Chapter One

Introduction

1.1 THE BEGINNING

Many problems in number theory have the form: *Prove that there exist infinitely many primes in a set* $\mathcal{A}$ or *prove that there is a prime in each set* $\mathcal{A}^{(n)}$ *for all large* n. Examples of the first include:

The twin-prime conjecture. Here one takes $\mathcal{A} = \{p + 2 : \ p \text{ a prime}\}$.

Primes represented by polynomials. A typical problem here is whether the quadratic $n^2 + 1$ is infinitely often prime. So one takes $\mathcal{A} = \{n^2 + 1 : n \in \mathbb{Z}\}$.

Examples of the second problem include:

Goldbach's conjecture. In this case $\mathcal{A}^{(n)} = \{2n - p : p \text{ a prime}, p \le n\}$.

Is there a prime between consecutive squares? For this problem we take $\mathcal{A}^{(n)} = \{m : n^2 \le m \le (n + 1)^2\}$.

I must stress that we will not be able to get as far as a proof of these results in this book! As is well known, the solution to these problems seems to be well beyond all our current methods. Nevertheless, the methods we present here have produced remarkable progress in our knowledge of the distribution of primes in "thin" sequences. It should be clear from reading this book just what information we lack to tackle the above problems.

If we write $\pi(x)$ for the number of primes up to x, the prime number theorem tells us that $\pi(x) \sim x/\log x$. Thus we might hope that if the integers in a set $\mathcal{A}$ are about x in size, and assuming that there are no obstacles preventing primes from belonging to $\mathcal{A}$ (like $\mathcal{A}$ consisting only of even numbers), then the number of primes in $\mathcal{A}$ is about $|\mathcal{A}|/\log x$ (perhaps times some factor depending on the likelihood of small primes dividing integers in $\mathcal{A}$). We shall discover that our hopes are realized so long as we have the two types of arithmetical information introduced in Section 1.6. We shall find that when the information available is strong, we can obtain an asymptotic formula for the number of primes in a given set. When the information is not quite so strong, we can often still obtain a non-trivial lower bound for this quantity. The common thread running through this book is the use of sieve methods: either identities or inequalities. Indeed our inequalities are simply identities where we bound below by zero sums that are in all probability positive. We shall see that the sieve method can be traced back to Eratosthenes in antiquity. Even modern formulations of sieve identities that apparently have no connection with the ancient Greek mathematician turn out to be intimately related through the

ubiquitous identity (1.3.1), which underlies all our work. From another point of view, our work here could be regarded simply as the inclusion/exclusion principle pushed to the nth degree (with $n \to \infty$!).

There are four basic types of problem that we will consider and we introduce them now to whet the reader's appetite.

1. Diophantine approximation. This is the easiest problem to deal with since much progress can be made with relatively elementary arithmetical information. We know that if α is irrational, then there are infinitely many pairs of coprime integers m, n with

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{n^2}.$$

See [57, Theorem 171], for example. Indeed, the right-hand side above can be improved by a factor up to $5^{-1/2}$. Now what if we wanted to have fractions with prime denominator? This is the sort of question a number theorist naturally asks. We would hope to get infinitely many solutions to

$$\left|\alpha - \frac{m}{p}\right| < \frac{1}{p^{1+\theta}},$$

and, as θ increases from 0 (trivial: there is a solution in m for every p) to 1, (the result is false for one: see note below), the problem presumably increases in difficulty. Indeed, no one has any idea how to increase θ above $\frac{1}{3}$ unless one assumes very strong results on primes in arithmetic progressions (stronger than the Generalized Riemann Hypothesis, which gives the $\frac{1}{3}$ exponent). Taking a different perspective on this problem, αn is dense (mod 1) if α is irrational, and one can consider Kronecker's theorem [57, Theorem 440] in the form

$$|\alpha n - m + \beta| < 3n^{-1}.$$

One would then ask about obtaining infinitely many solutions to

$$|\alpha p - m + \beta| < p^{-\theta}.$$

Taking $\beta = 0$ we recover our original problem. It will be useful to write

$$\|x\| = \min_{m \in \mathbb{Z}} |x - m|.$$

Thus the above problems correspond to small values of

$$\|\alpha p\| \quad \text{or} \quad \|\alpha p + \beta\|.$$

Remark. In [61] it is shown that there are uncountably many α such that

$$\|\alpha p\| < \frac{\log p}{500 p \log\log p}$$

has only finitely many solutions in primes p.

