Rabin test to determine whether 113 is prime or composite. First, we write $n-1 = 112 = 2^s t$, where $s = 4$ and $t = 7$. We pick 2 as the base, and check that $\gcd(2, 113) = 1$ and $1 < 2 < 112$.

We now calculate $x \equiv 2^7 \equiv 15 \pmod{113}$. Since $x \not\equiv \pm 1$, we repeatedly square x, at most $s = 4$ times.

$$s = 1: \quad x^2 \equiv 15^2 \equiv -1 \pmod{113}.$$

At this stage, we already stop because $x \equiv -1 \pmod{113}$. Thus 113 passes the Miller-Rabin test for base $b = 3$, and so 113 is possibly prime.

We say that $a \in \mathbb{Z}_N^*$ is a strong witness that N is composite (or simply a strong witness) if

(1) $a^t \not\equiv \pm 1 \pmod{N}$ and
(2) $a^{2^i t} \not\equiv -1 \pmod{N}$ for all $i \in \{1, \ldots, s-1\}$

For the following Lemma, we say $x \in \mathbb{Z}_N^*$ is a square root of 1 modulo N if $x^2 \equiv 1 \pmod{N}$.

Lemma 7.10. *If N is an odd prime, then the only square roots of 1 modulo N are $\pm 1 \pmod{N}$.*

Proof. Suppose $x^2 \equiv 1 \pmod{N}$ for some $x \in \{1, \ldots, N-1\}$. Then $0 = x^2 - 1 = (x+1)(x-1) \pmod{N}$, implying that $N \mid (x+1)$ or $N \mid (x-1)$ by Proposition 5.16. This can only possibly occur if $x \equiv \pm 1 \pmod{N}$. □

Theorem 7.11. *Let N be an odd number that is not a prime power. Then at least half the elements of $\mathbb{Z}_N^*$ are strong witnesses that N is composite.*

PROOF. Let $B \subseteq \mathbb{Z}_N^*$ denote the set of elements that are not strong witnesses. We define a set B' and show that: (1) B is a subset of B', and (2) B' is a strict subgroup of $\mathbb{Z}_N^*$. This suffices because by combining (2) and Lemma 6.7 we have that $|B'| \leq |\mathbb{Z}_N^*|/2$. Furthermore, by (1) it holds that $B \subseteq B'$, and so $|B| \leq |B'| \leq |\mathbb{Z}_N^*|/2$ as in Theorem 7.1. Thus, at least half the elements of $\mathbb{Z}_N^*$ are strong witnesses. (We do not claim that B is a subgroup of $\mathbb{Z}_N^*$.)

Note first that $-1 \in B$ since t is odd and so $(-1)^t \equiv -1 \pmod{N}$. Let $0 \leq i \leq s-1$ be the largest integer j for which there exists an $a \in B$ with $a^{2^j t} \equiv -1 \pmod{N}$, alternatively, i is the largest integer for which there exists an $a \in B$ with

$$(a^t, a^{2t}, \ldots, a^{2^s t}) = (*, \ldots, *, -1, 1, \ldots, 1)\,.$$

where the $(i+1)$st term is -1. Since $-1 \in B$ and $(-1)^{2^0 t} \equiv -1 \pmod{N}$, some such i exists. We fix i as above, and define

$$B' := \{a \mid a^{2^i t} \equiv \pm 1 \pmod{N}\}.$$

We now prove what that B' has the desired properties.

Claim 1: $B \subseteq B'$.

Let $a \in B$. Then either $a^t \equiv 1 \pmod{N}$ or $a^{2^j t} \equiv -1 \pmod{N}$ for some $0 \leq j \leq s-1$. In the first case, $a^{2^i t} = (a^t)^{2^i} \equiv 1 \pmod{N}$ and so $a \in B'$. In the second case, we have $j \leq i$ by choice of i. If $j = i$, then clearly $a \in B'$. If $j < i$, then $a^{2^i t} = (a^{2^j t})^{2^{i-j}} \equiv 1 \pmod{N}$ and $a \in B'$. Since a was arbitrary, this shows $B \subseteq B'$.

Claim 2: B' is a subgroup of $\mathbb{Z}_N^*$.

Clearly $1 \in B'$. Furthermore, if $a, b \in B'$ then

$$(ab)^{2^i t} = a^{2^i t} b^{2^i t} \equiv (\pm 1)(\pm 1) \equiv \pm 1 \pmod{N}\,,$$

and so $ab \in B'$. By Proposition 6.6, B' is a subgroup.

Claim 3: B' is a strict subgroup of $\mathbb{Z}_N^*$.

If N is an odd, composite integer that is not a prime power, then N can be written as $N = N_1 N_2$ with $N_1, N_2 > 1$ odd and $\gcd(N_1, N_2) = 1$. Appealing to the Chinese remainder theorem, let $a \leftrightarrow (a_1, a_2)$ denote the representation of $a \in \mathbb{Z}_N^*$ as an element of $\mathbb{Z}_{N_1}^* \times \mathbb{Z}_{N_2}^*$; that is $a_1 \equiv a \pmod{N_1}$ and $a_2 \equiv a \pmod{N_2}$. Take $a \in B'$ such that $a^{2^i t} \equiv -1 \pmod{N}$ (such an a must exist by the way we defined i), and $a \leftrightarrow (a_1, a_2)$. Since $-1 \leftrightarrow (-1, -1)$, we have

$$(a_1, a_2)^{2^i t} = (a_1^{2^i t}, a_2^{2^i t}) \equiv (-1, -1) \pmod{N}\,,$$

and so

$$a_1^{2^i t} \equiv -1 \pmod{N_1} \quad \text{and} \quad a_2^{2^i t} \equiv -1 \pmod{N_2}$$

Consider the element $b \in \mathbb{Z}_N^*$ with $b \leftrightarrow (a_1, 1)$. Then

$$b^{2^i t} \leftrightarrow (a_1, 1)^{2^i t} \equiv (a_1^{2^i t}, 1) = (-1, 1) \not\leftrightarrow \pm 1.$$

That is, $b^{2^i t} \neq \pm 1 \pmod{N}$ and so we have found an element $b \notin B'$. This proves that B' is a strict subgroup of $\mathbb{Z}_N^*$ and so, by Lemma 6.7, the size of B', and thus the size of B, is at most half the size of $\mathbb{Z}_N^*$. □

Notice that, although Theorems 7.1 and 7.11 are quite similar, there is one major difference between them. In the former theorem, Carmichael numbers are exceptional and so we need to assume that there is at least one witness before we can draw any conclusion. By contrast, in the latter theorem, we do not need to assume that a witness exists since it always does, at least as long as N is not a prime power.

We claim that, although it looks more complicated, the amount of computation for fixed n and base b is no greater (and may be less!) in Miller-Rabin than it is in Fermat. This is because an effective way of organising the computation of the reduced form of $b^{n-1} \bmod n$ is to first reduce $b^t \bmod n$ (using the fast exponentiation algorithm in Section 5.1), and then square and reduce s times; here, we assume that $n - 1 = 2^s t$ as before. If we do it this way, we can examine the numbers we get along the way and stop if either the initial reduced power $b^t \bmod n$ equals 1 or if it or any later reduced power equals -1.

Consider, for instance, $n = 73$. Here, $n - 1 = 72 = 2^3 \cdot 9$. If we want to compute b^{72}, we could compute the reduced form of b^{2^i} for $1 \leq i \leq 6$ and then $b^{72} = b^{2^6} \cdot b^{2^3}$: this involves seven multiply-and-reduce operations mod 73 (squaring six times and one other multiplication) but doing the calculation this way does not facilitate Miller-Rabin. If instead we compute and reduce $b^9 = b^1 \cdot b^{2^3} \bmod 9$, and then square and reduce b^9 three more times, we have done the same amount of work (i.e. seven multiply-and-reduce operations), but we computed b^9, b^{18}, and b^{36} along the way, so we can use these to carry out the Miller-Rabin test. If any of these reduced numbers equals $n - 1$ (i.e. if it is equivalent to $-1 \bmod n$), then we stop early and we have done less computation than the Fermat test requires us to do.

8. Primality Testing in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$

In this chapter, we investigate primality testing for $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$. It turns out that this readily reduces to primality testing for $\mathbb{Z}$, and so we can use Fermat's primality test or the Miller-Rabin test to solve this problem.

We first consider primality testing in $\mathbb{Z}[i]$. Recall that, up to multiplication by a unit, there are two main families of $\mathbb{Z}[i]$-primes: the $\mathbb{Z}$-primes p of the form $4n+3$, and conjugate prime pairs $\{\pi, \bar{\pi}\}$ where $\pi\bar{\pi} = p$ for a $\mathbb{Z}$-prime p of the form $4n+1$; in both cases, $n \in \mathbb{Z}$. The only other $\mathbb{Z}[i]$-prime – again up to multiplication by a unit – is $1+i$.

Thus, if we want to determine whether a nonzero Gaussian integer $z = a+bi$, $a, b \in \mathbb{Z}$, is possibly prime or not, we can use the following algorithm.

If either $a = 0$ or $b = 0$, then z is a $\mathbb{Z}[i]$-prime if and only if it is a unit multiple of a $\mathbb{Z}$-prime p that is congruent to 3 modulo 4. Thus, we should check this congruence and, if it is valid, then z is prime if and only if $|z|$ is a $\mathbb{Z}$-prime, something that we can investigate using the methods of the previous chapter.

If instead both a and b are nonzero, we calculate the norm $N(z)$. By Corollary 3.4 and Theorem 3.6, z is a $\mathbb{Z}[i]$-prime if and only if $N(z)$ is a $\mathbb{Z}$-prime (and necessarily $N(z)$ is either 2 or it is equivalent to 1 modulo 4).

In the following examples, we use Fermat's primality test and the Miller-Rabin test to investigate the primality of some Gaussian integers.

Example 8.1. We investigate whether $4+5i$ is prime or not in $\mathbb{Z}[i]$. First, note that $N(4+5i) = 41$. Because 41 is such a small number, we readily see that it is a $\mathbb{Z}$-prime, and so $4+5i$ is a $\mathbb{Z}[i]$-prime.

However, if we were uncertain of the $\mathbb{Z}$-primality of 41, we could for instance apply Fermat's primality test. Let us pick 3 as the base number (since $\gcd(41, 3) = 1$). By Fermat's little theorem, we need to calculate $3^{40} \pmod{41}$. Using fast exponentiation, we find $30^4 \equiv 1 \pmod{41}$. Hence, 41 is possibly prime by Fermat's primality test.

Example 8.2. We investigate whether 97 is prime or not in $\mathbb{Z}[i]$. Since $97 \not\equiv 3 \pmod{4}$, no further work is required: 97 is not a $\mathbb{Z}[i]$-prime.

Example 8.3. We investigate whether $16+9i$ is prime or not in $\mathbb{Z}[i]$. First, note that $N := N(16+9i) = 337$. Because 337 is not such a large number, and $\sqrt{337} < 19$, its primality is readily verified by checking that it is not divisible by any of the $\mathbb{Z}$-primes p for which $2 \leq p \leq 17$. Thus, $16+9i$ is a $\mathbb{Z}[i]$-prime.

However, if we were uncertain of the $\mathbb{Z}$-primality of 337, we could for instance apply the Miller-Rabin test. First, we write $N - 1 = 336 = 2^s t$, where $s = 4$ and $t = 21$. Let us pick 5 as the base, noting that $\gcd(5, 337) = 1$ and $1 < 5 < 336$. We reduce $x := 5^{21}$ modulo 337 to get $x \equiv 191 \pmod{337}$. As $x \not\equiv \pm 1$, we continue the algorithm,

squaring x at most $s = 4$ times.

$$\begin{aligned} s = 1: &\quad x^2 \equiv 191^2 \equiv 85 \pmod{337} \\ s = 2: &\quad x^4 \equiv 85^2 \equiv 148 \pmod{337} \\ s = 3: &\quad x^8 \equiv 148^2 \equiv -1 \pmod{337}. \end{aligned}$$

Hence, by the Miller-Rabin test, 337 is possibly prime, and $16 + 9i$ is possibly prime in $\mathbb{Z}[i]$. We could test with other bases to increase our confidence in the primality of $16 + 9i$.

Primality testing for the Eisenstein integers follows along similar lines. As for the Gaussian integers, there are two main families of $\mathbb{Z}[\omega]$-primes, up to multiplication by a unit: the $\mathbb{Z}$-primes p of the form $3n + 2$, and conjugate prime pairs $\{\pi, \bar{\pi}\}$ where $\pi\bar{\pi} = p$ for a $\mathbb{Z}$-prime p of the form $3n + 1$; in both cases, $n \in \mathbb{Z}$. The only other $\mathbb{Z}[\omega]$-prime—again up to multiplication by a unit—is $2 + \omega$.

Thus, if we want to determine whether a nonzero Eisenstein integer $z = a + b\omega$, $a, b \in \mathbb{Z}$, is possibly prime or not, we can use the following algorithm.

If either $a = 0$ or $b = 0$, then z is a $\mathbb{Z}[\omega]$-prime if and only if it is a unit multiple of a $\mathbb{Z}$-prime p that is congruent to 2 modulo 3. Thus, we should check this congruence and, if it is valid, then z is prime if and only if $|z|$ is a $\mathbb{Z}$-prime, something that we can investigate using the methods of the previous chapter.

If instead both a and b are nonzero, we calculate the norm $N(z)$. By Corollary 4.4 and Theorem 4.6, z is a $\mathbb{Z}[\omega]$-prime if and only if $N(z)$ is a $\mathbb{Z}$-prime (and necessarily $N(z)$ is either 3 or it is equivalent to 1 modulo 3).

9. Isomorphisms between certain quotient rings

In this chapter, we show that certain quotients of the rings $\mathbb{Z}$ and $\mathbb{Z}[i]$ are isomorphic, as are certain quotients of the rings $\mathbb{Z}$ and $\mathbb{Z}[\omega]$. This will later allow us to simplify the calculations for the RSA algorithm in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$.

9.1. Isomorphisms between quotients of $\mathbb{Z}$ and $\mathbb{Z}[i]$.

We will establish this isomorphism map based on **[5]**.

Recall that the units in $Z[i]$ are ± 1 and $\pm i$. Hence, given $a, b \in \mathbb{Z}$, the ideals $(a+bi)$, $(-a-bi)$, $(-b+ai)$ and $(b-ai)$ are all the same.

Fact 9.1. $\mathbb{Z}[i]/(a+bi) \cong \mathbb{Z}[i]/(-a-bi) \cong \mathbb{Z}[i]/(-b+ai) \cong \mathbb{Z}[i]/(b-ai)$.

Fact 9.2. $\mathbb{Z}[i]/(0) \cong \mathbb{Z}[i]$ *and* $\mathbb{Z}[i]/(1) \cong \{0\}$.

For our next theorem, we define $\mathbb{Z}_a[i]$ for an integer $a > 1$ to be the set of formal sums $x + yi$, $x, y \in \mathbb{Z}_a$, with the ring operations defined as in $\mathbb{Z}[i]$ except using the ring operations of $\mathbb{Z}_a$ in place of those of $\mathbb{Z}$.

