NUMBER THEORETIC BACKGROUND

ZEÉV RUDNICK

Contents

Date: March 5, 2002.

1. Preface

This paper is an expanded version of some of the lectures given at the summer school in Bologna. In these lectures we gave an introduction to very basic number theory, assuming practically no background. The lectures were intended for graduate students in Math and Physics and while the material is completely standard, we tried to make the presentation as elementary as possible.

Some of the easier proofs are included, others are relegated to exercises, but several of the deeper facts are stated without proof. Most of the material may be found in classic texts such as [1] and [3].

2. Divisibility

2.1. **Basics on divisibility.** We denote by $\mathbf{Z} = \{\dots, -2, -2, 0, 1, 2, \dots\}$ the set of integers.

The Euclidean property: If $b \neq 0$ then for any a, we can write

$$a = qb + r$$

with remainder $0 \leq r < |b|$ and quotient q. [1]

Given a pair of integers a, b with $b \neq 0$, we say that b divides a, denoted as $b \mid a$, if the remainder $r = 0$, that is if $a = bq$ for some integer q. We will also say that b is a divisor of a.

The basic properties of the divisibility relation are

1. $b \mid 0$ for all nonzero b and $1 \mid a$ for all a.
2. Transitivity: $b \mid a$ and $c \mid b$ implies $c \mid a$.
3. if $d \mid a$ and $d \mid b$ then $d \mid ax + by$ for all integers x, y.
4. units: $a \mid 1$ if and only if $a = \pm 1$. Indeed, since nonzero integers have absolute value at least 1, the only solutions of the equation $xy = 1$ in integers are $(x, y) = (1, 1)$ or $(-1, -1)$.
5. For nonzero integers a, b we have both $a \mid b$ and $b \mid a$ if and only if $b = \pm a$.

   Indeed, if $b \mid a$ the we can write $a = bx$ for some integer x, and from $a \mid b$ we can write $b = ay$ for some $y \in \mathbf{Z}$. Thus $axy = a$ and since $a \neq 0$ this means $xy = 1$ which forces $x = \pm 1$.

2.2. **The greatest common divisor.** Given a pair of integers a *common divisor* is an integer d which divides both. For instance the common divisors of 4 and 6 are the integers $\pm 1, \pm 2$.

Definition 2.1. *A* greatest common divisor *of a pair of nonzero integers a, b is a common divisor d which is maximal in the sense that if δ is any common divisor of a, b then $\delta \mid d$.*

[1] or $-|b|/2 < r \leq |b|/2$.

An inspection shows that the greatest common divisors of 4 and 6 are ± 2.

The basic fact is the existence (not apriori obvious from our definition) and essential uniqueness of greatest common divisors. We will denote by $\gcd(a, b)$ a choice of greatest common divisor, which is unique if require that it be positive:

Theorem 2.2. *Any two nonzero integers a, b admit a greatest common divisor, unique up to sign. Furthermore, one can always find integers x, y so that $\gcd(a, b) = ax + by$.*

Below we will see a proof of this which gives an efficient algorithm for both computing the gcd and finding integer solutions of the equation $\gcd(a, b) = ax + by$.

We first derive a few properties of the gcd:

Definition 2.3. *A pair of integers a, b is coprime if $\gcd(a, b) = 1$*

A useful criterion for coprimality is

Lemma 2.4. *a, b are coprime if and only if there are integers x, y such that $ax + by = 1$.*

We will need to use the following:

Lemma 2.5. *a) If a, b are coprime and $a \mid bc$ then $a \mid c$.*

b) If a, b are coprime and both divide c then their product ab divides c.

Proof. Indeed, if a, b are coprime then we can write $1 = ax + by$ for integers x, y. Multiplying this equation by c we find

$$c = a \cdot xc + y \cdot bc \; . \tag{2.1}$$

For part (a), we are assuming $a \mid bc$ and so both summands on the RHS of (2.1) are divisible by a, hence so is the LHS, namely c.

For part (b), we are assuming that $a \mid c$ and so $ab \mid yb \cdot c$ and likewise since we assume that $b \mid c$, we have $ab \mid ax \cdot c$ and thus ab divides the LHS of (2.1) and so divides the LHS, namely c. □

2.3. **The Euclidean algorithm.** The Euclidean algorithm gives an efficient method for finding the gcd as well as for computing x, y so that $\gcd(a, b) = ax + by$. The method is as follows: Assume $|a| \geq |b|$. It will be convenient to set $r_{-1} := |a|$, and $r_0 := |b|$. We use division with remainder to write

$$a = q_1 b + r_1, \qquad 0 \leq r_1 < |b| \; .$$

If $r_1 = 0$ then $b \mid a$ and $\gcd(a, b) = b$. Otherwise, iterate this step with a replaced by b and b replaced by the remainder r_1 to write

$$b = q_2 r_1 + r_2, \quad 0 \le r_2 < r_1 .$$

Continuing in this fashion, we get after k steps

$$r_{k-2} = q_k r_{k-1} + r_k, \qquad 0 \le r_k < r_{k-1} .$$

Since the sequence of remainders $|b| = r_0 > r_1 > r_2 > \dots$ is a strictly decreasing sequence of non-negative integers, this process has to terminate in a finite number of steps, say after n steps we have

$$r_{n-2} = q_n r_{n-1} + r_n, \qquad r_n \ne 0$$

and

$$r_{n-1} = q_{n+1} r_n$$

We claim that

$$\gcd(a, b) = r_n .$$

Moreover, the process gives integers x, y so that

$$\gcd(a, b) = ax + by .$$

To see this, we will show by descending induction on $i = n, n-1, \dots, 0, -1$ that

$$r_i \mid r_n \tag{2.2}$$

and that

$$r_n = x_i r_{i-1} + y_i r_i \tag{2.3}$$

Once we have this, taking $i = 0, -1$ will give $r_n \mid r_0 = b$ and $r_n \mid r_{-1} = a$, and taking $i = 0$ in (2.3) will give $r_n = x_0 a + y_0 b$.

For $i = n$ we clearly have $r_n \mid r_n$ and $r_n = 1r_n + 0r_{n-1}$. For $i = n-1$ we have $r_{n-1} = q_n r_n$ giving both (2.2) and (2.3) in that case. Assuming we know that $r_n \mid r_k$ and $r_n \mid r_{k-1}$, we use $r_{k-2} = q_k r_{k-1} + r_k$ to find that $r_n \mid r_{k-2}$. Further, assuming we know (2.3) for $i = k$ gives

$$\begin{aligned} r_n &= x_k r_{k-1} + y_k r_k \\ &= x_k r_{k-1} + y_k (r_{k-2} - q_k r_{k-1}) \\ &= y_k r_{k-2} + (x_k - q_k y_k) r_{k-1} \end{aligned}$$

that is (2.3) holds for $i = k-1$ with $x_{k-1} = y_k$, $y_{k-1} = x_k - q_k y_k$.