2. Primes in short intervals. We would like to know that the interval $[n, n+n^{1/2})$ contains primes for all large n. There is no method known at present that could tackle this problem (unless we assume an extraordinary hypothesis on the existence of Siegel zeros [40]). If we knew the Riemann Hypothesis were true, then we would obtain the expected asymptotic formula for the number of primes in the interval

$[n, n + n^{1/2+\epsilon})$. Here is a case where ϵ makes all the difference! To be more precise, Dirichlet series/polynomial methods only work when the intervals have this greater length. If we are unable to prove that there are primes in the interval $[n, n + n^{1/2+\epsilon})$ without the Riemann Hypothesis, how much larger do we have to make the interval to get an unconditional result? To keep with the convention of the first problem that increasing θ means increasing difficulty, how big can we make θ and the interval $[n, n + n^{1-\theta})$ still contains primes for all large n ? Can we get the expected asymptotic formula

$$\pi(n + n^{1-\theta}) - \pi(n) \ (= \Pi(n, \theta), \text{say}) \ \sim \frac{n^{1-\theta}}{\log n}?$$

We can in fact obtain this formula for $\theta < \frac{5}{12}$ (by Huxley's work [97]) and obtain good upper and lower bounds for larger θ. For example, we can get

$$1.01 > \frac{\Pi(n, \theta) \log n}{n^{1-\theta}} > 0.99$$

for $\theta \leq \frac{9}{20}$, as will be demonstrated in Chapter 10.

Instead of making the intervals longer and asking for primes, we can keep the intervals short and look for "prime-like" numbers. There are two likely candidates for such numbers: almost-primes (with a limited number of prime factors) and numbers with a large prime factor. We shall consider the latter only since conventional sieve methods provide the best answers for almost-primes. We can ask for an integer $m \in [n, n + n^{1/2+\epsilon})$ to have a large prime factor, say $> n^{\theta}$. Again, increasing the value of θ increases the difficulty of the problem, but θ can be taken quite close to 1 ($> \frac{25}{26}$; see Chapter 5 here). For this problem one can reduce the size of the interval to $[n, n + n^{\alpha})$ with $\alpha \leq \frac{1}{2}$, but the best exponent to date for θ, even with $\alpha = \frac{1}{2}$, is now substantially smaller (0.738 proved in [120], we give an improved result in Chapter 6 here).

Instead of considering *all* intervals, we can ask what happens almost always. That is, we consider intervals $[x, x + y(x))$ with $x \leq X$ but allow $o(X)$ exceptions if $x \in \mathbb{N}$. Equivalently, if $x \in [1, \infty)$, we allow a set of exceptional x of measure $o(X)$. Now the Riemann Hypothesis furnishes us with an almost perfect answer: One can take $y(x) = (\log x)^2$. Even without this hypothesis one can still take y as quite a small power of x. Later in this work we shall require both the "all" and the "almost-all" results with primes restricted to arithmetic progressions with small modulus in order to apply the circle method to consider the distribution of Goldbach numbers (numbers represented as the sum of two primes) in short intervals. We shall also generalize our method to discuss Gaussian primes in small regions.