Theorem 9.3. *If a is a positive integer larger than 1, then*

$$\mathbb{Z}[i]/(a) \cong \mathbb{Z}_a[i]\,.$$

PROOF. Define $\phi : \mathbb{Z}[i] \to \mathbb{Z}_a[i]$ by $\phi(x+yi) = [x]_a + i\,[y]_a$, where $[\,\cdot\,]_a$ denotes the equivalence class modulo a. This mapping is clearly a surjective ring homomorphism. Since $\phi(a) = [a]_a = [0]_a = 0$, a belongs to $\ker(\phi)$, and hence $(a) \subseteq \ker(\phi)$.

On the other hand, if $\phi(x+yi) = 0$, then both x and y are congruent to 0 modulo a, so we can write $x = ax'$ and $y = ay'$ for some integers x' and y'. Thus, $x+yi = ax'+ay'i = a(x'+y'i)$ lies in (a). Therefore, $\ker(\phi) = (a)$, which implies that $\mathbb{Z}[i]/(a) \cong \mathbb{Z}_a[i]$. □

Theorem 9.4. *If a is a positive integer larger than 1, then $\mathbb{Z}_a[i]$ is a field if and only if a is a prime in $\mathbb{Z}$ that is congruent to* 3 mod 4.

PROOF. Suppose first that $\mathbb{Z}_a[i]$ is a field. Since there are no zero divisors of the form $c+0i$, $c \in \mathbb{Z}_a \setminus \{0\}$, it follows that a must be prime. Moreover, a cannot be 2 since, if it were, then $(1+i)^2 \equiv 0 \pmod{a}$, giving a contradiction.

So, let a be an odd prime, and let $\mathbb{Z}_a[x]$ be the polynomial ring over $\mathbb{Z}_a$ as usual. Consider the usual ring homomorphism $\phi : \mathbb{Z}_a[x] \to \mathbb{Z}_a[i]$ given by $\phi(x) = i$. It is clear that $\ker(\phi) = (x^2+1)$. By the isomorphism theorem for rings, $\mathbb{Z}_a[i] \cong \mathbb{Z}_a[x]/(x^2+1)$, and this is a field if and only if x^2+1 is irreducible modulo a. This is equivalent to stating that there are no solutions to $x^2 \equiv -1 \pmod{a}$. This equation has solutions for a an odd prime if and only if $a \equiv 1 \pmod{4}$. Thus, we conclude that $a \equiv 3 \pmod{4}$.

Now, suppose that a is a prime congruent to 3 modulo 4, and consider again the ring homomorphism ϕ. Since x^2+1 is irreducible, the kernel (x^2+1) is a maximal ideal. Thus, $\mathbb{Z}_a[i]$ is a field. □

As a corollary to this theorem, we get the following result that we established previously.

Corollary 9.5. *If a is a positive integer larger than 1, then a is prime in $\mathbb{Z}[i]$ if and only if a is a prime in $\mathbb{Z}$ that is congruent to 3 modulo 4.*

We now come to our main isomorphism theorem for quotients of $\mathbb{Z}$ and $\mathbb{Z}[i]$.

Theorem 9.6. *If a and b are coprime integers, then $\mathbb{Z}[i]/(a+bi)$ is isomorphic to $\mathbb{Z}_{a^2+b^2}$.*

PROOF. Using Fact 9.1, we can assume without loss of generality that a and b are both positive integers. We observe that b is coprime to a^2+b^2, so b^{-1} exists in $\mathbb{Z}_{a^2+b^2}$. (Here, b^{-1} is an element of the inverse of the equivalence class of b modulo a^2+b^2.)

Since $a^2+b^2 \equiv 0 \pmod{a^2+b^2}$, we have $a^2 \equiv -b^2 \pmod{a^2+b^2}$, implying that $(ab^{-1})^2 \equiv -1$. We define $\psi : \mathbb{Z}[i] \to \mathbb{Z}_{a^2+b^2}$ by $\psi(x+yi) = x-(ab^{-1})y \pmod{a^2+b^2}$. Clearly, ψ is surjective and preserves addition.

We next show that ψ preserves multiplication. Let $\alpha = x+yi$ and $\beta = w+zi$. Then

$$\begin{aligned}
\psi(\alpha)\cdot\psi(\beta) = \psi(x+yi)\ \cdot\psi(w+zi) &= (x-ab^{-1}y)\cdot(w-ab^{-1}z)\\
&\equiv (xw)+a^2b^{-2}(yz)-ab^{-1}(xz+yw)\\
&\equiv (xw-yz)-ab^{-1}(xz+yw)\\
&= \psi((xw-yz)+(xz+yw)i)\\
&= \psi((x+yi)\cdot(w+zi))\\
&= \psi(\alpha\cdot\beta),
\end{aligned}$$

as required. Moreover, because $\psi(a+bi) = a-ab^{-1}b \equiv 0$, $(a+bi) \subseteq \ker(\psi)$.

For the converse containment, suppose $c+di \in \ker(\psi)$. Since $0 \equiv \psi(c+di) = c-ab^{-1}d$, we have $c \equiv ab^{-1}d$, and so $b(c+di) \equiv d(a+bi)$. We conclude that $b(c+di) \in (a+bi)$, and so $c+di \in (a+bi)$ because a and b are coprime. Thus, $\ker(\psi) \subseteq (a+bi)$, which means that $\ker(\psi) = (a+bi)$, and so $\mathbb{Z}[i]/(a+bi)$ is isomorphic to $\mathbb{Z}_{a^2+b^2}$. □

Again, this theorem leads to a result that we have already discussed in Chapter 3.

Corollary 9.7. *If a and b are coprime integers, then $a+bi$ is a prime in $\mathbb{Z}[i]$ if and only if a^2+b^2 is prime in $\mathbb{Z}$.*

Suppose we want to understand better the isomorphism $\phi : \mathbb{Z}[i]/I \to \mathbb{Z}/J$ implicitly constructed in the proof of Theorem 9.6, where $I = (a+bi)$ and $J = (a^2+b^2)$ are the ideals of interest in $\mathbb{Z}[i]$ and $\mathbb{Z}$, respectively. (ϕ is implicitly constructed via the equivalence between $\phi(z+I) = n+J$ and $\psi(z) = n+J$, where ψ is as in the proof of Theorem 9.6.)

Suppose first that we are given $c, d \in \mathbb{Z}$ and we want to compute e, where $\phi(c+di+I) = e+J$, or equivalently $\psi(c+di) = e+J$. Then, as in the above proof, e is any integer that is equivalent to $c-(ab^{-1})d \bmod (a^2+b^2)$.

There is also the inverse problem: suppose we are given e and we want to find c and d. Again, we want to solve the equivalence $c-(ab^{-1})d \equiv e \bmod (a^2+b^2)$. Taking $d=0$, we get the solution $c+di = e+0i$. All other solutions differ from this one by a (Gaussian integer) multiple of $a+bi$.

Example 9.8. Taking $a+bi = 3+2i$ and $I=(3+2i)\subset\mathbb{Z}[i]$, Theorem 9.6 says that $\mathbb{Z}[i]/I \cong \mathbb{Z}_{3^2+2^2} = \mathbb{Z}_{13} = \mathbb{Z}/J$, where $J=(13)\subset\mathbb{Z}$. Writing $\psi:\mathbb{Z}[i]\to\mathbb{Z}/J$ as before, suppose we want to find $\psi(7+4i)$. By the proof of Theorem 9.6,

$$\psi(7+4i) = 7-(3\cdot 2^{-1})\cdot 4 + J = 7-(3\cdot 7)\cdot 4 + J = 1+J\,,$$

and so $\psi(7+4i) = 1+J$.

Suppose next that we want to find (x,y) such that $\psi(x+yi) = 2+J$. Hence, we want to solve the equation

$$x-(3\cdot 7)\cdot y \equiv 2 \pmod{13}\,.$$

Letting $y=0$, we get the solution $x=2$. Hence, $\psi(2+0i) = 2+J$. The other solutions, of course, are the Gaussian integers that differ from 2 by a Gaussian multiple of $3+2i$, so examples include $-1-2i$, $5+2i$, $8+4i$, $14+8i$, $3i$, etc.

9.2. Isomorphisms between quotients of $\mathbb{Z}$ and $\mathbb{Z}[\omega]$.

The units in $Z[\omega]$ are ± 1, $\pm\omega$ and $\pm\omega^2$. Hence, we know that for integers a and b, the ideals $(a+b\omega)$, $(-a-b\omega)$, $(a\omega+b\omega^2)$, $(-a\omega-b\omega^2)$, $(a\omega^2+b)$ and $(-a\omega^2-b)$ are the same.

Fact 9.9. $\mathbb{Z}[\omega]/(a+b\omega) \cong \mathbb{Z}[\omega]/(-a-b\omega) \cong \mathbb{Z}[\omega]/(a\omega+b\omega^2) \cong \mathbb{Z}[\omega]/(-a\omega-b\omega^2) \cong \mathbb{Z}[\omega]/(a\omega^2+b) \cong \mathbb{Z}[\omega]/(-a\omega^2-b)$

Fact 9.10. $\mathbb{Z}[\omega]/(0) \cong \mathbb{Z}[\omega]$ *and* $\mathbb{Z}[\omega]/(1) \cong \{0\}$.

The following theorem has almost the same proof as its Gaussian equivalent, Theorem 9.3, so we omit the proof.

Theorem 9.11. *If a is a positive integer larger than 1, then*

$$\mathbb{Z}[\omega]/(a) \cong \mathbb{Z}_a[\omega]$$

Theorem 9.12. *If a is a positive integer larger than 1, then $\mathbb{Z}_a[\omega]$ is a field if and only if a is a prime in $\mathbb{Z}$ that is congruent to 2 modulo 3.*

PROOF. Suppose first that $\mathbb{Z}_a[\omega]$ is a field. Since there are no zero divisors of the form $c+0\omega$, $c\in\mathbb{Z}_a\setminus\{0\}$, it follows that a must be prime. Moreover, a cannot be 3 since, if it were, then $(2+\omega)^2 \equiv 0 \pmod a$, giving a contradiction.

So, let a be a prime $\neq 3$, and let $\mathbb{Z}_a[x]$ be the polynomial ring over $\mathbb{Z}_a$ as usual. It is clear that $\ker(\phi) = (x^2-x+1)$. By the isomorphism theorem for rings, $\mathbb{Z}_a[\omega] \cong \mathbb{Z}_a[x]/(x^2-x+1)$, and this is a field if and only if x^2-x+1 is irreducible modulo a.

This is equivalent to stating that there are no solutions to $x^2-x\equiv -1 \bmod a$. As we showed in the proof of Theorem 4.6, this equivalence has solutions for a prime a, $a\neq 3$, if and only if $a\equiv 1 \pmod 3$. Thus, we conclude that $a\equiv 2 \pmod 3$.

Now, suppose that a is a prime congruent to 2 modulo 3, and consider again the ring homomorphism ϕ. Since $x^2 - x + 1$ is irreducible, the kernel $(x^2 - x + 1)$ is a maximal ideal. Thus, $\mathbb{Z}_a[\omega]$ is a field. □

As a corollary to this theorem, we get the following result that we established in Chapter 4.

Corollary 9.13. *If a is a positive integer larger than 1, then a is prime in $\mathbb{Z}[\omega]$ if and only if a is a prime in $\mathbb{Z}$ that is congruent to 2 modulo 3.*

Theorem 9.14. *If a and b are coprime integers, then $\mathbb{Z}[\omega]/(a + b\omega)$ is isomorphic to $\mathbb{Z}_{a^2-ab+b^2}$.*

PROOF. Since $(z) = (zu)$ whenever u is a unit, and multiplication by a unit corresponds geometrically to rotation by any multiple of 60 degrees, we can assume without loss of generality that a and b are both positive integers. We observe that b is coprime to $a^2 - ab + b^2$, so b^{-1} exists in $\mathbb{Z}_{a^2-ab+b^2}$. (Here, b^{-1} is an element of the inverse of the equivalence class of b modulo $a^2 - ab + b^2$.)

Since $a^2 - ab + b^2 \equiv 0 \pmod{a^2 - ab + b^2}$, we have $a^2 - ab \equiv -b^2 \pmod{a^2 - ab + b^2}$, implying that $(ab^{-1})^2 \equiv -1 + ab^{-1}$. We define $\psi : \mathbb{Z}[\omega] \to \mathbb{Z}_{a^2-ab+b^2}$ by $\psi(x + y\omega) = x - (ab^{-1})y \pmod{a^2 - ab + b^2}$. Clearly, ψ is surjective and preserves addition.

We next show that ψ preserves multiplication. Let $\alpha = x + y\omega$ and $\beta = u + v\omega$. Since

$$\begin{aligned}
\psi(\alpha) \cdot \psi(\beta) = \psi(x + y\omega) \cdot \psi(u + v\omega) &= (x - ab^{-1}y) \cdot (u - ab^{-1}v) \\
&\equiv xu - ab^{-1}xv - ab^{-1}yu + (ab^{-1})^2yv \\
&\equiv xu - yv - ab^{-1}(xv + yu - yv) \\
&= \psi(xu - yv) - ab^{-1}(xv + yu - yv) \\
&= \psi((xu - yv) + (xv + yu - yv)\omega) \\
&= \psi(\alpha \cdot \beta),
\end{aligned}$$

as required. Moreover, because $\psi(a + b\omega) = a - ab^{-1}b \equiv 0$, $(a + b\omega) \subseteq \ker(\psi)$.

For the converse containment, suppose $c + d\omega \in \ker(\psi)$. Since $0 \equiv \psi(c + d\omega) = c - ab^{-1}d$, we have $c \equiv ab^{-1}d$, and so $b(c + d\omega) \equiv d(a + b\omega)$. We conclude that $b(c + d\omega) \in (a + b\omega)$, and so $c + di \in (a + b\omega)$ because a and b are coprime. Thus, $\ker(\psi) \subseteq (a + b\omega)$, which means that $\ker(\psi) = (a + b\omega)$, and so $\mathbb{Z}[i]/(a + b\omega)$ is isomorphic to $\mathbb{Z}_{a^2-ab+b^2}$. □

Again, this theorem leads to a result that we have already proved.

Corollary 9.15. *If a and b are coprime integers, then $a + b\omega$ is a prime in $\mathbb{Z}[\omega]$ if and only if $a^2 - ab + b^2$ is prime in $\mathbb{Z}$.*

Suppose we want to understand better the isomorphism $\phi : \mathbb{Z}[\omega]/I \to \mathbb{Z}/J$ implicitly constructed in the proof of Theorem 9.14, where $I = (a + b\omega)$ and $J = (a^2 + b^2)$ are the ideals of interest in $\mathbb{Z}[\omega]$ and $\mathbb{Z}$, respectively. (ϕ is implicitly constructed via the

equivalence between $\phi(z+I) = n+J$ and $\psi(z) = n+J$, where ψ is as in the proof of Theorem 9.14.)