Example: $a = 8$, $b = 5$: To compute $\gcd(8,5)$, we carry out the following steps:

$$\begin{aligned}
8 &= 1 \cdot 5 + 3 \\
5 &= 1 \cdot 3 + 2 \\
3 &= 1 \cdot 2 + 1 \\
2 &= 2 \cdot 1 \ .
\end{aligned}$$

Thus $\gcd(8,5) = 1$. To find integer solutions of $8x+5y=1$, we proceed backwards:

$$\begin{aligned}
1 &= 3 - 1 \cdot 2 \\
&= 3 - (5 - 1 \cdot 3) = 2 \cdot 3 - 5 \\
&= 2 \cdot (8 - 1 \cdot 5) - 5 = 2 \cdot 8 - 3 \cdot 5
\end{aligned}$$

and thus we found the solution $x = 2$, $y = -3$ to $8x + 5y = 1$.

An examination of the Euclidean algorithm shows that the number of steps is at most $2\log_2 |b| + 1 = O(\log \min(|a|, |b|))$, that is the complexity is **linear** in the number of bits needed to represent the input.

Exercise 2.6. *Prove this estimate on the number of steps.*

Hint: Show that $r_k \le r_{k-2}/2$.

Exercise 2.7. *Show that one cannot significantly improve this estimate.*

Hint: Take a, b to be consecutive Fibonacci numbers. These are defined recursively as $F_0 = 0$, $F_1 = 1$, and for $n \ge 1$ by $F_{n+1} = F_n + F_{n-1}$. Show that the number of steps for computing $\gcd(F_{n+1}, F_n)$ is n (?) and that $n \sim \log F_n / \log((1+\sqrt{5})/2)$ as $n \to \infty$.

2.4. **The Diophantine equation $ax + by = c$.** Given integers a, b, c, we wish to find *integer* solutions to the equation

$$ax + by = c \ .$$

As the example $4x + 6y = 1$ illustrates, such solutions need not exist, since the RHS is odd while the LHS is even! More generally, if $\gcd(a,b)$ does not divide c then there will be no integer solutions. It turns out that this divisibility condition is the only obstruction to the existence of solutions, and once this obstruction vanishes then there are infinitely many integer solutions:

Theorem 2.8. *The equation $ax + by = c$ has integer solutions if and only if* $\gcd(a,b) \mid c$.

If there is one solution (x_0, y_0) *then there are infinitely many integer solutions, and they are all given by*

$$x_k = x_0 + k\frac{b}{\gcd(a,b)}, \quad y_k = y_0 - k\frac{a}{\gcd(a,b)}$$

where k *runs over all integers.*

The proof of this theorem is very easy once we know the existence of one solution. To find a solution, first use the Euclidean algorithm to solve the equation $au + bv = \gcd(a, b)$ and then take

$$x_0 = u\frac{c}{\gcd(a,b)}, \quad y_0 = v\frac{c}{\gcd(a,b)} .$$

3. Prime Numbers

3.1. **The fundamental theorem of arithmetic.** A *prime* is a natural number $p > 1$ which has no proper divisors (that is no divisors other than $\pm p$ and ± 1).

The sequence of primes is thus $p_1 = 2,\ p_2 = 3,\ p_3 = 5, \dots$.

Lemma 3.1. *If* p *is prime which divides a product:* $p \mid bc$*, then it has to divide one of the factors:* $p \mid b$ *or* $p \mid c$*.*

Indeed, if p is prime and p does not divide b then automatically p and b are coprime, hence the result follows from Lemma 2.5.

We claim that every integer $n > 1$ is a product of primes: Indeed, $n > 1$ is either a prime, in which case we are done, or decomposable: $n = ab$, with $a, b > 1$. In the latter case, we have $a, b < n$ and arguing by induction, both a, b are products of primes and hence so is $n = ab$.

Since every integer is a product of primes, they are thus the building blocks of all the integers. In fact, more is true - the factorization into products of primes is *unique*:

Theorem 3.2 (Fundamental Theorem of Arithmetic)**.** *Every natural number is* uniquely *decomposable into a product of prime powers.*

Exercise 3.3. *Show that if the prime factorization of a pair of integers is given by* $a = \prod p^{\alpha(p)}$ *and* $b = \prod p^{\beta(p)}$ *then*

$$\gcd(a, b) = \prod p^{\min(\alpha(p),\beta(p))} .$$

While we know that every integer factors into a product of primes, a basic problem is how to find this factorization efficiently. Currently, there is no known algorithm which will give the prime factorization of an integer in a number of steps which is polynomial in the input (quantum algorithms aside).

TABLE 1. Comparison between $\pi(x)$ and $\mathrm{Li}(x)$.

x	$\pi(x)$	$[\mathrm{Li}(x)] - \pi(x)$
10^8	5,761,455	754
10^{10}	455,052,511	3,104
10^{12}	37,607,912,018	38,263
10^{14}	3,204,941,750,802	314,890
10^{16}	279,238,341,033,925	3,214,632

3.2. **There are infinitely many primes.**

Theorem 3.4 (Euclid). *There are infinitely many primes.*

Proof. Argue by *reductio ad absurdum*: If there were finitely many primes, say only M of them, then form the integer $Q = p_1 \cdot p_2 \cdot \dots \cdot p_M + 1$. It is either a prime or decomposable. Since Q is greater than all the primes $p_1, \dots, p_M$, it cannot be a prime. However, Q clearly leaves remainder 1 on division by each of the available primes p_i, and thus being divisible by no prime, cannot decompose into a product of primes! We thus arrive at a contradiction. ☐

Exercise 3.5. *Show that there infinitely many primes of the form* $4k+3$.

3.3. **The density of primes.** After knowledge that there are infinitely many primes, one can try to assess their density. Gauss recounted that in 1792, as a boy of 15, he arrived at the conjecture that the density of primes near x is about $1/\log x$ and so if we denote by $\pi(x)$ the number of primes up to x

$$\pi(x) := \#\{n : p_n \leq x\}$$

then $\pi(x)$ is asymptotically equal to the logarithmic integral, given for $x > 2$ by

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

In turn, $\mathrm{Li}(x)$ has an asymptotic expansion

$$\mathrm{Li}(x) = \frac{x}{\log x} + \frac{x}{(\log x)^2} + \dots + c_n \frac{x}{(\log x)^n} + O(\frac{x}{(\log x)^{n+1}})$$

To check the strength of $\mathrm{Li}(x)$ as an approximation to $\pi(x)$, we examine Table 1 (writing $[y] :=$ integer part of y). As is seen from this table, $\mathrm{Li}(x)$ is a remarkably good approximation to $\pi(x)$ in this range.