3. Primes in arithmetic progressions. Let $a, q \in \mathbb{N}, (a, q) = 1$. By a famous result due to Dirichlet we know that there are infinitely many primes in the arithmetic progression $a \pmod q$. The next natural questions to ask are: How big is the smallest such prime? How many such primes are there up to N? The smallest known value of C such that $p < q^C, p \equiv a \pmod q$ for all large q has become known as Linnik's constant since Linnik was the first to show that such a C exists.

We would expect that

$$\sum_{\substack{p\le N\\ p\equiv a \bmod q}} 1 \sim \frac{\pi(N)}{\phi(q)}, \tag{1.1.1}$$

where $\phi(q)$ is Euler's totient function. The Generalized Riemann Hypothesis implies this result for $N > q^{2+\epsilon}$. Mention should also be made of recent work by Friedlander and Iwaniec assuming the existence of Siegel zeros [39]. Unfortunately (1.1.1) is only known to be true unconditionally for N substantially larger than q. See [27] for a thorough discussion of this question. However, it is possible to show that (1.1.1) is true on average for $q \le N^{1/2}(\log N)^{-A}$ for some A (the Bombieri-Vinogradov theorem, which we prove in Chapter 2 here). As is well known, it is our ignorance concerning possible zeros of Dirichlet L-functions near the line $\operatorname{Re} s = 1$ that causes a lot of trouble. We are able to replace (1.1.1) with upper and lower bounds over larger ranges of q with our methods. We can also show that for most q Linnik's constant is not much larger than 2. Indeed, for almost all q we can show that it is actually less than 2. These results, although having importance in themeselves, have significance for other problems. For example, we can use average results on primes in arithmetic progressions to give good lower bounds for the greatest prime factor of $p + a$ where p is a prime (see Chapter 8 here). Also, we can use results on primes in most arithmetic progressions to study Carmichael numbers [65]. The techniques we use in this monograph do not appear capable of improving the best known value of Linnik's constant (see [81]), however. The reader will see why later when we consider primes in individual arithmetic progressions subject to a certain condition.

4. Primes represented by additive forms. It is conjectured that if $f(x) \in \mathbb{Z}[x]$ is nonconstant with a positive leading term and irreducible, and if $f(n)$ has no fixed prime divisor for $n \in \mathbb{N}$, then $f(n)$ will take on infinitely many prime values. Dirichlet's theorem shows that this is true for linear polynomials, but there are no known results for higher degree. If one is allowed two variables, the case $p = m^2 + n^2$ is well understood, and the more general case of the sum of two quadratic polynomials has been studied [100]. The first progress toward analogues for higher-degree forms came with the Friedlander and Iwaniec result that $m^2 + n^4$ takes on prime values infinitely often [37]. Indeed they were even able to furnish an asymptotic formula for the number of prime values taken as the region allowed for (m, n) expands. Further work was performed by Heath-Brown [82], who showed a similar result for $x^3 + 2y^3$. We shall prove both of these results in this book, although we do not have the space to provide all the details.

1.2 THE SIEVE OF ERATOSTHENES

The ancient Greek mathematician and astronomer Eratosthenes is credited with being the first to observe that the primes up to a given number, say N, can be found simply as follows. We write down all the integers up to N and take 2 as the first prime; then we cross out all subsequent multiples of 2. In general, find the next

uncrossed number as the next prime (so after 2 we find 3, of course) and cross out all of its multiples. This is easily demonstrated on a piece of paper or an overhead projector with the numbers up to 100, say, but I don't know how to make it look exciting in print! (See Tables 1.1 and 1.2 below) The reader with an internet search should soon find a site that will give an animated version of the sieve, or one could quickly write one's own program to do this task. By the 13th century A.D. it had been noticed that one needs only to cross out multiples of primes up to $\sqrt{N}$ since all composite numbers up to N must have at least one prime factor not exceeding $\sqrt{N}$. The reader will note the problem with the number 1 — we must not cross out all its multiples, yet it is not a prime! Sometimes, even with quite sophisticated arguments it is still necessary to deal with the number 1 separately. Historically, 1 was originally considered to be a prime, of course.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Table 1.1 The integers from 1 to 100 before sieving