Suppose first that we are given $c, d \in \mathbb{Z}$ and we want to compute e, where $\phi(c+d\omega+I) = e+J$, or equivalently $\psi(c+d\omega) = e+J$. Then e is any integer that is equivalent to $c-(ab^{-1})d \bmod (a^2-ab+b^2)$. Solving for e given (c,d), or vice versa, is then handled in a manner very similar to the Gaussian case.

10. Public vs Private Key Encryption: overview

In this chapter, we establish what public key encryption is and how it differs from private key encryption, see [**8**]. We start by defining the correctness requirement that any encryption scheme must satisfy.

Definition 10.1. Let M be the set of all plaintext messages m and K be the set of all possible keys k. For a cipher system to be correct, we require $D_k(E_k(m)) = m$, $\forall m \in M$, $\forall k \in K$.

In other words, encrypting and then decrypting a message always yields the original message.

Private key encryption enables two parties to communicate with each other. The communicating parties share a private key for encrypting and decrypting messages. Private key cryptosystems are symmetric cryptosystems: here, *symmetric* means that the shared key is used for both encryption and decryption.

In contrast to symmetric private key encryption, we have asymmetric public key encryption. *Asymmetric* here means that we use two different keys: a public key for encryption and a private key for decryption.

A party wishing to join a public key encryption system generates a pair of keys: a public key and a private key. The public key is published and allows anyone who knows it to encrypt messages. The private key, however, is kept secret and is known only to the party that generated the key. This private key is used for decryption and enables the party to recover the original message from the ciphertext. Only the person that knows the private key can decrypt messages, but anyone who knows the public key can encrypt messages. That is, when two parties want to communicate with each other, they each generate their own public key and private key and they exchange their public keys.

One of the most important public key encryption systems is the RSA algorithm, which we will discuss in the next chapter.

11. RSA Algorithm in $\mathbb{Z}$

The RSA algorithm **[14]** is named after Ronald Rivest, Adi Shamir and Leonard Adleman, who discovered it in 1977. Up until now, the RSA algorithm has been one of the most important and best-known asymmetric encryption methods. The security of this public key cryptosystem is based on the difficulty of decomposing large numbers into their prime factors.

To encrypt or decrypt messages using the RSA algorithm, we perform the following steps:

First, we pick two large primes p and q and calculate their product $N = pq$. We then calculate $\phi(N)$, where $\phi(N)$ is the Euler ϕ-function. Let the *encryption exponent* e be chosen so that $1 < e < \phi(N)$ and $\gcd(e, \phi(N)) = 1$. The pair (N, e) is the public key. The *decryption exponent* d is the inverse of $e \mod \phi(N)$, i.e. $ed \equiv 1 \pmod{\phi(N)}$. The pair (N, d) is the private key.

To encrypt a numerical message M, where $0 \leq M < N$, we compute $C \equiv M^e \pmod{N}$. To decrypt the encrypted message C, we calculate $M \equiv C^d \pmod{N}$.

To verify the correctness of the RSA algorithm, we need to verify the correctness requirement in Definition 10.1, i.e. we need to check that $(M^e)^d \equiv M \pmod{N}$. This follows from the equivalence $ed \equiv 1 \pmod{\phi(N)}$, as we now show.

The key fact is that $ed - 1$ is a multiple of $\phi(N) = (p-1)(q-1)$, and so is a multiple of both $p - 1$ and $q - 1$. We want to show that $M^{ed} - M$ is divisible by N, which is equivalent to showing that it is divisible by both p and q.

If M is a multiple of p, then it is trivial that $M^{ed} - M$ is divisible by p. Otherwise, M is a unit mod p and so, by Theorem 5.43 (Fermat's Little Theorem), $M^{p-1} \equiv 1 \pmod{p}$. It follows that $M^{ed-1} \equiv 1 \pmod{p}$, and so $M^{ed} - M$ is a multiple of p. The fact that $M^{ed} - M$ is a multiple of q follows similarly.

Note that the above shows that we could slightly improve our method and work with smaller exponents: we need only that $ed-1$ is a multiple of both $p-1$ and $q-1$, and so it should be a multiple of the least common multiple $\lambda(n)$ of $p-1$ and $q-1$. Note that $\phi(n)$ is clearly a multiple of $\lambda(n)$ but $\lambda(n)$ is smaller since 2 is a common factor of $p-1$ and $q-1$ (p and q are "large primes", and so certainly odd). Thus, we can compute d as the inverse of $e \mod \lambda(n)$ rather than mod $\phi(n)$. This makes the algorithm more efficient (although typically only a little more efficient because $\gcd(p-1, q-1)$ is usually much smaller than p or q), but makes no significant theoretical difference, so we stick with the $\phi(n)$-variant that was used in the original RSA paper.

The following pseudocode shows one way to generate the public key (N, e) and the private key (N, d) for two given primes p and q.

$N := p \cdot q$
$phi := (p-1) \cdot (q-1)$
$e := 2$
while $e < phi$ **do**

```
    if gcd(e, phi) = 1 then
        break
    else
        e := e + 1
    end if
end while
k := 2
d := (k · phi + 1)/e
while k < (phi − 1) do
    if d is an integer then
        break
    else
        k := k + 1
    end if
end while
```

The above method is the easiest one to write down but let us note that the computation of d can actually be made much more efficient by using "backward induction" in the Euclidean algorithm as in the proof of Theorem 5.11 to prove Bezout's identity. The idea is that, because e and $\phi(N)$ are coprime, we can find integers x, y such that $ex + \phi(N)y = 1$, and now x is essentially the desired decryption exponent, although we might want to add/subtract some multiple of $\phi(N)$ to x in order to get a decryption exponent d satisfying $0 < d < \phi(N)$.

With our public and private keys in hand, the task of encryption and decryption is straightforward. Specifically, to encrypt a numerical message $0 \leq M < N$, we calculate:

$$c \equiv M^e \pmod{N}.$$

To decrypt a given ciphertext $0 \leq c < N$, we calculate:

$$m \equiv M^d \pmod{N}.$$

As usual, all exponentiation should be done using the fast exponentiation algorithm in Section 5.1.

Example 11.1. In this example, we call our two participating parties Alice and Bob. Bob wants to send Alice a message using the RSA algorithm. Alice must first generate a pair of keys to enable the communication.

Alice picks two primes $p = 37$ and $q = 53$ and calculates their product $N = 37 \cdot 53 = 1961$. She also calculates $\phi(1961) = \phi(37)\phi(53) = 36 \cdot 52 = 1872$. Alice then picks her encryption exponent e such that $\gcd(e, 1872) = 1$ and $1 < e < 1872$; Alice picks $e = 5$. Now, Alice publishes her public key $(N, e) = (1961, 5)$, allowing Bob to send her an encrypted message.

Alice also generates her private key so that she can decrypt Bob's message. Therefore, she solves the equation $5d \equiv 1 \pmod{1872}$ for d to get $d = 749$. Alice keeps her private key $(N, d) = (1961, 749)$ a secret.

Bob wants to send the message $M = 21$ to Alice. Thus, he calculates $C \equiv 21^5 \equiv 1299 \pmod{1961}$. Bob sends $C = 1299$ to Alice.

Alice decrypts this message to determine the original message. She therefore calculates $M \equiv 1299^{749} \pmod{1961}$ and finds that Bob's message was $M = 21$.

Lastly, note that, although a typical N used for encryption is fairly large—the minimum recommendation for N is currently at least 2048 "bits", where a bit is a base-2 digit—it is probably much shorter than any real-life message that we wish to encrypt. Consequently, we must break the full message M into blocks of bit-length at most L, where $2^L \leq N$, and then it is these blocks that we encode, and which the receiver decodes and reassembles into the full message. In fact, ensuring that there is a gap between L and $\log_2 N$ is typical so that we can pad the hash with some random bits for security purposes.

11.1. RSA Digital Signature.

In the previous example, we illustrated how Bob can send a message to Alice using the RSA algorithm. Now, we discuss the case where Alice sends Bob a message. Rather than being concerned with encrypting the message, suppose she wants to allow Bob to verify that the message came from Alice. This time, she encrypts with her own private key, rather than Bob's public key.

Suppose as before that Alice has the public key (N, e) and the private key (N, d). Alice wants to send the numerical message M, $0 \leq M < N$. She can calculate $S \equiv M^d \pmod{N}$, where S is the signature. Then, Alice sends the message (S, M) to Bob. Bob has Alice's public key, so he can then calculate $S^e \pmod{N}$. Since $S \equiv M^d \pmod{N}$, $S^e \equiv (M^d)^e \equiv M \pmod{N}$.

If $S^e \equiv M \pmod{N}$, this gives a valid signature, while $S^e \not\equiv M \pmod{N}$ gives an invalid signature. Thus, if Bob verifies a valid signature, he can be confident that the message comes from Alice.

For various reasons, including efficiency, the full message is not signed in this manner. Instead of breaking the message M into blocks and signing each one, a so called *hashing function* is first applied to replace the long message by a much shorter *message digest*. Mathematically, this hashing function is $f_N : \mathbb{N} \to \mathbb{N}$, whose values are all less than our desired integer $N \in \mathbb{N}$. In practice, f_N takes a message of arbitrarily long bit-length and replaces it by a message digest of bit-length at most L, where $2^L \leq N$; as for encryption, it is common to ensure that there is a gap between L and $\log_2 N$ is typical so that we can pad the digest for security purpose.

Hashing and padding the message, and then encrypting the much shorter digest (using Alice's private key), is typically much faster than encrypting the full message, making this method is more efficient.

Suppose now that Alice has public and private keys (N_A, e_A) and (N_A, d_A), respectively, and Bob's corresponding keys are (N_B, e_B) and (N_B, d_B). Using a mutually agreed hashing function f_N, Alice computes $f_N(M)$ and raises it to the power d_A to get the digital signature S. Separately, Alice raises each padded block of the message M to

the power e_B and sends them to Bob. Once Bob receives, decrypts, and reassembles the message M from these encrypted blocks, he can compute $f_N(M)$ and compare it with S^{e_A}. If they match, he can read the message, safe in the knowledge that it is from Alice.

A third party can also use the public key to compute $S^{e_A} = f_N(M)$, so we must make sure that f_N is a so-called *cryptographic hashing function* (CFA). One important property that a CFA f_N must have is that $f_N(M)$ should give essentially no useful information about M. A CFA should also have the property that for two randomly chosen messages M and M', the probability that $f_N(M) = f_N(M')$ is extremely small: this ensures that Alice will almost certainly produce different digital signatures for different messages so that the signature is useless to a third party.

Such CFAs (with acronyms such as SHA and DSA) exist and are commonly used in cryptography, but we will not discuss them further here. Note that f_N does not have to be different for different N: it suffices that it corresponds to a bit-length that is less than $\log_2 N$, typically with a gap in this difference to allow padding. Thus, if we choose N to have bit-length at least 2048, we could for instance use the common SHA-1 hashing method since it gives hash values with at most 160 base-2 digits.

Example 11.2. In this toy example, we illustrate how the RSA algorithm works, focusing on the digital signatures, but not including breaking a long message into blocks or hashing.

Alice wants to send Bob a message $M = 15$ and wants to ensure that Bob knows that the message is from her. She computes a number E, $0 \leq E < N_B$, that is equivalent to $M^{e_B} \bmod N_B$. She also computes a signature S, as detailed below, and then transmits (E, S) to Bob.

Let us assume that Alice's keys are $(N_A, e_A) = (1961, 5)$ and $(N_A, d_A) = (1961, 749)$, as in Example 11.1. Alice calculates $15^{749} \equiv 1054 \pmod{1961}$, and so $S = 1054$.

Bob decrypts the message by computing M as the number satisfying $0 \leq M < N_B$ that is equivalent to $E^{d_B} \bmod N_B$. Bob verifies the signature by computing $S^{e_A} \bmod N_A$. Since $15 \equiv 1054^5 \bmod 1961$, the signature matches and the message is likely to be from Alice.

Note that in the above example, an eavesdropping third party could decrypt the signature to get the message, so the procedure is defective. Thus hashing is an essential step to ensure that such signed communication preserves the secrecy of the message.

An alternative method, described in the original RSA paper **[14]**, uses Alice's private key to compute the signed message S as above, and then uses Bob's public key to encrypt S, yielding a twice-encrypted message T that Alice sends to Bob. Bob uses his own private key to recover S, and then Alice's public key to recover the original message. This method also provides a secure communication and a verifiable signature, but the message digest method mentioned earlier is the standard nowadays.

12. Attacking RSA

Now that we have explained how the RSA algorithm works, we will discuss how this public key encryption scheme can be attacked.

12.1. Factorization.

A common way to attack the RSA algorithm is by trying to factorize the modulus $N = pq$ into its prime decomposition.

In the following proposition, see **[1]**, we show that knowing the private key d and factoring N are equivalent. Thus, as soon as the factorization of N is known, d can be determined and vice versa.

Proposition 12.1. *Let (N, e) be an RSA public key. Given the private key d, we can efficiently factor the modulus $N = pq$. Conversely, given the factorization of N, we can efficiently recover d.*

PROOF. We assume first that we know the factorization of N. The factorization of N gives us $\phi(N)$. As e is known, we can easily solve the equation $ed \equiv 1 \pmod{\phi(N)}$ to determine d; see the pseudocode and subsequent discussion in Chapter 11.

To prove the converse, we show knowing d allows us to factorize N. Given d, we compute $k = de - 1$. By definition of e and d, we know that k is a multiple of $\phi(N)$. Since $\phi(N)$ is even, we have $k = 2^t r$ for some odd r and $t \geq 1$. We know that $g^k = 1$ for every $g \in \mathbb{Z}_N^*$, and therefore $g^{k/2}$ is a square root of unity modulo N. By the Chinese remainder theorem, 1 has four square roots modulo $N = pq$. Two of these square roots, namely ± 1, do not help us to factorize N. The other two are $\pm x$, where x satisfies $x \equiv 1 \bmod p$ and $x \equiv -1 \bmod q$. Using either of these last two square roots, we can factorize N, since $\gcd(x - 1, N)$ equals one of the two prime factors.

Thus, the idea is that we randomly choose $g \in \mathbb{Z}_N^*$. With probability at least $1/2$ (over the choice of g), one of the elements in the sequence $g^{k/2}, g^{k/4}, \ldots, g^{k/2^t} \bmod N$ is a square root of unity that reveals the factorization of N. So, by investigating this sequence for sufficiently many randomly chosen g, we can factorize N with probability as close to 1 as we wish. □

We will now discuss some commonly used methods for factorization.

12.1.1. Fermat's Method. Fermat's factorization method derives from the idea that we can write integers as sums or differences of squares.