As a measure of the quality of the approximation, note that the width of the third column is about a half of the width of the second one, that is to say that the remainder is approximately square root of the main term!

The statement that $\pi(x) \sim \mathrm{Li}(x)$ is known as the Prime Number Theorem. It was proved in 1896 by Hadamard and de la Vallée Poussin, by using the Riemann zeta function. The empirical statement made above from the data in Table 1 as to the magnitude of the remainder in this approximation is a form of the celebrated Riemann Hypothesis, see Section 8.2.

3.4. **Primes in arithmetic progressions.** An important issue is the existence of primes in a given arithmetic progression: Given a and $q > 1$, to find a large prime p with $p \equiv a \mod q$. Clearly, in some instances it cannot be done, say the progression $\{2, 4, 6, 8, \dots\}$ contains no large primes as all primes except 2 are odd. Likewise, if a and q have a common factor $d > 1$ then it divides every element of the progression $a, a+q, a+2q \dots$ and so there are no primes in it (excepting perhaps if $a = d$ is prime). We should thus restrict attention to the case that a and q are co-prime. It turns out that this is the only obstruction to the existence of primes in arithmetic progression, as was proved by Dirichlet in 1837. In fact there are arbitrarily large primes in every progression not excluded by such reasoning:

Theorem 3.6 (Dirichlet's Theorem). *For $q > 1$ and any a co-prime to q, there are infinitely many primes of the form $a + kq$.*

One can try to give an argument for this along the lines of Euclid's argument for the existence of infinitely many primes (Theorem 3.4). This works in a few cases of small q, and for some special progressions such as $p \equiv 1 \mod q$, but this line of attack has not yielded Dirichlet's theorem in its full force.

A quantitative version of Dirichlet's theorem is the Prime Number Theorem for arithmetic progressions, which asserts that for fixed $q > 1$, every progression $a \mod q$ has asymptotically the same density of primes. Thus setting $\pi(x; q, a) := \#\{p \le x : p \equiv a \mod q\}$, we have for a coprime to q

$$\pi(x; q, a) \sim \frac{1}{\phi(q)} \mathrm{Li}(x) .$$

4. Continued Fractions

Continued fractions are expressions such as

$$1+\cfrac{1}{2+\cfrac{1}{3+\cfrac{1}{4}}}\,.$$

We now study them systematically.

Given integers $a_0 \in \mathbf{Z}$, $a_1, a_2, \dots \geq 1$, consider the finite continued fraction

$$[a_0; a_1, \dots, a_m] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_m}}}}\,.$$

To compute this fraction, one defines integers p_m, q_m by the recursion ($m \geq 1$):

$$p_m = a_m p_{m-1} + p_{m-2}$$
$$q_m = a_m q_{m-1} + q_{m-2}$$

with $p_{-1} = 1$, $p_0 = a_0$, $q_{-1} = 0$, $q_0 = 1$. These satisfy the relations

$$p_m q_{m-1} - p_{m-1} q_m = (-1)^{m-1}$$

and

$$p_m q_{m-2} - p_{m-2} q_m = (-1)^m a_m\,.$$

On then shows that

$$[a_0; a_1, \dots, a_m] = p_m / q_m\,.$$

The infinite simple continued fraction $[a_0; a_1, a_2, \dots]$ is the limit of the "convergents" p_m/q_m. Every irrational α has a unique continued fraction expansion.

Example: We will obtain the continued fraction expansion of the quadratic irrationality $\sqrt{3}$: Since $\sqrt{3}$ lies between 1 and 2, we write $\sqrt{3} = 1 + 1/x_1$ with

$$x_1 = \frac{1}{\sqrt{3}-1} = \frac{\sqrt{3}+1}{2}$$

Thus $x_1 = 1 + 1/x_2$ with

$$x_2 = \frac{2}{\sqrt{3}-1} = \sqrt{3}+1$$

and thus we can write $x_2 = 2 + 1/x_3$ with

$$x_3 = \frac{1}{\sqrt{3}-1} = x_1 \ .$$

The procedure has thus cycled back and we may continue it indefinitely to find that $\sqrt{3}$ has the *periodic* continued fraction expansion

$$\sqrt{3} = 1 + \cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{\ddots}}} = [1;1,2,1,2,\dots] = [1;\overline{1,2}] \ .$$

This is something of a rarity; it turns out that an irrational has a periodic continued fraction expansion if and only if it is a *quadratic irrationality*, that is of the form $r + s\sqrt{d}$, with r, s rational and $d > 1$ an integer which is not a perfect square.

The convergents give very good rational approximations to α: We have

$$\frac{1}{2}\frac{1}{q_m q_{m+1}} < |\alpha - \frac{p_m}{q_m}| < \frac{1}{q_m q_{m+1}} \ .$$

The convergents p_m/q_m are the "best" rational approximations to α, in the following senses: If p/q satisfies $|\alpha - p/q| < 1/2q^2$ then $p/q = p_m/q_m$ for some m. Moreover, for $m > 1$, if $0 < q \le q_m$ and $p/q \ne p_m/q_m$ then $|\alpha - p/q| > |\alpha - p_m/q_m|$.

5. Modular arithmetic

5.1. **Congruences.** Given $N > 1$, we say that two integers a, b are congruent modulo N, written as $a \equiv b \mod N$, if $N \mid a - b$.

Congruence modulo N is an "equivalence relation" on the set of integers, that is as follows immediately from the definition,

1. $a \equiv a \mod N$
2. $a \equiv b \mod N$ if and only if $b \equiv a \mod N$
3. $a \equiv b \mod N$ and $b \equiv c \mod N$ implies $a \equiv c \mod N$

Given $N > 1$, every integer is congruent to precisely one of the N integers $\{0, 1, \dots, N-1\}$, as is seen by writing $a = qN + r$ with remainder $0 \le r < N$. Thus denoting by $\mathbf{Z}/N\mathbf{Z}$ the set of congruence classes of integers modulo N, we see that a complete set of representatives can be taken to be $\{0, 1, \dots, N-1\}$, though other choices are equally valid.

The set of congruence classes modulo N also admits algebraic operations of addition and multiplication. To see this, one needs to check that if $a \equiv a' \mod N$ and $b \equiv b' \mod N$ then $a + b \equiv a' + b' \mod N$

TABLE 2. Multiplication modulo 4

	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

TABLE 3. Multiplication modulo 5

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

and likewise $a \cdot b \equiv a' \cdot b' \mod N$, that is the sum/product of congruence classes is independent of the choice of representatives.

Example: See tables 2 and 3 for the multiplication tables modulo 4 and 5.

The existence of addition and multiplication satisfy the usual laws of integer arithmetic, that is commutativity, associativity, distributivity etc. This is formally expressed by saying that $\mathbf{Z}/N\mathbf{Z}$ is a *"ring"*.