1	2	3	x	5	x	7	x	x	x
11	x	13	x	x	x	17	x	19	x
x	x	23	x	x	x	x	x	29	x
31	x	x	x	x	x	37	x	x	x
41	x	43	x	x	x	47	x	x	x
x	x	53	x	x	x	x	x	59	x
61	x	x	x	x	x	67	x	x	x
71	x	73	x	x	x	x	x	79	x
x	x	83	x	x	x	x	x	89	x
x	x	x	x	x	x	97	x	x	x

Table 1.2 After the sieve of Eratosthenes has been applied

Clearly the simple principle inherent in the sieve of Eratosthenes is not restricted just to finding the primes up to N. One can similarly "sieve" any given set of integers $\mathcal{A}$ by crossing out multiples of primes less than the square root of each number concerned. Nor need one only sieve to obtain primes. One could strike out all multiples of primes that divide some given integer q and thereby obtain the

integers in a set coprime to q. As another example, one could cross out multiples of primes congruent to $3 \pmod 4$ to obtain those members of a set that are properly represented as a sum of two squares.

1.3 THE SIEVE OF ERATOSTHENES-LEGENDRE

In 1808 Legendre showed how the Sieve of Eratosthenes could be used to count the number of primes up to x. The crucial point is that one needs to distinguish between numbers that are crossed off once, twice, three times, and so on. If one estimates $\pi(x)$ by

$$\sum_{n\le x} 1 - \sum_{p\le\sqrt{x}} \sum_{p|n\le x} 1,$$

one gets too small a number because many numbers are crossed off more than once. If one uses

$$\sum_{n\le x} 1 - \sum_{p\le\sqrt{x}} \sum_{p|n\le x} 1 + \sum_{p\le\sqrt{x}} \sum_{q<p} \sum_{pq|n\le x} 1,$$

one obtains too large a number because numbers crossed out three times, for example, are counted three times by the final sum. The formula Legendre produced needed a whole string of multiple sums that increase in number with x. Clearly this was in need of some notation to tidy up the expression. Some years later Möbius defined the function $\mu(n)$, which bears his name, by writing

$$\mu(n) = \begin{cases} (-1)^r & \text{if } n \text{ has } r \text{ distinct prime factors,} \\ 0 & \text{otherwise.} \end{cases}$$

We take $\mu(1) = 1$ since 1 has no prime factors and note that $\mu(n)$ is zero whenever n has a square factor exceeding 1. In other words, $\mu(n)$ is only nonzero for *square-free* n.

We then obtain

$$\pi(x) - \pi(x^{\frac{1}{2}}) + 1 = \sum_{\substack{d\le x \\ d|P(x^{1/2})}} \mu(d) \sum_{n\le x/d} 1 = \sum_{\substack{d\le x \\ d|P(x^{1/2})}} \mu(d) \left[\frac{x}{d}\right],$$

where

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

Note that the left-hand side of the Eratosthenes-Legendre formula is $\pi(x)-\pi(x^{1/2})+1$ and not $\pi(x)$ since we "sieve out" the primes up to $x^{1/2}$ and the number 1 is not sieved out at all on the right-hand side.

The formula can be proved directly by noting that

$$\sum_{d|n} \mu(d) = \prod_{p|n} (1-1) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \ge 2. \end{cases} \tag{1.3.1}$$

We shall find that this simple formula lies at the heart of much that follows throughout this book.