Proposition 12.2. *Let n be a positive odd integer. There is a 1-to-1 correspondence between factorizations of n in the form $n = ab$, where $a \geq b > 0$, and representations of n in the form $t^2 - s^2$, where s and t are nonnegative integers. The correspondence is given by the equations*

$$t = \frac{a+b}{2}, \quad s = \frac{a-b}{2}, \quad a = t + s, \quad b = t - s.$$

PROOF. Given such a factorization, we can write $n = ab = ((a+b)/2)^2 - ((a-b)/2)^2$, so we obtain the representation as a difference of two squares. Conversely, given $n = t^2 - s^2$ we can factor the right side as $(t+s)(t-s)$. The equations in the proposition explicitly give the 1-to-1 correspondence between the two ways of writing n. □

If $n = ab$, where a and b are close together, then $s = (a-b)/2$ is small, and so t is only slightly larger than $\sqrt{n}$. In this case, we can find a and b by trying all values for t starting with $\lceil \sqrt{n} \rceil$ until we find one for which $t^2 - n = s^2$ is a perfect square.

When the two factors a and b are relatively close, Fermat's factorization method can be fast, but when they are further apart, this method is not efficient enough to use. Assuming that n is not a perfect square, the following pseudocode illustrates how to use Fermat's factorization method to decompose an odd integer n into its prime factors p and q.

$x := \lceil \sqrt{n} \rceil$
$y := \sqrt{x^2 - n}$
while y is not a perfect square **do**
    $x := x + 1$
end while
Print(n equals (x-y) * (x+y)

We will illustrate Fermat's factorization method with an example.

Example 12.3. In this example, we let $n = 481$. Therefore, $x = \lceil \sqrt{481} \rceil = 22$.

$$y = \sqrt{22^2 - 481} = \sqrt{3}$$
$$y = \sqrt{23^2 - 481} = \sqrt{48}$$
$$y = \sqrt{24^2 - 481} = \sqrt{95}$$
$$y = \sqrt{25^2 - 481} = \sqrt{144} = 12$$

Hence, $n = (25-12)(25+12) = 13 \cdot 37$.

12.1.2. Pollard Rho Method of Factorization. The Pollard rho method of factorization was invented by John Pollard in 1975, see **[12]**. For this method, we choose a polynomial f whose iterates generate a large set of integers modulo n. It is important to choose a polynomial f that maps $\mathbb{Z}/n\mathbb{Z}$ to itself in a "random" way; in particular, linear polynomials should not be used, as the mapping is "not random enough". Furthermore, no polynomial that defines an injective map on $\mathbb{Z}/n\mathbb{Z}$ should be used. Any polynomial of degree at least 2 can be used, and typically, polynomials of the form $x^2 + c$, where $c \neq 0, -1, -2$ are used.

We pick a starting value x_0 and calculate the sequence of iterated f-values

$$f(x_0), f(f(x_0)), f(f(f(x_0))), \ldots$$

For brevity, we write this sequence as

$$x_1, x_2, x_3, \ldots$$

i.e. $x_i = f(x_{i-1})$ for every $i \in \mathbb{N}$. Often, x_0 is taken to equal 2, but other starting values can be used.

We are interested in the above sequence modulo both n and p, where p is an (unknown) prime factor of a composite number n. More generally, it usually suffices to find any factor d of n satisfying $1 < d < n$. If d is not prime and we need a prime factor, we can simply repeat the process, but with n replaced by the smaller number d. Clearly, such a repeated process gives a prime after a finite number of steps.

We would like to find indices $k < i$ such that $x_i \equiv x_k \bmod p$, but $x_i \not\equiv x_k \bmod n$. If we find this, then $\gcd(x_i - x_k, n) = d > 1$. Since d is a non-trivial divisor of n, we have factorized n, as required. Since p is unknown, in practice we reverse the process: we calculate $\gcd(x_i - x_k, n)$ for various values of $k < i$, and keep going until this gcd is different from 1.

The "rho" in the name of this method reflects the fact that the initial terms in the sequence $x_1, x_2, x_3, \ldots$ may be non-periodic, giving the "tail" of the Greek letter rho, but eventually it becomes periodic, giving the "circle" of rho. If $\gcd(x_i - x_k, n) = 1$ for our current choice of integers $0 \leq k < i$, it may be because k is too small (i.e. it is in the tail of the rho) or it may be because we have not yet tried a large enough i (i.e. we haven't gone "all the way around the circle"). Thus, it is important to try various consecutive values of i for each k, but also to let k increase if we fail to find an i with $\gcd(x_i - x_k, n) > 1$ for many consecutive values of i.

More details on Pollard's rho method of factorization, often called the "Monte Carlo" method of factorization, can be found in [**12**]. There are different variations of the algorithm based around how we choose the various values of i and k to test.

For the remainder of this chapter, we will discuss a variation due to Brent, which we will call the *Pollard-Brent method*, and we will give an example. A detailed comparison of the Pollard-Brent method with the original Pollard rho method can be found in [**2**].

Every variation of Pollard's rho method is based on the following observation:

Let $S = 0, 1, 2, \ldots, N-1$, where N is a (large) integer. Pseudorandom numbers are often generated by an iteration of the form

$$x_{i+1} = f(x_i),$$

where $f : S \to S$, and $x_0 \in S$, are as described above. Since S is finite, there exists $\mu \geq 0$, and $\nu \geq 1$ such that

$$x_{\mu+\nu} = x_\mu \,,$$

and so it follows that

$$x_{k+\nu} = x_k \,, \qquad \text{for all } k \geq \mu \,. \tag{12.4}$$

The minimal such ν is called the period of the sequence (x_i), but in general the sequence is not actually periodic until we have passed an initial non-periodic part of the sequence.

Above, we should think of S as being the residue classes modulo some number j. The numbers ν and μ are likely to be smaller when j is an (unknown) factor d of a composite number n than they are for $j = n$, so we can likely find i and k such that

$x_i \equiv x_k \bmod d$, but $x_i \not\equiv x_k \bmod n$. If, however, the period is the same for n as for its factors $1 < d < p$, then the method has failed *for this choice of starting value* x_0. In that case, we can try again with a different starting value.

The Pollard-Brent variation is based on using powers of 2. More precisely, we choose $k = 2^l - 1$ for some fixed l, and then try out i for all $2^l \leq i < 2^{l+1}$, calculating $m = |x_i - x_k|$ and $\gcd(m, n) = d$. Note that we lose nothing by not looking at smaller values of i for a given k, since if $x_j \equiv x_k \bmod p$ for some $0 \leq j < k = 2^l - 1$, then the same equivalence holds with j replaced by $i = 2k - j$ since $i - k = k - j$. Note also that in this case, $k < i < 2^{l+1}$.

The relative efficiency of this method is a consequence of the fact that there is no need to minimise μ in (12.4): we simply want to go far enough into the sequence so that we are in the periodic part. Thus, the value of k jumps almost to the next power of 2 whenever we have computed that far into the sequence. By contrast, it is important to find the period ν (working modulo an unknown factor $d > 1$) rather than some multiple of it because a multiple of it might also be a multiple of the period modulo n causing the method to fail. For that reason, we try every value of i from $k + 1 = 2^l$ up to $2k + 1 = 2^{l+1} - 1$, corresponding to a fresh search for any period between 1 and k, rather than looking for periods longer than we looked for with our previous values of k.

Since the algorithm revolves around finding $k < i$ such that $\gcd(x_i - x_k, n) > 1$, we can force the inequalities $0 \leq x_i, x_k < n$ by replacing these numbers with representatives modulo n. This leaves $\gcd(x_i - x_k, n)$ unchanged since f is a polynomial so theoretically it makes no difference. Practically, however, this is of great benefit in terms of efficient calculations since if we do not do this, the values in this sequence get very large rather quickly (for a typical polynomial of degree at least 2).

Here are the detailed steps of the Pollard-Brent algorithm. It involves storing only two values of the sequence (x_j) at any one time. We call these two values x and y. Mostly, x will correspond to x_k above, and y will correspond to x_i, but see also the commentary below the algorithm where the "computer-friendly" algorithm is interpreted in more traditional mathematical language.

(1) Given the number n that we are trying to factorize, choose a polynomial f of degree at least 2, and treat f as a map from $\mathbb{Z}_n$ to $\mathbb{Z}_n$.
(2) Let $l := 0$ and $i := 0$. Choose an initial value $y \in \mathbb{Z}_n$.
(3) Let $k := i$ and $x := y$.
(4) Increase i by 1 and replace y by $f(y)$.
(5) Let $m := |y - x| \bmod n$ and $d := \gcd(n, m)$.
(6) If $d > 1$, stop and report d as a factor of n.
(7) If $i < 2^{l+1} - 1$, go back to Step 4.
(8) Increase l by 1 and go back to Step 3.

Step 2 initialises the variable l that keeps count of our powers of 2. It also assigns an initial value, previously called x_0, to the variable y. In fact, since $i = 0$, we initially have $y = x_i = x_0$. We will have $y = x_i$ throughout the algorithm.

Step 3 is the initialization of the "outer loop". In this loop, we try out different values of k. Whenever we reach this step, i is one less than a power of 2. This is the one step where we redefine x and k so $x = x_k$ throughout the algorithm.

Step 4 is where we increment i and, for the new i, we compute $y = x_i$ to be $f(x_{i-1})$. This is the start of the inner loop, a loop where we try out different values of i for the current value of k. Note that the first time that this step is executed is after Step 3, where we defined k to be i, so the first iteration of the inner loop corresponds to $i = k + 1$.

In Step 5, the absolute value makes no mathematical difference, but we take it in case we wish to use available computer code for the gcd function that requires non-negative function arguments.

Step 7 is the end of the inner loop and Step 8 is the end of the outer loop. Note that the change to k in Step 8 ensures that $k = 2^l - 1$ for our new value of l. The redefinition of x ensures that x equals x_k for the new value of k.

We now give an example to illustrate the Pollard-Brent method.

Example 12.5. Suppose $n = 481$, $f(x) = x^2 + 1$, and $x_0 = 2$.

$l = k = 0$, $i = 1$:
$x_1 = f(x_0) = 2^2 + 1 = 5$
$m = |x_1 - x_0| = |5 - 2| = 3$
$d = \gcd(481, 3) = 1$, so we continue

$l = k = 1$, $i = 2$:
$x_2 = f(x_1) = 5^2 + 1 = 26$
$m = |x_2 - x_1| = |26 - 5| = 21$
$d = \gcd(481, 21) = 1$, so we continue

$l = k = 1$, $i = 3$:
$x_3 = f(x_2) = 26^2 + 1 = 196 \pmod{481}$
$m = |x_3 - x_1| = |196 - 5| = 191$
$d = \gcd(481, 191) = 1$, so we continue

$l = 2$, $k = 3$, $i = 4$:
$x_4 = f(x_3) = 196^2 + 1 = 418 \pmod{481}$
$m = |x_4 - x_3| = |418 - 196| = 222$
$d = \gcd(481, 222) = 37$, so we stop

Hence, 37 is a factor of 481, and we get $481 = 13 \cdot 37$.

12.2. Encrypting short messages using small e.

Let M be the numerical message we want to encrypt and let $N = pq$. We observe that when $M < N^{1/e}$, raising M to the e^{th} power and reducing it modulo N involves no modular reduction since $M^e < N$. Given the ciphertext $C = M^e \bmod N$, an attacker can determine M by calculating $M = C^{1/e}$ in $\mathbb{Z}$. This can easily be done: finding e^{th} roots is easy over $\mathbb{Z}$ and only difficult when working modulo N.

Hence, for small exponents e, the RSA algorithm is vulnerable to attacks. We illustrate this idea with an example.

Example 12.6. Let $(10379, 3)$ be the public key. We want to encrypt the message $M = 15$. Hence, we get $C \equiv 15^3 \equiv 3375 \pmod{10379}$. As $15 < \sqrt[3]{10379}$, the reduction of 15^3 modulo 10379 is itself. If an eavesdropper wants to determine the original message M, he simply needs to compute $M = \sqrt[3]{3375} = 15$.

12.3. Encrypting related messages.

For this attack, we assume that two related messages were sent to the same recipient. Suppose we send messages M and $M + \delta$ to a recipient using the public key (N, e), where δ is known but the message M is not. By encrypting M and $M + \delta$, we obtain $C_1 \equiv M^e \pmod{N}$ and $C_2 \equiv (M + \delta)^e \pmod{N}$, respectively. An eavesdropper can define the two polynomials $p_1(x) := x^e - C_1 \pmod{N}$ and $p_2(x) := (x + \delta)^e - C_2 \pmod{N}$, both of degree e. We observe that $x = M$ is a root modulo N of both polynomials. Thus, both polynomials have the linear factor $x - M$. Hence, if the greatest common divisor of $p_1(x)$ and $p_2(x)$ modulo N is linear, this will reveal the message M. Since the greatest common divisor can be computed efficiently, this attack is feasible for small exponents e.

Example 12.7. In this example, we let the public key be $(145, 3)$, the message be $M = 15$ and we let $\delta = 2$. Thus, we get

$$C_1 \equiv 15^3 \equiv 40 \pmod{145} \quad \text{and} \quad C_2 \equiv (15 + 2)^3 \equiv 17^3 \equiv 128 \pmod{145}.$$

We define the two $\mathbb{Z}_{145}$-polynomials $p_1(x) = x^3 - 40$ and $p_2(x) = (x + 2)^3 - 128 \equiv x^3 + 6x^2 + 12x + 25$. Both polynomials have the common factor $x - 15$. Thus, factorizing both polynomials (in $\mathbb{Z}_{145}[x]$), we get $p_1(x) = (x - 15)(x^2 + 15x + 80)$ and $p_2(x) = (x - 15)(x^2 + 21x + 37)$. And indeed, $x - 15$ is the greatest common divisor of $p_1(x)$ and $p_2(x)$. This shows that the original message was $M = 15$.

12.4. Sending the same message to multiple receivers.

Suppose we want to encrypt the message M and send it to k different parties, all of whom have chosen the same encryption exponent e but different N, i.e. we need to use the public keys (N_i, e) for $1 \leq i \leq k$. If $\gcd(N_i, N_j) \neq 1$ for some $i \neq j$, we can factorize N_i and N_j and easily recover the message M, so let us suppose that $\gcd(N_i, N_j) = 1$ for all $1 \leq i < j \leq k$. An eavesdropper sees

$$C_i \equiv M^e \pmod{N_i}, \qquad i = 1, \ldots, k.$$

Let $N = \prod_{i=1}^{k} N_i$. Using the Chinese remainder theorem, there exists a unique non-negative integer $C^* < N$ such that

$$C^* \equiv C_i \pmod{N_i}, \qquad i = 1, \ldots, k.$$

Suppose now that $k \geq e$. Then M^e satisfies the above set of k equations, and $M^e < N$ as $M < \min\{N_i \mid 1 \leq i \leq k\}$. Thus, $C^* = M^e$ over the integers, and so the message M can be easily recovered by computing the e^{th} root of C^*.