5.2. **Modular inverses.** An integer a is *invertible modulo* N if there is an integer b so that $ab \equiv 1 \mod N$. We say that b is an inverse of a modulo N, denoted by $b \equiv a^{-1} \mod N$.

For instance 1 is always invertible and $1^{-1} \equiv 1 \mod N$. Invertibility makes sense for congruence classes modulo N (why?). We will denote the set of invertible residue classes modulo N by $(\mathbf{Z}/N\mathbf{Z})^*$.

Examining the multiplication tables (2), (3), we see that modulo 4, the invertible residue classes are $1, 3$ with $3^{-1} \equiv 3 \mod 4$ while all nonzero residue classes modulo 5 are invertible, and $2^{-1} \equiv 3 \mod 5$, $3^{-1} \equiv 2 \mod 5$ and $4^{-1} \equiv 4 \mod 5$.

As these examples indicate, an inverse modulo N, of it exists, is *unique* modulo N. Indeed, if $ab \equiv 1 \mod N$ and $ac \equiv 1 \mod N$ then using commutativity and associativity of modular multiplication we have

$$b \equiv b \cdot 1 \equiv b \cdot ac \equiv (ab) \cdot c \equiv 1 \cdot c \equiv c \mod N$$

and thus $b \equiv c \mod N$ as claimed.

Moreover the product of invertible classes is still invertible. Thus we get a commutative group structure on the set $(\mathbf{Z}/N\mathbf{Z})^*$ of invertible residue classes modulo N. We will hence refer to $(\mathbf{Z}/N\mathbf{Z})^*$ as the *multiplicative group modulo N.*

A criterion for invertibility modulo N, as well as an algorithm for finding the modular inverse, is given by

Lemma 5.1. *A necessary and sufficient condition for an integer a to be invertible modulo N is that a and N are coprime:* $\gcd(a, N) = 1$.

Indeed, a and N are coprime if and only if we can solve $ax + Ny = 1$ (Lemma 2.4), and the latter equation is equivalent to solubility of the congruence $ax \equiv 1 \mod N$.

As an immediate corollary, we see that if N is *prime* then all nonzero residue classes modulo N are invertible, since the only divisors of N are 1 and N and so the only alternative to $\gcd(a, N) = 1$ in this case is $\gcd(a, N) = N$, and thus $a \equiv 0 \mod N$.

Moreover, as described in (2.3), the Euclidean algorithm provides for an efficient method of finding a solution of $ax + Ny = 1$, that is of polynomial time in the input, and thus an efficient method for finding modular inverses.

Definition 5.2. *Euler's totient function $\phi(N)$ is the number of invertible residue classes modulo N.*

As we saw above, of p is prime then all nonzero residue classes mod p are invertible and thus $\phi(p) = p - 1$ in this case.

Exercise 5.3. *Show that for prime p we have $\phi(p^k) = p^k - p^{k-1}$ ($k \geq 1$).*

5.3. The Chinese Remainder Theorem.

Given $x \mod mn$, we get a pair of residue classes $(x \mod m, x \mod n)$ in $\mathbf{Z}/mZ \times \mathbf{Z}/n\mathbf{Z}$. If m, n are coprime, then we may recover $x \mod mn$ from this pair.

Theorem 5.4 (CRT)**.** *If m, n are coprime then for a, b we can solve the congruence*

$$\begin{cases} x \equiv a \mod m \\ x \equiv b \mod n \end{cases} \tag{5.1}$$

the solution is unique modulo mn.

Proof. To see uniqueness, note that if x, x' are solutions of (5.1), then both m and n divide $x - x'$ and since m, n are coprime this forces $mn \mid x - x'$, that is $x \equiv x' \mod mn$.

TABLE 4. Orders of elements mod 5

a	1	2	3	4
$\mathrm{ord}(a,5)$	1	4	4	2

The method of solution is constructive: Since m, n are coprime, m is invertible modulo n. Denote by $\bar{m}$ the inverse of $m \mod n$, and likewise let $\bar{n}$ be the inverse of $n \mod m$. Then a solution of the system of congruences (5.1) is given by

$$x = n\bar{n}a + m\bar{m}b \mod mn .$$

Indeed, modulo m we have $n\bar{n} \equiv 1 \mod m$ while $m \equiv 0 \mod m$ and thus $x = n\bar{n}a + m\bar{m}b \equiv 1 \cdot a + 0 \cdot \bar{m}b \equiv a \mod m$, and likewise $x \equiv b \mod n$. □

Exercise 5.5. *If $x \equiv a \mod 3$, $x \equiv b \mod 5$, $x \equiv c \mod 7$ find $x \mod 105$.*

Exercise 5.6. *Show that if m, n are coprime then x is invertible modulo mn if and only if it is invertible both modulo m and modulo n.*

Exercise 5.7. *Show that if m, n are coprime then $\phi(mn) = \phi(m)\phi(n)$.*

Exercise 5.8. *Show that for $N > 1$ we have $\phi(N) = N\prod_{p|N}(1 - 1/p)$ where the product is over all primes dividing N.*

5.4. **The structure of the multiplicative group $(\mathbf{Z}/N\mathbf{Z})^*$.** In this section, we will investigate the structure of multiplication in the group of invertible residue classes modulo N.

Exercise 5.9. *If $a, b \in \mathbf{Z}/N\mathbf{Z}$ then we may compute $a \cdot b \mod N$ in $O(\log a \log b)$ bit operations.*

A fundamental aspect is the power operation, of raising an element to a power: $a \mapsto a^k \mod N$. It turns out that this is computationally easy:

Exercise 5.10 (Divide and conquer)**.** *Show that we can compute $a^k \mod N$ in at most $O(\log k(\log N)^2)$ elementary bit operations.*

Definition 5.11. *The order of $a \in (\mathbf{Z}/N\mathbf{Z})^*$ is the least integer $k \geq 1$ for which $a^k \equiv 1 \mod N$.*

We denote this integer by $\mathrm{ord}(a, N)$.

Exercise 5.12. *Make up tables of orders of all elements modulo* $5, 8, 11$.

That $\mathrm{ord}(a, N)$ exists is guaranteed by:

Table 5. Orders of elements mod 8

a	1	3	5	7
$\mathrm{ord}(a, 8)$	1	2	2	2

Table 6. Orders of elements mod 11

a	1	2	3	4	5	6	7	8	9	10
$\mathrm{ord}(a, 11)$	1	10	5	5	5	10	10	10	5	2

Theorem 5.13 (Fermat-Euler). *For any $a \in (\mathbf{Z}/N\mathbf{Z})^*$ we have*

$$a^{\phi(N)} \equiv 1 \mod N .$$

In particular, we see that $\mathrm{ord}(a, N) \leq \phi(N)$. In fact more is true:

Proposition 5.14. *Let $a \in (\mathbf{Z}/N\mathbf{Z})^*$ be invertible modulo N.*
a) Suppose $a^k \equiv 1 \mod N$. Then $\mathrm{ord}(a, N)$ divides k.
b) In particular, $\mathrm{ord}(a, N)$ divides $\phi(N)$.