Example 12.8. In this example, we send the message $M = 15$ to three different parties using the public keys $(187, 3)$, $(161, 3)$ and $(145, 3)$, respectively. Encrypting

$M = 15$, we obtain

$$C_1 \equiv 15^3 \equiv 9 \pmod{187}$$
$$C_2 \equiv 15^3 \equiv 155 \pmod{161}$$
$$C_3 \equiv 15^3 \equiv 40 \pmod{145}.$$

As 187, 161 and 145 are all coprime, we can now solve the following system of linear congruences using the Chinese remainder theorem:

$$x \equiv 9 \pmod{187}$$
$$x \equiv 155 \pmod{161}$$
$$x \equiv 40 \pmod{145}.$$

Thus, we get $x \equiv 3375 \pmod{4365515}$. We know that $M^3 = x$. Therefore, $M = \sqrt[3]{x}$ and we get $M = \sqrt[3]{3375} = 15$. And indeed, $M = 15$ was the message that we encrypted.

12.5. Blinding.

Another way to attack the RSA algorithm is a technique called blinding, see **[1]**. To illustrate the idea of blinding, we call the two parties involved Alice and Bob. Alice obtains a valid signature on a message of her choice by asking Bob to sign a random "blinded" message. Bob has no idea what message he is actually signing. Using the signature from the "blinded" message, Alice can obtain Bob's signature from the "blinded" message and use it to sign the original message in Bob's name.

Let (N, e) be Bob's public key and (N, d) be his corresponding private key. Suppose an attacker called Alice wants Bob's signature on a message $M \in \mathbb{Z}_N^*$. Alice knows that Bob would refuse to sign the message once he reads it, so Alice can try the following: she chooses a random $r \in \mathbb{Z}_N^*$ and defines $M' := r^e M \bmod N$. Then she asks Bob to sign the message M'. Bob may be willing to sign M' since it seems meaningless, so he sends the signature S' back to Alice, where $S' \equiv (M')^d \bmod N$. Alice merely has to compute $S := S'/r \pmod N$ in order to obtain Bob's signature S on the original message M. Indeed,

$$S^e \equiv (S')^e/r^e \equiv (M')^{ed} \equiv M'/r^e \equiv M \pmod N.$$

As discussed in Chapter 11.1, most signature schemes do not directly sign a message, but rather a hashed message digest, so this attack is not a serious concern to the RSA algorithm. Nevertheless, let us illustrate the idea of blinding with an example of a directly signed message.

Example 12.9. Suppose Bob's public key is $(91, 5)$ and his private key is $(91, 29)$. Alice wants Bob to sign the message $M = 35$, but Bob refuses. Alice picks $r = 18$ and calculates $M' \equiv 18^5 \cdot 35 \equiv 84 \pmod{91}$. This time, Bob agrees to sign the message and so he computes $S' \equiv 84^{29} \equiv 28 \pmod{91}$ and sends this signature to Alice. Alice now computes $S = 28/18 \equiv 42 \pmod{91}$. Thus, Alice now knows that the signature to her original message $M = 35$ is $S = 42$. And indeed, $35^{29} \equiv 42 \bmod 91$.

13. RSA Algorithm in $\mathbb{Z}[i]$

Having discussed the RSA algorithm in the integers, our goal now is to extend this algorithm to the Gaussian integers.

Recall that modular arithmetic in $\mathbb{Z}[i]$ was discussed in Chapter 5. One difference in notation below from that in Chapter 5 is that we will write $N_{\mathbb{Z}[i]}(x)$ for the Gaussian norm of $x \in \mathbb{Z}[i]$. This is consistent with our use of $\phi_{\mathbb{Z}[i]}(x)$ for Gaussian ϕ-function values, and also helps to distinguish the norm from the product of two primes, which we call N as we did in $\mathbb{Z}$.

Recall from Theorem 3.6 that the prime factorization of a $\mathbb{Z}$-prime p in $\mathbb{Z}[i]$ is as given by the following three cases:

- If $p = 2$, then $2 = (1+i)(1-i)$ is a product of two associate Gaussian primes.
- If $p \equiv 3 \pmod 4$, then p is a Gaussian prime.
- If $p \equiv 1 \pmod 4$, then p is a product $\pi\bar{\pi}$ of two conjugate non-associate Gaussian primes.

Since the primes used in the RSA algorithm must be sufficiently large, we omit the case $p = 2$, as the Gaussian prime $1 + i$ is simply too small to be used in the RSA algorithm. Thus, we only deal with the primes p and π, where p is a $\mathbb{Z}$-prime of the form $4n + 3$ and π is a non-real Gaussian prime.

When choosing p and q for the RSA algorithm in $\mathbb{Z}[i]$, where $p \neq q$, we have 3 different options:

- p, q are both $\mathbb{Z}$-primes of the form $4n + 3$;
- p, q are both non-real Gaussian primes;
- one of p and q is a $\mathbb{Z}$-prime and the other is a non-real Gaussian prime.

We now discuss each of the three cases in detail.

13.1. Case 1: A $\mathbb{Z}$-prime and a non-real Gaussian prime.

We start with discussing the case where one of the primes p and q is a $\mathbb{Z}$-prime of the form $4n+3$ and the other is a non-real Gaussian prime. This case is completely unsuitable for encryption because its prime decomposition can be obtained very efficiently. To see this, suppose we pick a $\mathbb{Z}$-prime p and a non-real Gaussian prime $\pi = a + bi$, $a, b \in \mathbb{Z}$. Computing their product gives $N = p(a + bi) = pa + pbi$. Since a and b are coprime, p is the greatest common divisor of the real and imaginary parts of N, and so the division algorithm can be used to factorize N efficiently.

Example 13.1. Let $N = 1270 + 1651i$ be the product of the primes p and π, where p and π are unknown. We use the division algorithm to find the prime factors of N.

$$\begin{aligned} 1651 &= 1270 + 381 \\ 1270 &= 3 \cdot 381 + 127 \\ 381 &= 3 \cdot 127 + 0 \end{aligned}$$

Hence, $\gcd(1651, 1270) = 127$, and so one of the factors equals 127. Dividing both 1270 and 1651 by 127 gives the other prime factor, $10 + 13i$.

As this case of the RSA algorithm in $\mathbb{Z}[i]$ cannot be implemented securely, we do not discuss it further.

13.2. Case 2: Two $\mathbb{Z}$-primes.

We next discuss the RSA algorithm in $\mathbb{Z}[i]$ for two $\mathbb{Z}$-primes of the form $4n+3$. Since we are working with $\mathbb{Z}$-primes, this approach works in a fairly similar fashion to the RSA algorithm in $\mathbb{Z}$. However, both the original and the encrypted messages now have both real and imaginary parts.

We start with an example of how to encrypt and decrypt a message of the form $a+bi$.

Example 13.2. In this example, we pick our $\mathbb{Z}$-primes to be $p = 19$ and $q = 23$ and we want to encrypt the message $M = 4 + 7i$.

We start by calculating the product of p and q: $N = 19 \cdot 23 = 437$. Next, we calculate

$$\phi_{\mathbb{Z}[i]}(437) = \phi_{\mathbb{Z}[i]}(19)\phi_{\mathbb{Z}[i]}(23) = (N_{\mathbb{Z}[i]}(19) - 1)(N_{\mathbb{Z}[i]}(23) - 1) = 360 \cdot 528 = 190080\,.$$

We choose $e = 7$ to be our encryption exponent; note that $\gcd(7, 190080) = 1$ and $1 < 7 < 190080$.

To find the decryption exponent d, we solve the equation $7 \cdot d \equiv 1 \bmod 190080$ to get $d = 81463$.

Encrypting the message $M = 4+7i$ using the RSA algorithm, we reduce $(4+7i)^7 \bmod 437$ to get $C = 141 + 265i$. To decrypt the ciphertext $C = 141 + 265i$, we reduce $(141 + 265i)^{81463} \pmod{437}$, giving the original message $M = 4 + 7i$.

The above example illustrates that encryption and decryption over $\mathbb{Z}[i]$ works in a similar fashion as over $\mathbb{Z}$. Note though that we must use the $\phi_{\mathbb{Z}[i]}$-function rather than the Euler ϕ-function.

And indeed, as the operations are the same as in $\mathbb{Z}$, i.e. we are only using multiplication and reduction modulo N, the correctness requirement from Definition 10.1 also holds for this case of the RSA algorithm in $\mathbb{Z}[i]$, i.e. $D_k(E_k(M)) = M$ also holds over $\mathbb{Z}[i]$. Verifying this, we get $(M^e)^d \equiv M^{ed} \equiv M \pmod{N}$, as $ed \equiv 1 \pmod{\phi_{\mathbb{Z}[i]}(N)}$.

As both p and q are $\mathbb{Z}$-primes, the attacks that we discussed in Chapter 12 for the RSA algorithm over $\mathbb{Z}$ work exactly the same way over $\mathbb{Z}[i]$.

What might give hope that the RSA algorithm over $\mathbb{Z}[i]$ in this case may be a little more secure than the RSA algorithm over $\mathbb{Z}$ for the same value of $N = pq$ is the fact that $\phi_{Z[i]}(N) > \phi_{\mathbb{Z}}(N)$. Therefore, we have a larger set from which we can choose the encryption exponent e and we get a larger set of possible decryption exponents d.

13.3. Case 3: Two non-real Gaussian primes.

The final case for the RSA algorithm in $\mathbb{Z}[i]$ that we need to discuss is the case where we choose two non-real Gaussian primes, say π and σ.

As the only operations we use here are multiplication and reduction modulo N, this case of the RSA algorithm, using two Gaussian primes, works similarly to the case with two $\mathbb{Z}$-primes discussed above. Furthermore, we can verify that the correctness requirement from Definition 10.1 holds, i.e. $D_k(E_k(M)) = M$ also holds for this case of the RSA algorithm over $\mathbb{Z}[i]$.

However, we have two different ways of carrying out the work here. The first option is to perform all calculations in $\mathbb{Z}[i]$. The second option is to use the isomorphism map between quotients of $\mathbb{Z}$ and $\mathbb{Z}[i]$ that we established in Theorem 9.6.

We will start with an example where we perform all the calculations in $\mathbb{Z}[i]$.

Example 13.3. In this example, we choose the two Gaussian primes be $\pi = 3 + 2i$ and $\sigma = 6 + 5i$, and we want to encrypt the message $M = -23 + 17i$ using the RSA algorithm over $\mathbb{Z}[i]$. We first compute the product $N := \pi\sigma = (3+2i)(6+5i) = 8+27i$, and then calculate

$$\begin{aligned}\phi_{Z[i]}(8 + 27i) &= \phi_{Z[i]}(3 + 2i)\phi_{Z[i]}(6 + 5i)\\ &= (N_{\mathbb{Z}[i]}(3 + 2i) - 1)(N_{\mathbb{Z}[i]}(6 + 5i) - 1) = 12 \cdot 60 = 720\,.\end{aligned}$$

We choose $e = 7$ to be our encryption exponent, noting that $\gcd(\phi_{Z[i]}(8 + 27i), 7) = 1$ and $1 < 7 < \phi_{Z[i]}(8 + 27i)$. Thus, the public key is $(8 + 27i, 7)$. Now, we solve the equation $ed \equiv 1 \bmod 720$ for d, i.e. we solve $7 \cdot d \equiv 1 \bmod 720$, and obtain that our decryption exponent is $d = 103$. Thus, the private key is $(8 + 27i, 103)$.

To encrypt the message M, we need to reduce $(-23 + 17i)^7 \pmod{8 + 27i}$. Below, all equivalences are mod $8 + 27i$. Following the fast exponentiation algorithm, we get

$$\begin{aligned}(-23 + 17i)^2 &\equiv -19 + 29i\,,\\ (-23 + 17i)^4 &\equiv (-19 + 29i)^2 \equiv -1 + 19i\,,\end{aligned}$$

and so,

$$(-23 + 17i)^7 \equiv (-1 + 19i)(-19 + 29i)(-23 + 17i) \equiv -18 + 29i\,.$$

Thus, $(-23 + 17i)^7 \equiv -18 + 29i$ is the encrypted message.

To decrypt the ciphertext $C \equiv -18 + 29i$, we reduce $(-18 + 29i)^{103} \pmod{8 + 27i}$. Using fast exponentiation again, we get

$$\begin{aligned}(-18 + 29i)^2 &\equiv 7i\\ (-18 + 29i)^4 &\equiv -14 + 19i\\ (-18 + 29i)^8 &\equiv -5 + 8i\\ (-18 + 29i)^{16} &\equiv -7 + 28i\\ (-18 + 29i)^{32} &\equiv 7i\\ (-18 + 29i)^{64} &\equiv -14 + 19i\end{aligned}$$

and so $(-18+29i)^{103} \equiv (-14+19i)(7i)(-14+19i)(7i)(-18+29i) \equiv -23+17i$, which was our original message M.

We will now discuss how to carry out an equivalent encryption using the isomorphism map. For this approach, we first pick the two Gaussian primes, say π and σ and, as usual, define $N := \pi\sigma$. Doing arithmetic mod N means working over the quotient ring $\mathbb{Z}[i]/(N)$. Since neither π nor σ are in $\mathbb{Z}$, $N = a + bi$, where a and b are coprime. Consequently, Theorem 9.6 tells us that $\mathbb{Z}[i]/(N) \cong \mathbb{Z}_{N_{\mathbb{Z}[i]}(N)}$; this isomorphism allows us to recast the Gaussian RSA modulo N in terms of the normal RSA algorithm for $\mathbb{Z}$ modulo $N_{\mathbb{Z}[i]}(N)$.

As N is a Gaussian integer of the form $a + bi$, where a and b are coprime, we can compute b^{-1} in $\mathbb{Z}_{N_{\mathbb{Z}[i]}(N)}$. We then obtain the isomorphism map ψ, where $\psi(x + yi) = x - (a \cdot b^{-1}) \cdot y \pmod{N_{\mathbb{Z}[i]}(N)}$. Note that ψ is the map we obtained by proving Theorem 9.6.

We now illustrate this idea by redoing Example 13.3, but using our isomorphism.

Example 13.4. Let $\pi = 3 + 2i$ and $\sigma = 6 + 5i$ as in Example 13.3. As before, $N = 8 + 27i$, and so we are working in $Z[i]/(8 + 27i)$. Since $N_{\mathbb{Z}[i]}(8 + 27i) = 793$, we have $\mathbb{Z}[i]/(8 + 27i) \cong \mathbb{Z}_{793}$.

To write down the isomorphism map ψ, we first need to compute 27^{-1} in $\mathbb{Z}_{793}$. Solving $27x \equiv 1 \bmod 793$, we find $x \equiv 235 \bmod 793$. Thus our desired isomorphism is

$$\psi(x + yi) \equiv x - (8 \cdot 235)y \equiv x - 294y \pmod{793}.$$

As in Example 13.3, we encrypt the message $M = -23 + 17i$. We first compute

$$\psi(-23 + 17i) \equiv -23 - 294 \cdot 17 \equiv 530 \pmod{793}.$$

Now, let $e = 7$ as in Example 13.3, and we compute $530^7 \equiv 179 \bmod 793$. Recall that in Example 13.3, our solution was $C = -18 + 29i$. And indeed, applying the isomorphism map to this answer, we get

$$\psi(-18 + 29i) = -18 - (294) \cdot 29 \equiv 179 \bmod 793.$$

Thus, in fact, we get the same answer.