Proof. Write $k = q\,\mathrm{ord}(a, N) + r$, with $0 \leq r < \mathrm{ord}(a, N)$. Then $1 = a^k = (a^{\mathrm{ord}(a,N)})^q \cdot a^r \equiv 1^q \cdot a^r = a^r \mod N$. Since $r < \mathrm{ord}(a, N)$, this forces $r = 0$ by the minimality of $\mathrm{ord}(a, N)$, that is $\mathrm{ord}(a, N) \mid k$. Part (b) follows from part (a) and Fermat-Euler. □

Exercise 5.15. *Let $p \neq 2, 5$ be a prime, and consider the decimal (base 10) expansion of the rational $1/p$. This expansion is periodic of the form $1/p = 0.\overline{a_1 \dots a_T}$ where T denotes the minimal period. For instance, $1/3 = 0.333\cdots = 0.\overline{3}$ ($T = 1$), $1/7 = 0.\overline{142857}$ ($T = 6$), $1/11 = 0.\overline{09}$ ($T = 2$) etc.*
Show that $T = \mathrm{ord}(10, p)$. Generalize.

5.5. **Primitive roots.** The maximal order of an invertible element is $\phi(N)$. We will say that $a \in (\mathbf{Z}/N\mathbf{Z})^*$ is a *primitive root modulo N* if $\mathrm{ord}(a, N) = \phi(N)$.

An examination of the tables of orders of elements modulo 5, 8 ,... indicates that this sometimes does indeed happen, though when the modulus is 8 the maximal order is 2 rather than $4 = \phi(8)$.

The following theorem explains this empirical finding:

Theorem 5.16. *If p is a prime then there is a primitive root modulo p.*

For composite moduli, it is relatively rare to have primitive roots:

Exercise 5.17. *a) Show that if $n > 2$ then $\phi(n)$ is even.*
b) If $m, n > 2$ are coprime then there is no primitive root modulo mn.

It turns out that there is a primitive root modulo N if and only if $N = 2, 4$ or $N = p^k, 2p^k$ where p is an odd prime.

While Theorem 5.16 guarantees the existence of a primitive root modulo a prime, one does not know of an efficient algorithm that given a (large) prime p, finds a primitive root modulo p.

In this context, there is a conjecture of Emil Artin from the 1920's which among other things says that the number 2 is a primitive root modulo infinitely many primes, and the same is true for any integer $a \neq \pm$ and not a perfect square - see the survey by Ram Murty [4] for further details.

The importance of the notion of primitive roots comes from the following observation: For $g \in (\mathbf{Z}/N\mathbf{Z})^*$, the function $x \mapsto g^x \mod N$ has period exactly $\operatorname{ord}(g, N)$ by Proposition 5.14, and in particular if g is a primitive root modulo N then this period is precisely $\phi(N)$. Thus if g is a primitive root modulo N then for any $a \in (\mathbf{Z}/N\mathbf{Z})^*$ we may write $a \equiv g^x \mod N$ with x unique modulo $\phi(N)$. Since there are $\phi(N)$ invertible residues and the same number of (necessarily invertible) powers $g^x \mod N$, we find:

Lemma 5.18. *An invertible element $g \in (\mathbf{Z}/N\mathbf{Z})^*$ is a primitive root modulo N if and only if every element $a \in (\mathbf{Z}/N\mathbf{Z})^*$ can be written as $a \equiv g^x \mod N$.*

This lemma allows us to convert modular *multiplication* in the group $(\mathbf{Z}/N\mathbf{Z})^*$ into modular *addition* in $\mathbf{Z}/\phi(N)\mathbf{Z}$, assuming we have a primitive root, since if $a \equiv g^x$ and $b \equiv g^y$ then $a \cdot b \equiv g^{x+y} \mod N$.

If $a \equiv g^x \mod N$, we will write $x := \operatorname{Ind}_g(a, N)$. We may think of $\operatorname{Ind}_g(a, N)$ as a *discrete logarithm*, since by the above reasoning $\operatorname{Ind}_g(ab) \equiv \operatorname{Ind}_g(a) + \operatorname{Ind}_g(b) \mod \phi(N)$. Given g, x it is easy to compute $g^x \mod N$, there is no known efficient method of determining $x = \operatorname{Ind}_g(a, N)$ from a, g and N. This is known as the *discrete logarithm problem.*

6. Quadratic congruences

6.1. **Euler's criterion.** Let p be an odd prime. We will study the congruence

$$x^2 \equiv a \mod p \tag{6.1}$$

where a is invertible modulo p. One algorithm for deciding the solubility of this congruence is given by Euler's criterion:

Proposition 6.1 (Euler's criterion)**.** *There is a solution of $x^2 \equiv a \mod p$ if and only if $a^{(p-1)/2} \equiv 1 \mod p$.*

Proof. Indeed, if $a \equiv x^2 \mod p$ then $a^{(p-1)/2} = x^{p-1} \equiv 1 \mod p$. For the reverse direction, we may use the existence of a primitive root g mod p to try and solve the congruence (6.1) by writing

$$a \equiv g^b \mod p, \qquad 0 \le b < p-1$$

It will suffice to show that $b \equiv 2z \mod p-1$ since then $x \equiv g^z \mod p$ solves the congruence (6.1). Substituting $a \equiv g^b \mod p$ in $a^{(p-1)/2} \equiv 1$ mod p gives

$$g^{\frac{p-1}{2}b} = 1 \mod p \, .$$

By Proposition 5.14 this forces $(p-1)b/2 = 0 \mod \operatorname{ord}(g,p)$ and since g is a primitive root, $\operatorname{ord}(g,p) = p-1$ and we find that $b \equiv 0 \mod 2$, that is $b = 2z$ as required. □

As an immediate consequence, we see that for p odd, $-1 \equiv x^2 \mod p$ if and only if $p \equiv 1 \mod 4$, since $(-1)^{(p-1)/2} = 1$ precisely in that case.

6.2. **The Legendre symbol and Quadratic Reciprocity.** For $p \neq 2$ an odd prime and a invertible modulo p, the Legendre symbol $\left(\frac{a}{p}\right)$ is defined as

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & a \equiv x^2 \mod p \\ -1 & \text{otherwise} \end{cases}$$

It is sometime convenient to extend the definition to include non-invertible residues by requiring $\left(\frac{a}{p}\right) = 0$ if $p \mid a$.

Below are some simple properties of the Legendre symbol:

1. If $a \equiv b \mod p$ then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.
2. $\left(\frac{x^2}{p}\right) = +1$. Both these follow from the very definition of the Legendre symbol.
3. Euler's criterion can be reformulated to read

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \mod p$$

This gives a computationally efficient method of finding the Legendre symbol, as we can compute $a^{(p-1)/2} \mod p$ in $O(\log^3 p)$ steps.

A simple consequence is a rule for when -1 is a square modulo p:

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} +1, & p \equiv 1 \mod 4 \\ -1, & p \equiv 3 \mod 4 \end{cases}$$

4. Multiplicativity:

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$

This follows from Euler's criterion!

More profound is the celebrated law of Quadratic Reciprocity, conjectured by Euler and proved by Gauss:

Theorem 6.2. *If $p \neq q$ are distinct odd primes then*

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

An addendum is the quadratic character of 2:

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} +1, & p \equiv \pm 1 \mod 8 \\ -1, & p \equiv \pm 3 \mod 8 \end{cases}$$

Example: To illustrate the power of the law of quadratic reciprocity, we will compute the Legendre symbol $\left(\frac{5}{p}\right)$ for primes $p \neq 2, 5$. We have

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right)(-1)^{(p-1)/2\cdot(5-1)/2} = \left(\frac{p}{5}\right) .$$

To compute $\left(\frac{p}{5}\right)$, we only need to know the remainder of p after division by 5 and then check that the invertible squares modulo 5 are 1 and 4. This gives that $\left(\frac{5}{p}\right) = 1$ if $p \equiv \pm 1 \mod 5$ and $\left(\frac{5}{p}\right) = -1$ for $p \equiv \pm 2 \mod 5$.

7. Pell's equation

Given a positive integer d, not a perfect square, we wish to find the integer solutions of the equation

$$x^2 - dy^2 = 1 .$$

This equation was studied several centuries ago. Its study in modern times was championed by Fermat (Pell had nothing to do with this equation, and owes it being named after him to Euler).

Exercise 7.1. *Show that if d is a perfect square then the equation $x^{-}dy^2 = 1$ has only finitely many integer solutions.*

Obvious solutions are $(x, y) = (\pm 1, 0)$, called the *trivial solutions*. The nontrivial solutions come in quadruples: If (x, y) are solutions then so are $(\pm x, \pm y)$. We will say that a solution is *positive* if $x, y > 0$.

TABLE 7. Solutions of $x^2 - 2y^2 = 1$

n	y_n
1	2
2	12
3	70
4	408
5	2378
6	13860
7	80782
8	470832
9	2744210
10	15994428

It turns out that nontrivial solutions always exist, though they are quite sparse. We may clearly order the positive solutions by increasing size of their y coordinate, or equivalently by the size of their x-coordinate. The first nontrivial solution (x_1, y_1) will be called the *fundamental solution.*

Example: Suppose $d = 2$. To find (positive) solutions of the corresponding Pell equation, we rewrite it as

$$1 + dy^2 = x^2$$

and the proceed to search through values of $y = 1, 2, \dots,$ to find those for which $1 + dy^2$ is a perfect square. This process quickly yields the solutions $(x, y) = (3, 2), (17, 12), (99, 70), (577, 408), \dots$. In this way we can clearly find all solutions up to any given value of y.

In Table 7 we enumerate the y-coordinate of first few positive solutions (x_n, y_n) in the case $d = 2$. One can clearly see an exponential increase of y_n with n.

Here it was easy to find the first few solutions by a brute-force search. This is not always the case. For instance, for $d = 61$ the equation was already treated in the 12-th century in India, and it was found that the fundamental solution has $y_1 = 226, 153, 980$. It is unlikely that this was found by hand merely by brute force!

Another example of a large fundamental solution for a small value of d is given by $d = 109$, when $y_1 = 15, 140, 424, 455, 100$.

7.1. **The group law.** A remarkable feature of the Pell equation is the existence of a composition law on the set of solutions, turning them into a commutative group. One way of seeing this is to first endow the set of *real* solutions with a group law. Geometrically, the real solutions form a hyperbola $x^2 - dy^2 = 1$ with two sheets, and we do this just for the right sheet $(x > 0)$, via the following parameterization by means of the hyperbolic functions:

$$x(t) := \cosh(t), \quad y(t) := \frac{\sinh(t)}{\sqrt{d}}$$

If we denote by $P_t = (x(t), y(t))$ the point corresponding to t then the group law is the one inherited from the additive group of the reals, namely

$$P(t) * P(t') := P(t + t') .$$

More generally, we can write any real solution as $\pm P(t)$ and then declare the group law to be $\epsilon P(t) * \epsilon' P(t') := \epsilon\epsilon' P(t + t')$, $\epsilon, \epsilon' = \pm 1$.

In this form this does little except to demonstrate the existence of the group law, since recovering $t+t'$ from t and t' involves a transcendental inversion problem. However, we may use the addition formulae for the hyperbolic functions to compute the x and y coordinates of the composition. If we set $P = (x, y)$, $P' = (x', y')$ then $P * P' = (x'', y'')$ with

$$x'' = xx' + dyy', \qquad y'' = xy' + x'y . \tag{7.1}$$

In particular, the inverse of (x, y) is $(x, -y)$.

Note that from (7.1) it is not transparent to see that the addition law is associative!

In the form (7.1) we immediately see that the composition of *rational* solutions is still rational and ditto for *integer* solutions, which are our goal. Thus we see that the integer solutions form a group.

An easy way to recall the composition law (7.1) for rational solutions is to map them to quadratic irrationalities: $(x, y) \mapsto \alpha = x + \sqrt{d}y$, in which case composition is given by ordinary multiplication.

Rational solutions are easy to find by the "secant method".

Exercise 7.2. *Show that all rational solutions of* $x^2 - dy^2 = 1$ *are given by* $(-1, 0)$ *together with*

$$\{(\frac{1 + dt^2}{1 - dt^2}, \frac{2t}{1 - dt^2}) : t \ \textit{rational}\} . \tag{7.2}$$

Hint: Start with the trivial solution $(-1, 0)$. Given a point $(x, y) \neq (-1, 0)$ on the hyperbola $x^2 - dy^2 = 1$, draw the line connecting the two points $(-1, 0)$ and (x, y). This line will intersect the y-axis at a point

$(0, t)$. Show that $y = t(x + 1)$ and substitute back into the original equation $x^2 - dy^2 = 1$ to find the expression (7.2). Argue that as t varies over all rationals, we get all the rational solutions other than $(-1, 0)$.