13.4. RSA Digital Signature in $\mathbb{Z}[i]$.

Above, we discussed the three different cases for the RSA algorithm in the Gaussian integers and found that the algorithm works similarly to the RSA algorithm in $\mathbb{Z}$. In Chapter 11.1, we discussed the RSA algorithm including digital signatures and saw that the only operations included were multiplication and reduction modulo N, where N is the product of the two primes. Since the cases for the Gaussian integers work similarly to the integer case, we do not need to discuss the digital signatures for the Gaussian integers in detail as the operations are the same.

To sign a numerical message M, we again compute $S \equiv M^d \pmod{N}$ to obtain the signature S and send the message (S, M) to the other party. The receiver of the message checks whether $S^e \pmod{N}$ is equivalent to M to determine whether S is a valid signature or not.

13.5. Attacking RSA in $\mathbb{Z}[i]$.

The methods for attacking the RSA-algorithm in $\mathbb{Z}[i]$ can be adapted from those discussed in Chapter 12 for $\mathbb{Z}$. All other attacks work similarly to the integer case. Therefore, we need not elaborate on this.

13.6. Conclusion.

In summary, we have seen three different approaches to an RSA algorithm for $\mathbb{Z}[i]$.

The first approach is to take a $\mathbb{Z}$-prime and a non-real Gaussian prime. As we have already noted, this approach is hopeless: it can be easily attacked and is therefore not a secure implementation for the RSA algorithm in $\mathbb{Z}[i]$.

The second approach is to take two $\mathbb{Z}$-primes of the form $4n + 3$, while the third approach is to take two Gaussian primes. Both approaches work similarly to the RSA algorithm in $\mathbb{Z}$.

The third approach in particular is very similar to RSA over $\mathbb{Z}$, since we have an isomorphism between the quotient rings $\mathbb{Z}[i]/(I)$ and $\mathbb{Z}/(J)$, where $I = (a + bi)$ and $J = (a^2 + b^2)$, and so all calculations are essentially the same up to isomorphism. This approach is therefore just RSA over $\mathbb{Z}$ in disguise, so it can be discarded as it offers nothing new.

The second approach is analogous, but not isomorphic, to RSA over $\mathbb{Z}$. Thus, we have a new but related variant of RSA. It might even be a slightly more secure variant since we have a larger set from which to choose our encryption and decryption exponents, but this would need further investigation.

14. RSA Algorithm in $\mathbb{Z}[\omega]$

Having studied the RSA algorithm in the Gaussian integers, we now want to develop this algorithm in the context of the Eisenstein integers.

Recall from Theorem 4.6 that the prime factorization of a $\mathbb{Z}$-prime p in $\mathbb{Z}[\omega]$ is as given by the following three cases:

- if $p = 3$, $3 = (2+\omega)(2+\omega^2)$ is the product of two associate Eisenstein primes,
- if $p \equiv 2 \pmod 3$, then p is an Eisenstein prime,
- if $p \equiv 1 \pmod 3$, then p is a product of two conjugate non-associate Eisenstein primes $\pi\bar{\pi}$.

Similarly to the Gaussian integers, we ignore the case $p = 3$, as the Eisenstein prime $2+\omega$ is too small to be used in the RSA algorithm. Thus, we only look into the cases p and π, where p is a $\mathbb{Z}$-prime of the form $3n+2$ and π is a non-real Eisenstein prime.

When choosing p and q for the RSA algorithm in $\mathbb{Z}[\omega]$, where $p \neq q$, we have three different options:

- p, q are both $\mathbb{Z}$-primes;
- p, q are both non-real Eisenstein primes;
- one of p and q is a $\mathbb{Z}$-prime and the other is a non-real Eisenstein prime.

We now discuss each of the three cases in detail.

14.1. Case 1: A $\mathbb{Z}$-prime and a non-real Eisenstein prime.

The first case we will discuss is the case where we choose π to be an non-real Eisenstein prime and p to be a $\mathbb{Z}$-prime of the form $3n+2$. As for the analogous case for the Gaussian integers, this case for the Eisenstein integers cannot be safely implemented in the RSA algorithm. To see this, suppose we pick a $\mathbb{Z}$-prime p of the form $3n+2$ and a non-real Eisenstein prime $\pi = a+b\omega$. Calculating the product $N = p \cdot \pi = pa + pb\omega$, we see that p is a common factor. Hence, using the divison algorithm, we can efficiently factorize N and therefore easily attack this implementation of the RSA algorithm.

Example 14.1. Let $N = p \cdot q = 1507 + 8083\omega$, where p and q are unknown. Using the division algorithm, we determine the prime factors p and q:

$$\begin{aligned} 8083 &= 5 \cdot 1507 + 548 \\ 1507 &= 2 \cdot 548 + 411 \\ 548 &= 411 + 137 \\ 411 &= 3 \cdot 137 + 0 \end{aligned}$$

Hence, $\gcd(1507, 8083) = 137$ and so one of the prime factors of N is 137. Dividing both 1507 and 8083 by 137, we obtain the other prime factor, $11 + 59\omega$.

14.2. Case 2: Two $\mathbb{Z}$-primes.

We next discuss the RSA algorithm in $\mathbb{Z}[\omega]$ for two $\mathbb{Z}$-primes of the form $3n+2$. Since we are working with $\mathbb{Z}$-primes, this approach works in a fairly similar fashion to the RSA algorithm in $\mathbb{Z}$. However, both the original and the encrypted messages now lie in $\mathbb{Z}[\omega]$ (modulo some product N of two primes).

We start by illustrating this with an example.

Example 14.2. In this example, we pick our $\mathbb{Z}$-primes to be $p = 17$ and $q = 23$ and we want to encrypt the message $M = 5 + 3\omega$.

First, we calculate the product $N = p \cdot q = 17 \cdot 23 = 391$ and compute the $\phi_{Z[\omega]}(391)$ to find the number of Eisenstein integers coprime to 391:

$$\phi_{Z[\omega]}(391) = \phi_{Z[\omega]}(17)\phi_{Z[\omega]}(23) = (N_{Z[\omega]}(17) - 1)(N_{Z[\omega]}(23) - 1) = 288 \cdot 528 = 152064\,.$$

We choose $e = 13$, noting that $\gcd(13, 152064) = 1$ and $1 < 13 < 152064$. We want $ed \equiv 1 \pmod{\phi_{Z[\omega]}(N)}$, so we solve the equivalence $13d \equiv 1 \bmod 152064$ to get $d \equiv 46789$.

To encrypt the message $M = 5+3\omega$, we reduce $(5+3\omega)^{13} \bmod 391$ to get the ciphertext $C = 250 + 366\omega$. To check if the encryption system works, we decrypt the ciphertext by reducing $(250 + 366\omega)^{46789} \bmod 391$ to get the original message $M = 5 + 3\omega$.

The above example illustrates that encryption and decryption over $\mathbb{Z}[\omega]$, using two $\mathbb{Z}$-primes, works in a similar fashion as over $\mathbb{Z}$. Note though that we must use the $\phi_{\mathbb{Z}[\omega]}$-function rather than the Euler ϕ-function.

And indeed, as the operations are the same as in $\mathbb{Z}$, i.e. we are only using multiplication and reduction modulo N, the correctness requirement from Definition 10.1 also holds for this case of the RSA algorithm in $\mathbb{Z}[\omega]$, i.e. $D_k(E_k(M)) = M$ also holds over $\mathbb{Z}[\omega]$. Verifying this, we get $(M^e)^d \equiv M^{ed} \equiv M \pmod{N}$, as $ed \equiv 1 \pmod{\phi_{\mathbb{Z}[\omega]}(N)}$.

As both p and q are $\mathbb{Z}$-primes, the attacks that we discussed in Chapter 12 for the RSA algorithm over $\mathbb{Z}$ work exactly the same way over $\mathbb{Z}[\omega]$.

As for the corresponding case over $\mathbb{Z}[i]$, we could hope that the RSA algorithm over $\mathbb{Z}[\omega]$ in this case may be a little more secure than the RSA algorithm over $\mathbb{Z}$ for the same value of $N = pq$ due to the fact that $\phi_{Z[\omega]}(N) > \phi_{\mathbb{Z}}(N)$. Therefore, we have a larger set from which we can choose the encryption and decryption exponents e and d.

14.3. Case 3: Two non-real Eisenstein primes.

The final case for the RSA algorithm in $\mathbb{Z}[\omega]$ that we need to discuss is the case where we choose two non-real Eisenstein primes π and σ.

As the only operations we use here are multiplication and reduction modulo N, this case of the RSA algorithm works similarly to the case with two $\mathbb{Z}$-primes discussed above. Furthermore, we can verify that the correctness requirement from Definition 10.1 holds, i.e. $D_k(E_k(M)) = M$ also holds for this case of the RSA algorithm over $\mathbb{Z}[\omega]$.

As in the corresponding case for Gaussian integers, we have two different ways of carrying out the work here. The first option is to perform all calculations in $\mathbb{Z}[\omega]$. The second option is to use the isomorphism map between quotients of $\mathbb{Z}$ and $\mathbb{Z}[\omega]$ that we established in Theorem 9.14 and perform all calculations in $\mathbb{Z}$.

We will start with an example where we perform all the calculations in $\mathbb{Z}[\omega]$.

Example 14.3. In this example, we let $p = 5 + 2\omega$ and $q = 7 + 3\omega$ and we want to encrypt the message $M = 4 + 5\omega$ using the RSA algorithm over $\mathbb{Z}[\omega]$.

We start by calculating the product of the two primes and obtain $N = 29+23\omega$. Next,

$$\begin{aligned}\phi_{Z[\omega]}(29 + 23\omega) &= \phi_{Z[\omega]}(2 + 5\omega) \cdot \phi_{Z[\omega]}(7 + 3\omega) \\ &= (N_{Z[\omega]}(2 + 5\omega) - 1)(N_{Z[\omega]}(7 + 3\omega) - 1) = 18 \cdot 36 = 648\,.\end{aligned}$$

We let $e = 5$, noting that $\gcd(5, 648) = 1$ and $1 < 5 < 648$. Solving the equivalence $5d \equiv 1 \pmod{648}$ for d, we obtain $d = 389$.

To encrypt our message $M = 4 + 5\omega$, we reduce $(4 + 5\omega)^5 \bmod 29 + 23\omega$ to get the ciphertext $C = 11 + 23\omega$. To decrypt this message, we reduce $(11 + 23\omega)^{389} \bmod 29 + 23\omega$, yielding the original message $M = 4 + 5\omega$.

Next, we consider the same case, but handle it using the isomorphism map. We again choose two Eisenstein primes, say π and σ, and calculate their product N. Doing arithmetic mod N means working over the quotient ring $\mathbb{Z}[\omega]/(N)$. Since neither π nor σ are in $\mathbb{Z}$, $N = a + b\omega$, where a and b are coprime. Consequently, Theorem 9.14 tells us that $\mathbb{Z}[\omega]/(N) \cong \mathbb{Z}_{N_{\mathbb{Z}[\omega]}(N)}$; this isomorphism allows us to recast the Eisenstein RSA modulo N in terms of the normal RSA algorithm for $\mathbb{Z}$ modulo $N_{\mathbb{Z}[i]}(N)$.

As N is an Eisenstein integer of the form $a + b\omega$, where a and b are coprime, we can compute b^{-1} in $\mathbb{Z}_{N_{\mathbb{Z}[\omega]}(N)}$. We then obtain the isomorphism map ψ, where $\psi(x+y\omega) = x - (a \cdot b^{-1}) \cdot y \pmod{N_{\mathbb{Z}[\omega]}(N)}$. Note that ψ is the map we obtained by proving Theorem 9.14.

We now illustrate this idea by redoing Example 14.3, but using our isomorphism.

Example 14.4. Let $p = 5 + 2\omega$ and $q = 7 + 3\omega$ as in Example 14.3. As before, $N = 29+23\omega$, and so we are working in $\mathbb{Z}[\omega]/(29+23\omega)$. Since $N_{\mathbb{Z}[\omega]}(29+23\omega) = 703$, we have $\mathbb{Z}[\omega]/(29 + 23\omega) \cong \mathbb{Z}_{703}$.

To write down the isomorphism map ψ, we first need to compute 23^{-1} in $\mathbb{Z}_{703}$. Solving $23x \equiv 1 \bmod 703$, we find $x \equiv 214 \bmod 703$. Thus our desired isomorphism is

$$\psi(x + yi) \equiv x - (29 \cdot 214)y \equiv x - 582y \pmod{703}\,.$$

As in Example 14.3, we encrypt the message $M = 4 + 5\omega$. We first compute

$$\psi(4 + 5\omega) \equiv 4 - 582 \cdot 5 \equiv 609 \pmod{703}\,.$$

Now, let $e = 5$ as in Example 14.3, and we compute $609^5 \equiv 685 \bmod 703$. Recall that in Example 14.3, our solution was $C = 11 + 23\omega$. And indeed, applying the isomorphism map to this answer, we get

$$\psi(11 + 23\omega) = 11 - 582 \cdot (23) \equiv 685 \bmod 703\,.$$

Thus, in fact, we get the same answer.

14.4. Conclusion.

As for $\mathbb{Z}[i]$, we have seen three different approaches to an RSA algorithm for $\mathbb{Z}[\omega]$

The first approach is to take a $\mathbb{Z}$-prime and a non-real Eisenstein prime. As we have already noted, this approach is hopeless: it can be easily attacked and is therefore not a secure implementation for the RSA algorithm in $\mathbb{Z}[\omega]$.

The second approach is to take two $\mathbb{Z}$-primes of the form $3n + 2$, while the third approach is to take two Eisenstein primes. As with the Gaussian integers, both approaches work similarly to the RSA algorithm over $\mathbb{Z}$.

The third approach in particular is very similar to RSA over $\mathbb{Z}$, since we have an isomorphism between the quotient rings $\mathbb{Z}[\omega]/(I)$ and $\mathbb{Z}/(J)$, where $I = (a + b\omega)$ and $J = (a^2 - ab + b^2)$, and so all calculations are essentially up to isomorphism. This approach is therefore just RSA over $\mathbb{Z}$ in disguise, so it can be discarded as it offers nothing new.

The second approach is analogous, but not isomorphic, to RSA over $\mathbb{Z}$. Thus, we have a new but related variant of RSA. It might even be a slightly more secure variant since we have a larger set from which to choose our encryption and decryption exponents, but again this would need further investigation.

15. Diffie-Hellman Key Exchange in $\mathbb{Z}$

The Diffie-Hellman Key Exchange, published in 1976 by Whitfield Diffie and Martin Hellman [**4**], was the first asymmetric encryption scheme to be made public.