7.2. **Integer solutions.** As we saw on the basis of examples, integer solutions are harder to come by. From the shape of the composition law (7.1) we see that if there is one nontrivial integer solution, then there are automatically *infinitely many* integer solutions: We may assume that we have a positive solution $P = (x_1, y_1)$, $x_1, y_1 > 0$ and then composing it with itself we get the solutions $P^{*2} = P * P, \ldots, P^{*n} = P^{*(n-1)} * P = (x_n, y_n)$ and from (7.1) its is clear that $x_n \to \infty$ as $n \to \infty$.

Theorem 7.3. *a) If $d > 0$ is not a perfect square then there are infinitely many integer solutions of Pell's equation $x^2 - dy^2 = 1$.*

b) If we denote by $\epsilon_d = (x_1, y_1)$ the fundamental solution then all integer solutions are given by $x + \sqrt{d}y = \pm\epsilon_d^n$, $n \in \mathbf{Z}$.

7.3. **Finding the fundamental solution.** While Theorem 7.3 guarantees the existence of solutions and that all solutions are found from knowledge of the fundamental solution, it tells us nothing about how to find the fundamental solution. One method of course is to search as described above. However there is an alternative method which is more efficient and involves the continued fraction expansion of $\sqrt{d}$. The situation is as follows: The continued fraction expansion of $\sqrt{d}$ is periodic, of the form:

$$\sqrt{d} = [a_0; \overline{a_1, \ldots, a_h}]$$

where h is the minimal period of the expansion. Let p_{h-1}/q_{h-1} be the $(h-1)$-st partial convergent. Then

$$p_{h-1}^2 - dq_{h-1}^2 = (-1)^h.$$

If h is *even* then the fundamental solution is (p_{h-1}, q_{h-1}).

If h is *odd* then the fundamental solution ϵ_d is given by the $(2h-1)$-st partial convergent, and moreover

$$\epsilon_d = p_{2h-1} + \sqrt{d}q_{2h-1} = (p_{h-1} + \sqrt{d}q_{h-1})^2 .$$

Examples: $d = 7$: Then $\sqrt{7} = [2; \overline{1, 1, 1, 4}]$ has period $h = 4$. The 3-rd partial convergent is $[2; \overline{1, 1, 1}] = 8/3$ and the fundamental solution is $(8, 3)$.

$d = 61$: Then $\sqrt{61} = [7; \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}]$ and this allows us to compute the (large) fundamental solution.

8. The Riemann zeta function

The Riemann zeta function is defined for complex s with $\mathrm{Re}(s) > 1$ by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \, .$$

We give an introduction to its basic properties (see [2]).

A basic fact is Euler's product formula, which displays the connection between $\zeta(s)$ and primes:

Theorem 8.1. *For* $\mathrm{Re}(s) > 1$, $\zeta(s)$ *can be represented by the convergent product over all primes:*

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}} \, .$$

Proof. The idea is to expand each factor $(1 - p^{-s})^{-1}$ as a geometric series

$$\frac{1}{1 - p^{-s}} = \sum_{k=0}^{\infty} \frac{1}{p^{ks}}$$

and to multiply together the resulting series

$$\prod_p \frac{1}{1 - p^{-s}} = \sum \frac{1}{(p_1^{k_1} p_2^{k_2} \cdot \cdots \cdots \cdot p_r^{k_r})^s} \, .$$

We can write this as a sum

$$\sum_{n=1}^{\infty} \frac{a(n)}{n^s}$$

where $a(n)$ is the number of ways of expressing the integer n as a product of prime powers. By the Fundamental Theorem of Arithmetic 3.2, this can be done in one and only one way, i.e. $a(n) = 1$, which proves the product formula, once we check that everything is absolutely convergent if $\mathrm{Re}(s) > 1$. $\square$

As the above argument shows, the product formula is but a form of the Fundamental Theorem of Arithmetic.

8.1. **Analytic continuation and functional equation of $\zeta(s)$.** To further explore the connection between the theory of primes and $\zeta(s)$, we will analytically continue $\zeta(s)$ to all values of s. We use the Gamma function given for $\mathrm{Re}(s) > 0$ by the integral representation

$$\Gamma(s) = \int_0^{\infty} e^{-t} t^s \frac{dt}{t}$$

to define the *completed zeta function* by

$$\zeta^*(s) := \pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s)$$

The basic fact about this variant of $\zeta(s)$ is

Theorem 8.2. 1. *The completed zeta function $\zeta^*(s)$ has a meromorphic continuation to the entire s-plane.*
2. *$\zeta^*(s)$ is analytic except for simple poles at $s = 0, 1$.*
3. *It satisfies the functional equation*

$$\zeta^*(s) = \zeta^*(1-s)$$

As an immediate consequence of this fact, we observe that $\zeta^*(s)$ has no zeros outside the critical strip $0 \le \mathrm{Re}(s) \le 1$. This holds since $\Gamma(s)$ is never zero, and $\zeta(s)$ is analytic and nonzero in the region of convergence $\mathrm{Re}(s) > 1$, so that the completed zeta function $\zeta^*(s) \neq 0$ in $\mathrm{Re}(s) > 1$; by the functional equation, the same is true for the symmetric region $\mathrm{Re}(s) < 0$. Moreover, since $\Gamma(s)$ is analytic except for simple poles at $s = 0, -1, -2, \dots$, $\zeta(s)$ is nonzero in $\mathrm{Re}(s) < 0$ except for simple zeros at the negative even integers $s = -2, -4, \dots$ (to make up for the simple poles of $\Gamma(\frac{s}{2})$ at these points). These are called the *trivial zeros* of $\zeta(s)$; the nontrivial ones are the zeros of $\zeta^*(s)$ and as we have seen they all lie in the critical strip.

Proof. (Sketch) We start with the integral representation

$$\pi^{-s/2}\Gamma(\frac{s}{2})\frac{1}{n^s} = \int_0^\infty e^{-\pi n^2 t} t^{s/2}\frac{dt}{t}$$

which shows that we have an integral representation of $\zeta^*(s)$ for $\mathrm{Re}(s) > 1$ as

$$\zeta^*(s) = \int_0^\infty \frac{\theta(t)-1}{2} t^{s/2}\frac{dt}{t} \tag{8.1}$$

where the theta-function is given for $t > 0$ by

$$\theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t}$$

By Poisson summation, $\theta(t)$ has a transformation formula

$$\theta(\frac{1}{t}) = \sqrt{t}\theta(t) \tag{8.2}$$

Breaking up the region of integration in the integral representation 8.1 to an integral over $(0,1)$ and one over $(1,\infty)$, we change variables $t \mapsto 1/t$ to to transform the integral over (0.1) to one over $(1,\infty)$.

We then use the transformation formula 8.2 for $\theta(t)$ to find after some manipulation that

$$\zeta^*(s) = -\frac{1}{s} - \frac{1}{1-s} + \int_1^\infty \frac{\theta(t)-1}{2}\left(t^{s/2} + t^{(1-s)/2}\right)\frac{dt}{t} \tag{8.3}$$

Since $\theta(t)-1 = O(e^{-\pi t})$ as $t \to \infty$, the integral is absolutely convergent for all s and is therefore an entire function of s. Thus from (8.3) we get the meromorphic continuation, with the only poles being the simple ones at $s = 0, 1$. From the symmetry of (8.3) with respect to $s \mapsto 1-s$ we get the functional equation. □

8.2. **Connecting the primes and the zeros of $\zeta(s)$.** Riemann, in his seminal paper of 1858 [5], used $\zeta(s)$ to give a formula for $\pi(x)$ in terms of the zeros of $\zeta(s)$. His formula gives a clear understanding as to why $\mathrm{Li}(x)$ is the correct approximation to $\pi(x)$. Instead of a formula for $\pi(x)$, it is more convenient to give a formula for the weighted sum of prime *powers* $p^k \le x$, each prime power p^k weighted by the logarithm $\log p$ of the corresponding prime. One defines

$$\psi(x) := \sum_p \sum_{k:p^k \le x} \log p$$

The repetitions p^k for $k \ge 2$ give a contribution of the order of at most $\sqrt{x}$. The primes ($k = 1$) give a contribution which, if one believes Gauss' assertion that the density of primes near x is about $1/\log x$, is about x. Thus we expect (and the above argument is easily made rigorous) that the Prime Number Theorem is equivalent to the assertion that $\psi(x) \sim x$. This is made transparent by the formula (due to von Mangoldt)

$$\psi(x) = x - \sum \frac{x^\rho}{\rho} - \frac{\zeta'}{\zeta}(0) \tag{8.4}$$

where the sum is over all zeros ρ of $\zeta(s)$. Note that we cannot expect the formula to converge absolutely, since it would then define a *continuous* function of x, while $\psi(x)$ is a step function with jumps when $x = p^k$ is a prime power.

The contribution of the trivial zeros $\rho = -2, -4, -6, \dots$ is easily summed to equal $\frac{1}{2}\log(1 - x^{-2})$ and is negligible. The constant term is $\zeta'/\zeta(0) = \log 2\pi$. The important part is the sum over the nontrivial zeros, which we expect to be of smaller order than x. It is thus crucial to understand the distribution of the zeros.

8.3. **The Riemann Hypothesis.** As noted in section 8.1, the nontrivial zeros of $\zeta(s)$ all lie in the critical strip $0 \le \mathrm{Re}(s) \le 1$. If ρ is a zero then by the functional equation $\zeta^*(s) = \zeta^*(1-s)$, so is $1-\rho$, and since $\zeta(\bar{s}) = \overline{\zeta(s)}$ ($\bar{z}$ denoting complex conjugation), we get zeros at $\bar{\rho}$ and $1-\bar{\rho}$ (the two symmetries $s \mapsto \bar{s}$ and $s \mapsto 1-s$ coincide on the "critical line" $\mathrm{Re}(s) = 1/2$).

The first few zeros were computed by Riemann himself, and all lie on the critical line $\mathrm{Re}(s) = 1/2$. They are $\rho_n = 1/2 + it_n$ with $t_1 = 14.13\dots, t_2 = 21.02\dots, t_3 = 25.01\dots$ etc. (by symmetry, we only need to consider positive t).

Riemann's Hypothesis (RH): *All nontrivial zeros of $\zeta(s)$ lie on the critical line* $\mathrm{Re}(s) = 1/2$.

The Riemann Hypothesis has been checked extensively and is widely believed to be true, though an explanation and proof are still missing to date. Its significance to the theory of primes is immense. For instance, we can use RH to explain the small size of the remainder term $\mathrm{Li}(x) - \pi(x)$ in Table 1. To see this, it suffices to show that $\psi(x) - x$ is small, and in fact we shall argue that it is of order at most $\sqrt{x}\log^2 x$. This is reasonable if we look at the formula for $\psi(x)$ in (8.4), which we will write as

$$\psi(x) = x - \sum \frac{x^{1/2+it_n}}{1/2+it_n} + \dots$$

where the sum is now only over the nontrivial zeros, the omitted terms being negligible. If we assume the t_n are *real*, so $|x^{1/2+it_n}| = \sqrt{x}$, it is tempting to then use the triangle inequality to deduce

$$|\psi(x) - x| \le \sqrt{x} \sum \frac{1}{|1/2+it_n|}$$

and so say that $\psi(x) - x$ is of order $\sqrt{x}$. The argument is not quite correct, as it transpires that the sum of absolute values diverges: $\sum 1/|1/2+it_n| = \infty$. Nevertheless, this gives the essence of what is happening, and in fact taking more care and using more information on the distribution of zeros, one can show that $\psi(x) - x \ll \sqrt{x}\log^2 x$. This gives $\pi(x) - \mathrm{Li}(x) \ll \sqrt{x}\log x$ and so explains the observation regarding the size of the third column in Table 1.

The Riemann Hypothesis and its generalization (GRH) to other "L-functions" is one of the most important unsolved problems in Number Theory, and its validity has numerous implications. For instance, there are algorithms for *primality testing* of integers which are proved to require polynomial time (that is, testing if n is prime or not requires a number of operations polynomial in $\log n$), provided we assume GRH.

As to what was actually proved so far, the significant fact is that there are no zeros on the boundary of the critical strip: $\zeta(1+it) \neq 0$, so $0 < \mathrm{Re}(\rho) < 1$ for all nontrivial zeros. This is enough to prove the Prime Number Theorem, as was done (independently) by Hadamard and de la Vallée Poussin in 1896. One has in fact a zero-free region near the boundary of the critical strip, whose width shrinks to zero as we go up. However, to date we do not have a proof that there is any strip of the form $\mathrm{Re}(s) > 1 - \delta$ in which there are no zeros, for any $\delta > 0$.

References

[1] H. Davenport, *The higher arithmetic*, Seventh edition, Cambridge Univ. Press, Cambridge, 1999.

[2] H.M. Edwards *Riemann's Zeta Function*, Academic Press 1974.

[3] G.H. Hardy and E.M. Wright, *An introduction to the theory of numbers* (The Clarendon Press, Oxford University Press, New York, 1979).

[4] M. Murty *Artin's conjecture for primitive roots* Math. Intelligencer **10** (1988), no. 4, 59–67.

[5] B. Riemann *Über die Anzahl der Primzahlen unter einer gegebenen Größe*, Montasb. der Berliner Akad. (1858/60) 671–680, in *Gessamelte Mathematische Werke* 2nd edition, Teubner, Leipzig 1982 no. VII.

Raymond and Beverly Sackler School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel (rudnick@post.tau.ac.il)