Suppose two parties, whom we will call Alice and Bob, wish to create a shared secret message, typically an encryption key. To set up the Diffie-Hellman Key Exchange, Alice and Bob publicly agree on a prime number p and a generator g for the multiplicative group $\mathbb{Z}_p^*$. Alice chooses a random integer a to be her encryption exponent, where $1 \leq a \leq p-1$, and calculates $A \equiv g^a \bmod p$. Similarly, Bob chooses his encryption exponent b, where $1 \leq b \leq p-1$, and calculates $B \equiv g^b \bmod p$. They share their public keys (p, g, A) and (p, g, B) with each other, but keep the encryption exponents secret. Alice now computes $k_a \equiv B^a \bmod p$, and Bob computes $k_b \equiv A^b \bmod p$. Since

$$k_a \equiv B^a \equiv (g^b)^a \equiv g^{ab} \equiv (g^a)^b \equiv A^b \equiv k_b \pmod{p},$$

$k_a \equiv k_b$ is a shared secret message.

We now illustrate the Diffie-Hellman Key Exchange with an example.

Example 15.1. Alice and Bob decide to work with the prime number $p = 11$ and the generator $g = 2$. Alice chooses the encryption exponent $a = 7$ and calculates $A \equiv g^a \equiv 2^7 \equiv 7 \pmod{11}$. Alice's public key is now $(p, g, A) = (11, 2, 7)$. Bob chooses $b = 6$ to be his encryption exponent. He calculates $B \equiv g^b \equiv 2^6 \equiv 9 \pmod{11}$. Bob's public key is $(p, g, B) = (11, 2, 9)$.

To get the secret message, Alice calculates $k_a \equiv B^a \equiv 9^7 \equiv 4 \pmod{11}$, Bob calculates $k_b \equiv A^b \equiv 7^6 \equiv 4 \pmod{11}$. Then $k_a \equiv k_b \equiv 4$ is Alice and Bob's shared secret message.

If G is a cyclic group of order p, with generator g, then $\{g^0, g^1, \ldots, g^{p-1}\}$ is all of G. Thus, for every $h \in G$, there is a unique $x \in \mathbb{Z}_p$ such that $g^x = h$. When the underlying group G is understood from the context, we call this x the discrete logarithm of h with respect to g and write $x = \log_g h$. (Logarithms in this case are called *discrete* because they take values in a finite range.)

A Diffie-Hellman eavesdropper can discover Alice and Bob's shared secret by computing one of their private keys, i.e. either $a \equiv \log_g A$ or $b \equiv \log_g B$. For that reason, the Diffie-Hellman Key Exchange is called a *discrete logarithm encryption system.* It is effective because there is no known method of computing discrete logarithms that is efficient for large primes p.

16. Diffie-Hellman Key Exchange in $\mathbb{Z}[i]$

In this chapter, we describe a variant of the Diffie-Hellman Key Exchange for the Gaussian integers.

The discussion of the Diffie-Hellman Key Exchange in $\mathbb{Z}[i]$ can be divided into two cases. In the first case, we consider $\mathbb{Z}$-primes of the form $p \equiv 1 \bmod 4$, i.e. we work with Gaussian primes π, where π is of the form $a+bi$. In the second case, we consider $\mathbb{Z}$-primes of the form $p \equiv 3 \bmod 4$.

16.1. Case 1: Using a non-real Gaussian prime.

We begin by discussing the case where π is a non-real Gaussian prime. As for the RSA Algorithm, we have two options here. We can either work in $\mathbb{Z}[i]$, or use the isomorphism map between $\mathbb{Z}[i]$ and $\mathbb{Z}$ and perform all the calculations in $\mathbb{Z}$. By virtue of this isomorphism, the calculations work the same way as in $\mathbb{Z}$, and so this case is not genuinely new.

However, there is one interesting aspect to consider, namely the number of generators in $\mathbb{Z}[i]$. First, note that the number of generators of $\mathbb{Z}_p^*$ is $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}}(p))$, where $\phi_{\mathbb{Z}}$ is the Euler ϕ-function and p is an odd prime. To see this, first note that $n := \phi_{\mathbb{Z}}(p)$ is the order of $\mathbb{Z}_p^*$, and that $\mathbb{Z}_p^*$ is always cyclic. Thus, if g is a generator of $\mathbb{Z}_p^*$, then g^k is a generator if and only if k is coprime to p, and so there are $\phi_{\mathbb{Z}}(n)$ generators of $\mathbb{Z}_p^*$, as claimed.

In Chapter 5, we saw that many results on $\mathbb{Z}$ have analogues on $\mathbb{Z}[i]$ where we replace the prime π by the norm $N_{\mathbb{Z}[i]}(\pi)$ and $\phi_{\mathbb{Z}}$ by $\phi_{\mathbb{Z}[i]}$, so one might expect that there are $\phi_{\mathbb{Z}[i]}(\phi_{\mathbb{Z}[i]}(\pi))$ generators in the multiplicative group of $\mathbb{Z}[i]/(\pi)$.

However, this is incorrect: the actual number of generators of this group is $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[i]}(\pi))$. To see this, first note that it follows from the isomorphism theorem Theorem 9.6 that all nonzero elements of $\mathbb{Z}[i]/(\pi)$ form a cyclic group of order $n := \phi_{\mathbb{Z}[i]}(\pi) = N_{\mathbb{Z}[i]}(\pi)-1$, and now as before the number of generators of this group is $\phi_{\mathbb{Z}}(n)$.

We summarise the above observation as a corollary.

Corollary 16.1. *Let π be a non-real Gaussian prime and let $n := N_{\mathbb{Z}[i]}(\pi)$. The isomorphism map between the quotient rings $\mathbb{Z}[i]/(\pi)$ and $\mathbb{Z}/(n)$ maps the generators of the multiplicative group of $\mathbb{Z}[i]/(\pi)$ to the generators of $\mathbb{Z}_n^*$. Consequently, the number of generators is $\phi_{\mathbb{Z}}(n)$.*

We now illustrate this with an example.

Example 16.2. Let $\pi = 3 + 2i$ and note that $N_{\mathbb{Z}[i]}(3 + 2i) = 13$. We calculate $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[i]}(\pi))$ to obtain the number of generators in $\mathbb{Z}[i]/(3 + 2i)$. Now, $\phi_{\mathbb{Z}[i]}(3 + 2i) = N_{\mathbb{Z}[i]}(3 + 2i) - 1 = 13 - 1 = 12$ and $\phi_{\mathbb{Z}}(12) = 4$, so the number of generators of the multiplicative group of $\mathbb{Z}[i]/(3 + 2i)$ is 4.

16.2. Case 2: Using a $\mathbb{Z}$-prime.

Here, we use a $\mathbb{Z}$-prime p that is also a Gaussian prime, i.e. $p \equiv 3 \bmod 4$. Although we cannot now use the isomorphism map between $\mathbb{Z}[i]$ and $\mathbb{Z}$ here, the only operations used are multiplication and reduction modulo p. Thus, this case also works in a similar fashion to the integer case and the congruence

$$A^b \equiv (g^a)^b \equiv g^{ab} \equiv (g^b)^a \equiv B^a \pmod{p}$$

also holds in $\mathbb{Z}[i]$.

The number of generators can be determined as in Case 1 by calculating $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[i]}(p))$. Note that all generators are of the form $a + bi$, where $a, b \neq 0$ since if a or b were 0, powers thereof modulo a $\mathbb{Z}$-prime would not be equivalent to a Gaussian integer of the form $c + di$, where $c, d \neq 0$.

16.3. Conclusion.

In conclusion, there are two different approaches to developing a $\mathbb{Z}[i]$-variant of the Diffie-Hellman Key Exchange: we can either work with $\mathbb{Z}$-primes of the form $4n + 3$ or we can work with non-real Gaussian primes. Regardless of our choice of the prime, we have seen that the Diffie-Hellman Key Exchange works the same way in $\mathbb{Z}[i]$ as it does in $\mathbb{Z}$.

The interesting part we noted here is that we get the number of generators by computing $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[i]}(p))$, where p is our Gaussian prime. This works in both cases, that is, whether p is a $\mathbb{Z}$-prime or not. Depending on the type of prime, we get two different types of generators $a + bi$. When p is a $\mathbb{Z}$-prime, we must have $a, b \neq 0$, whereas if p is a non-real Gaussian prime, we can only say that at least one of a and b is nonzero. Apart from that, the $\mathbb{Z}[i]$-variant of the Diffie-Hellman Key Exchange does not bring any new insights.

17. Diffie-Hellman Key Exchange in $\mathbb{Z}[\omega]$

Analogously, we can discuss a variant of the Diffie-Hellman Key Exchange in the context of the Eisenstein integers $\mathbb{Z}[\omega]$. The discussion in $\mathbb{Z}[\omega]$ can be split up into two cases. In the first case, we consider $\mathbb{Z}$-primes of the form $p \equiv 1 \bmod 3$, i.e. we work with Eisenstein primes π, where π is of the form $a + b\omega$. In the second case, we consider $\mathbb{Z}$-primes of the form $p \equiv 2 \bmod 3$.

17.1. Case 1: Using a non-real Eisenstein prime.

Similar to the Gaussian case, this case for the Eisenstein integers offers no new insights. Again, we can either work in $\mathbb{Z}[\omega]$ or use the isomorphism between $\mathbb{Z}$ and $\mathbb{Z}[\omega]$ and calculate everything in $\mathbb{Z}$.

Suppose π is the non-real Eisenstein prime. As in the Gaussian case, the number of generators of the multiplicative group of $\mathbb{Z}[\omega]/(\pi)$ is $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[\omega]}(p))$. To see this, first note that it follows from the isomorphism theorem Theorem 9.14 that all nonzero elements of $\mathbb{Z}[\omega]/(\pi)$ form a cyclic group of order $n := \phi_{\mathbb{Z}[\omega]}(\pi) = N_{\mathbb{Z}[\omega]}(\pi) - 1$, and now as before the number of generators of this group is $\phi_{\mathbb{Z}}(n)$.

We summarise the above observation as a corollary.

Corollary 17.1. *Let π be a non-real Eisenstein prime and let $n := N_{\mathbb{Z}[\omega]}(\pi)$. The isomorphism map between the quotient rings $\mathbb{Z}[\omega]/(\pi)$ and $\mathbb{Z}/(n)$ maps the generators of the multiplicative group of $\mathbb{Z}[\omega]/(\pi)$ to the generators of $\mathbb{Z}_n^*$. Consequently, the number of generators is $\phi_{\mathbb{Z}}(n)$.*

The generators in this case of the Diffie-Hellman Key Exchange are of the form $a+b\omega$, where at least one of a and b is nonzero: a and b cannot both be equal to 0, as 0 cannot generate the group.

17.2. Case 2: Using a $\mathbb{Z}$-prime.

Here, we use a $\mathbb{Z}$-prime p that is also an Eisenstein prime, i.e. $p \equiv 2 \bmod 3$. Although we cannot now use the isomorphism map between $\mathbb{Z}[\omega]$ and $\mathbb{Z}$ here, the only operations used are multiplication and reduction modulo p. Thus, this case works in a similar fashion to the integer case and the congruence

$$A^b \equiv (g^a)^b \equiv g^{ab} \equiv (g^b)^a \equiv B^a \pmod{p}$$

also holds in $\mathbb{Z}[\omega]$.

The number of generators can be determined as in Case 1 by calculating $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[\omega]}(p))$. Note that all generators are of the form $a + b\omega$, where $a, b \neq 0$ since if a or b were 0, powers thereof modulo a $\mathbb{Z}$-prime would not be equivalent to an Eisenstein integer of the form $c + d\omega$, where $c, d \neq 0$.

17.3. Conclusion.

In conclusion, there are two different approaches to developing a $\mathbb{Z}[\omega]$-variant of the Diffie-Hellman Key Exchange: we can either work with $\mathbb{Z}$-primes of the form $3n+2$ or we can work with non-real Eisenstein primes. Regardless of our choice of the prime, we have seen that the Diffie-Hellman Key Exchange works the same way in $\mathbb{Z}[\omega]$ as it does in $\mathbb{Z}$.

The interesting part we noted here is that we get the number of generators by computing $\phi_{\mathbb{Z}}(\phi_{\mathbb{Z}[\omega]}(p))$, where p is our Eisenstein prime. This works in both cases, that is, whether p is a $\mathbb{Z}$-prime or not. Depending on the type of prime, we get two different types of generators $a+b\omega$. When p is a $\mathbb{Z}$-prime, we must have $a, b \neq 0$, whereas if p is a non-real Eisenstein prime, we can only say that at least one of a and b is nonzero. Apart from that, the $\mathbb{Z}[\omega]$-variant of the Diffie-Hellman Key Exchange does not bring any new insights.

18. El Gamal in $\mathbb{Z}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$

The El Gamal encryption scheme is an extension of the Diffie-Hellman Key Exchange to allow public key encryption. It is based on Diffie-Hellman, but allows two parties not only to share a secret message, but also to exchange encrypted messages. This encryption scheme was first published by Taher El Gamal in 1985.

Suppose that the two parties who want to share an encrypted message are called Alice and Bob. We discuss the El Gamal encryption scheme for the case where Bob sends Alice an encrypted message and Alice decrypts the message.

The El Gamal encryption scheme includes the following steps. (In all cases below, "calculating $x \equiv E \bmod p$", where E is some expression, means reducing E mod p, i.e. computing an integer x, $1 \leq x < p$, which is equivalent to E mod p.)

(1) Alice chooses a prime p and a generator g of $\mathbb{Z}_p^*$.
(2) Alice chooses her private key a, where $1 \leq a \leq p-1$, and creates her public key $A \equiv g^a \bmod p$.
(3) Alice makes the key (p, g, A) public knowledge.
(4) If Bob wants to send Alice a message M, he first chooses an integer b as his private key, where $1 \leq b \leq p-1$, and calculates $K \equiv A^b \bmod p$.
(5) Bob then encrypts M as the pair of numbers $B \equiv g^b \bmod p$ and $C \equiv KM \bmod p$ and sends (B, C) to Alice.
(6) To decrypt the message, Alice first calculates $K \equiv B^a \bmod p$.
(7) Alice then calculates $M \equiv K^{-1}C \bmod p$.

We illustrate the El Gamal encryption scheme with an example.

Example 18.1. Alice chooses her prime $p = 23$, her generator $g = 5$ for $\mathbb{Z}_{23}^*$ and her private key $a = 7$. Alice calculates $A \equiv g^a \equiv 5^7 \equiv 17 \bmod 23$ and publishes $(23, 5, 17)$ as her public key.

Bob wants to send the numerical message $M = 15$ to Alice. He picks $b = 9$ and calculates $K \equiv A^b \equiv 17^9 \equiv 7 \bmod 23$. Bob also calculates $B \equiv g^b \equiv 5^9 \equiv 11 \bmod 23$ and $C \equiv K \cdot M \equiv 7 \cdot 15 \equiv 13 \bmod 23$. He sends the pair $(11, 13)$ to Alice.

Alice can now calculate $K \equiv B^a \equiv 11^7 \equiv 7 \bmod 23$. She also knows that $M \equiv K^{-1} \cdot C \bmod p$. She finds K^{-1} by solving the equation $K \cdot K^{-1} \equiv 1 \bmod 23$, i.e. she solves $7K^{-1} \equiv 1 \bmod 23$ and finds that $K^{-1} \equiv 10 \bmod 23$. Alice now calculates $M \equiv K^{-1} \cdot C \equiv 10 \cdot 13 \equiv 15 \bmod 23$, to get the original message $M = 15$.

We remark that Alice could actually calculate K^{-1} directly, without calculating K. To see this, note first that $K \equiv B^a \bmod p$, and so $K^{-1} \equiv (B^a)^{-1} \equiv B^{-a} \bmod p$. By Fermat's Little Theorem, Theorem 5.43, we have $B^{p-1} \equiv 1 \bmod p$. Hence,

$$K^{-1} \equiv B^{-a} \equiv B^{-a}(1) \equiv B^{-a}B^{p-1} \equiv B^{p-a-1} \bmod p\,.$$

As discussed in the previous two chapters, the Diffie-Hellman Key Exchange works in $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$ in the same way as in $\mathbb{Z}$. Since the El Gamal system is based on the

Diffie-Hellman Key Exchange, there is thus no difference in the El Gamal encryption scheme in $\mathbb{Z}$, $\mathbb{Z}[i]$ and $\mathbb{Z}[\omega]$, Hence, we do not discuss this any further.

19. Attacking Diffie-Hellman and El Gamal

The Diffie-Hellman Key Exchange and the El Gamal encryption are both discrete logarithm encryption schemes. There are several methods to attack such schemes. In this chapter, we will discuss in detail some possible attacks on discrete logarithm encryption schemes.

19.1. Brute-force attack.

The idea behind the brute-force attack is simply to compute powers of the generator g until we find the result y.

$$\begin{aligned} g^1 &\stackrel{?}{=} y \\ g^2 &\stackrel{?}{=} y \\ &\vdots \\ g^x &\stackrel{?}{=} y \end{aligned}$$

This method is a very time-consuming way to calculate the discrete logarithm. For a random logarithm x, we expect to find the correct solution after checking half of the possible values for x. To thwart such brute-force attacks on discrete logarithmic encryption, we need to choose a group G whose order $|G|$ is sufficiently large. The Diffie-Hellman Key Exchange works with groups $\mathbb{Z}_p^*$, where p is prime. This means on average that $(p-1)/2$ computations are required to find the solution to the discrete logarithm. Bearing in mind today's computer technology, the group should at least be of order 2^{80} to render brute-force attacks infeasible.

19.2. Pohlig-Hellman algorithm.

This algorithm for attacking discrete logarithmic encryption was published in 1978 by Stephen Pohlig and Martin Hellman [**11**], who credit Roland Silver with having discovered the algorithm independently of them (and so the algorithm is sometimes called the Silver-Pohlig-Hellman algorithm).

With any attack on a discrete logarithm scheme, the aim is to find x such that $y \equiv g^x \bmod p$ for given y, g, and p; as usual, p is a prime and g is a generator of $\mathbb{Z}_p^*$. Thus, g has order $p-1$ and so x is defined mod $p-1$, so we may as well assume that $0 \leq x < p-1$.

Since p is a large prime, $p-1$ is certainly not prime. The idea of Pohlig-Hellman is to divide and conquer. Suppose

$$p - 1 = q_1^{a_1} q_2^{a_2} \dots q_r^{a_r} \tag{19.1}$$

is the prime decomposition of $p-1$, where $a_i \in \mathbb{N}$ for all i. Instead of computing $x \pmod{p-1}$, we will instead compute $x \pmod{q_i^{a_i}}$ for each i and use the Chinese Remainder Theorem then to efficiently recover $x \pmod{p-1}$. Effectively, we reduce the discrete log problem mod $p-1$ to a set of discrete log problems mod $q_i^{a_i}$. In fact, we

will see that we can do even better than this by reducing each of these latter problems more or less to a set of discrete log problems mod q_i. It follows that Pohlig-Hellman gives an efficient way to find x if all prime factors q_i are relatively small, but it is not particularly helpful if $p-1$ has one or more very large prime factors.

Thus, to protect against a Pohlig-Hellman attack we should aim to find such a prime p. We could for instance first choose a large prime q and then test for primality numbers of the form $p = 2kq+1$ for some positive integer k. The well-known Dirichlet theorem on arithmetic progressions indicates that there are infinitely many such primes and indeed the prime number theorem for arithmetic progressions gives us reassurance about how many such numbers we expect to have to test before we find such a prime p.

To eliminate unnecessary indices in our explanation of the Pohlig-Hellman algorithm, we write $q := q_i$ for a fixed but arbitrary i, and we write b in place of a_i, where a_i and a_i are as above. Our task is therefore to find u such that $0 \le u < q^b$ and $x \equiv u \bmod q^b$. To reduce this discrete log problem mod q^b to a set of similar problems mod q, we write u in its base-q expansion, i.e. $u = \sum_{k=0}^{b-1} u_k q^k$, where $0 \le u_k < q$ for all $0 \le k < b$, and so there exists an integer m such that

$$x = mq^b + \sum_{k=0}^{b-1} u_k q^k \,. \tag{19.2}$$

We will calculate the digits u_k one at a time.

To make this process work, we first compute and record all qth roots of unity mod p:

$$r_{q,j} \equiv g^{j(p-1)/q} \bmod p \quad \text{and} \quad 0 \le r_{q,j} < p\,, \qquad 0 \le j < q\,,$$

Since $x \equiv u \bmod q^b$, it follows from (19.2) that $x = u_0 + M_0 q$ for some integer M_0. Thus, defining $y_0 := y$ and writing equivalences mod p, we define $0 \le Y_0 < p$ by

$$Y_0 \equiv y_0^{(p-1)/q} \equiv g^{x(p-1)/q} = \left(g^{u_0+M_0 q}\right)^{(p-1)/q} = g^{u_0(p-1)/q} \cdot g^{M_0(p-1)} \equiv g^{u_0(p-1)/q}\,.$$

and so $u_0 = j$, where $0 \le j < q$ is the index satisfying $Y_0 = r_{q,j}$.

Assume inductively that we have computed u_k for $0 \le k < n < b$ and we want to compute u_n. Letting $y_n \equiv y \cdot g^{-s_n} \bmod p$, where $s_n = \sum_{k=0}^{n-1} u_k q^k$, it follows from (19.2) that $y_n \equiv g^{u_n q^n + M_n q^{n+1}} \bmod p$, for some integer M_n. Thus, writing equivalences mod p,, we define $0 \le Y_n < p$ by

$$Y_n \equiv y_n^{(p-1)/q^{n+1}} \equiv \left(g^{u_n q^n + M_n q^{n+1}}\right)^{(p-1)/q^{n+1}} = g^{u_n(p-1)/q} \cdot g^{M_n(p-1)} \equiv g^{u_n(p-1)/q}\,.$$

and so $u_n = j$, where $0 \le j < q$ is the index satisfying $Y_n = r_{q,j}$.

We will now illustrate the Pohlig-Hellman algorithm with an example in which we use the same notation as above.

Example 19.3. Let us solve the equation $3 \equiv 2^x \pmod{19}$ for x by using the Pohlig-Hellman algorithm. Thus, $y = 3$, $g = 2$, and $p = 19$.

First, we compute $\phi(19) = 18$ and write it in its canonical form $18 = 2 \cdot 3^2$.

We now calculate the qth roots of unity for $q \in \{2,3\}$; all equivalences in these calculations are mod 19.

For $q = 2$:
$r_{2,0} \equiv 2^{0\cdot 18/2} \equiv 1$ and $r_{2,1} \equiv 2^{1\cdot 18/2} \equiv 2^9 \equiv -1$.

For $q = 3$:
$r_{3,0} \equiv 2^{0\cdot 18/3} \equiv 1$, $r_{3,1} \equiv 2^{1\cdot 18/3} \equiv 2^6 \equiv 7$, and $r_{3,2} \equiv 2^{2\cdot 18/3} \equiv 2^{12} \equiv 11$.

For each $(q,b) = (q_i, a_i)$, we now wish to find $1 \le u < q^b$ so that $x \equiv u \bmod q^b$.

For $q = 2$, we calculate

$$y^{(p-1)/q} \equiv 3^9 \equiv -1 \bmod 19\,.$$

Since $r_{2,1} \equiv -1$, we see that $u = u_0 = 1$, so $x \equiv 1 \bmod 2$.

For $q = 3$, we wish to find $1 \le u < 3^2$ so that $x \equiv u \bmod 3^2$. We write $u = u_0 + 3u_1$, where $0 \le u_k < 3$ for $k \in \{0,1\}$. We first calculate

$$y_0^{(p-1)/q} \equiv 3^6 \equiv 7 \pmod{19}.$$

Since $r_{3,1} \equiv 7$, we have $u_0 = 1$.

We now calculate

$$y_1 \equiv y \cdot g^{-u_0} \equiv 3 \cdot 2^{-1} \equiv 3 \cdot 10 \equiv 11\,, \bmod 19$$

and so

$$y_1^{(p-1)/q^2} \equiv 11^{18/9} \equiv 11^2 \equiv 7 \bmod 19\,.$$

From $r_{3,1} \equiv 7$, we deduce that $x_1 = 1$. Hence, $x \equiv 1 + 1 \cdot 3 \equiv 4 \bmod 9$. Thus, we want to find x such that

$$x \equiv 1 \bmod 2 \quad \text{and} \quad x \equiv 4 \bmod 9$$

The solution is of course $x \equiv 13 \bmod 19$. Indeed, it is clear that $3 \equiv 2^{13} \pmod{19}$.

19.3. Shank's baby step giant step algorithm.

Shank's baby step giant step algorithm is a rather straightforward approach to attacking discrete logarithmic encryption, and is named after Daniel Shanks. As before, we aim to find x such that $y \equiv g^x \bmod p$, where g is a generator of the cyclic group of order p and p is prime. We know that x must lie somewhere in the cycle

$$1 = g^0, \quad g^1, \quad g^2, \quad \ldots, \quad g^{p-2}, \quad g^{p-1}, \quad g^p \equiv 1\,.$$

To solve the equation $y \equiv g^x \pmod p$ for x, where y, g, and p are known, we follow this algorithm:

(1) Define $n = \lceil \sqrt{p} \rceil$.
(2) Reduce $g^{-1} \bmod p$, i.e. find h such that $1 \le h < p$ and $g \cdot h \equiv 1 \pmod p$.
(3) Make a table of reduced values of $y \cdot g^{-r} \bmod p$ for all $0 \le r \le n-1$.
(4) Reduce $g^n \bmod p$.
(5) Reduce $(g^n)^k \bmod p$, for successive integers $k \ge 0$, until we get a match in the table of Step (4).
(6) Use $y \cdot g^{-r} \equiv (g^n)^k \pmod p$ to solve for $x = nk + r$.

Step (3) in this algorithm is the baby step phase, as we go through all possible r in the list. Step (5) refers to the giant step phase, where we raise (g^n) to powers of k.

Unlike Pohlig-Hellman, this algorithm does not typically lead to an efficient solution to the discrete logarithm problem for large p. Thus, we only need to pick p large to protect against it. The previously discussed protection against a Pohlig-Hellman attack will also protect against a baby step giant step attack.

Example 19.4. We use Shank's baby step giant step algorithm to solve the equation $3 \equiv 2^x \pmod{19}$. We are using the same equation as in Example 19.3 to illustrate the differences between the two attacks.

Let $n = \lceil \sqrt{19} \rceil = 5$.

We calculate $g^{-1} \equiv 2^{-1} \equiv 10 \bmod 19$. We now reduce $y \cdot g^{-r} \equiv 3 \cdot 10^r \bmod 19$ for $0 \leq r \leq 4$:

$$\begin{aligned} 3 \cdot 2^0 &\equiv 3 \mod 19 \\ 3 \cdot 2^{-1} &\equiv 11 \mod 19 \\ 3 \cdot 2^{-2} &\equiv 15 \mod 19 \\ 3 \cdot 2^{-3} &\equiv 17 \mod 19 \\ 3 \cdot 2^{-4} &\equiv 18 \mod 19 \end{aligned}$$

With this list of values in hand, we next compute $g^n \equiv 2^5 \equiv 13 \bmod 19$. We then start reducing $(g^n)^k \bmod p$ for $k \geq 0$, and compare the solutions with our list of values:

$$\begin{aligned} (2^5)^0 &\equiv 1 \mod 19 \\ (2^5)^1 &\equiv 13 \mod 19 \\ (2^5)^2 &\equiv 17 \mod 19 \end{aligned}$$

Since $(2^5)^2 \equiv 17 \bmod 19$ gives us a match, we have

$$3 \cdot 2^{-3} \equiv (2^5)^2 \bmod 19\,,$$

and so $x = 5 \cdot 2 + 3 = 13$. And indeed, if we check the solution, we get $3 \equiv 2^{13} \pmod{19}$. Note that we have obtained the same solution as in Example 19.3.

Bibliography

[1] D. Boneh, *Twenty years of attacks on the RSA cryptosystem*, available online: `https://crypto.stanford.edu/~dabo/pubs/papers/RSA-survey.pdf` (accessed 27 March 2023).

[2] R.P. Brent, *An improved Monte Carlo factorization algorithm*, BIT **20** (1980), 176–184.

[3] J.T. Cross, *The Euler ϕ-Function in the Gaussian Integers*, Amer. Math. Monthly **90** (1983), 518–528.

[4] W. Diffie and M.E. Hellman, *New Directions in Cryptography*, IEEE Trans. Inf. Theory **22** (1976), 644–654.

[5] G. Dresden and W.M. Dymàcek, *Finding Factors of Factor Rings over the Gaussian Integers*, Amer. Math. Monthly **112** (2005), 602–211.

[6] D.S. Dummit and R.M. Foote, *Abstract Algebra*, 3rd edition. J. Wiley & Sons Inc., Hoboken, NJ, 2004.

[7] P. Erdős, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen. **4** (1956), 201–206.

[8] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*, 2nd. edition. Taylor & Francis, Boca Raton, Florida, 2015.

[9] N. Koblitz, *A Course in Number Theory and Cryptography*, 2nd edition, Springer, New York, NY, 1994.

[10] C. Paar and J. Pelzl, *Understanding Cryptography: A Textbook for Students and Practitioners*, Springer, Berlin Heidelberg, 2010.

[11] S. Pohlig and M. Hellman *An Improved Algorithm for Computing Logarithms over GF(p) and its Cryptographic Significance* IEEE Trans. Inf. Theory **24** (1978), 106–110.

[12] J.M. Pollard, *A Monte Carlo Method for Factorization*, BIT **15** (1975), 331–334.

[13] M.O. Rabin, *Probabilistic algorithm for testing primality*, J. Number Theory **12** (1980), 128–138.

[14] R.L. Rivest, A. Shamir, L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Comm. ACM **21** (1978), 120–126.

[15] R.G. Stein, *Exploring the Gaussian Integers*, Two-Year Coll. Math. J. **7** (1976), 4–10.