Clearly we can modify this formula in many ways. Say we want to determine whether an integer n is free of prime factors from some finite set $\mathcal{P}$. Let Q be the product of the primes in $\mathcal{P}$. We then have

$$\sum_{d|(Q,n)} \mu(d) = \begin{cases} 1 & \text{if } p|n \Rightarrow p \notin \mathcal{P}, \\ 0 & \text{if there is some } p|n, p \in \mathcal{P}. \end{cases}$$

Hence

$$\begin{aligned} \sum_{\substack{n\in\mathcal{A} \\ (n,Q)=1}} 1 &= \sum_{n\in\mathcal{A}} \sum_{d|(Q,n)} \mu(d) \\ &= \sum_{d|Q} \mu(d) \sum_{\substack{n\in\mathcal{A} \\ d|n}} 1. \end{aligned}$$

As a particular example consider the number of integers coprime to a positive integer q in a given interval. Since $\mu(d) = 0$ if d has a squared factor, it makes no difference whether we use q or

$$Q = \prod_{p|q} p$$

(the *square-free kernel* of q) in the above. We get

$$\begin{aligned} |\{n : x < n \le x + y, (n, q) = 1\}| &= \sum_{d|q} \mu(d) \sum_{x<nd\le x+y} 1 \\ &= \sum_{d|q} \mu(d) \left(\left[\frac{x+y}{d} \right] - \left[\frac{x}{d} \right] \right) \\ &= \sum_{d|q} y \frac{\mu(d)}{d} + O\left(\sum_{d|q} 1 \right) \\ &= y \frac{\phi(q)}{q} + O(\tau(q)). \end{aligned}$$

Here we have written $\tau(q)$ for the number of divisors of q and noted that $\phi(q)$ satisfies

$$\frac{\phi(q)}{q} = \prod_{p|q} \left(1 - \frac{1}{p} \right) = \sum_{d|q} \frac{\mu(d)}{d}.$$

Since $\tau(q) = O(q^\epsilon)$ for any $\epsilon > 0$ [57, Theorem 315], we thus obtain a good result for the numbers in an interval that are coprime to q once the interval length is larger than a small power of q (assuming q is sufficiently large). However, this approach is hopeless if one tries to employ it to find the number of primes up to x. We would get

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

However, it is not difficult to show that the error term above is not even $o(x)$ — using Mertens' prime number theorem (Theorem 1.2) its size is seen to be asymptotic to $(6/\pi^2)(1-\log 2)x$. Since $\pi(x) = o(x)$ is trivial (this can be obtained by the above argument by sieving only by the primes up to $\log\log x$, say), we obtain no information. Despite this disappointment we shall find that the formula

$$\pi(x) - \pi\left(\tfrac{1}{2}x\right) = \sum_{\substack{d\le x\\ d\mid P(x^{1/2})}} \mu(d)\left(\left[\frac{x}{d}\right] - \left[\frac{x}{2d}\right]\right)$$

is fundamental to our work as we switch from primes in some given set to primes in the interval $[x/2, x)$ or $[x-y, x)$, where y is of a slightly smaller order than x.

1.4 THE PRIME NUMBER THEOREM AND ITS CONSEQUENCES

In 1860 Riemann [149] formulated a programme to establish the following result, which had been conjectured some 68 years earlier by Gauss.

We have

$$\pi(x) = \mathrm{Li}(x)(1+o(1))$$

as $x \to \infty$, where

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\log y}\,dy.$$

We shall refer to this result, perhaps as stated below with an explicit error term, as the *Prime Number Theorem* and frequently abbreviate this simply to the PNT. In 1896 Hadamard [54] and de la Vallée Poussin [29] independently proved this result. The crux of Riemann's programme is the study of the behaviour of the Riemann zeta-function defined for $\mathrm{Re}\, s > 1$ by

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}. \tag{1.4.1}$$

Riemann proved, by establishing a functional equation, that $\zeta(s)$ possesses an analytic continuation to the whole of the complex plane except for the simple pole at $s = 1$. It should be noted that Euler (1737) had previously used the fact that

$$\zeta(x) \to \infty \text{ as } x \to 1^+$$

to provide a proof of the infinitude of the set of primes.

From our perspective there are three important stages in the proof of the PNT:

1) Replace $\pi(x)$ by

$$\psi(x) = \sum_{n\le x} \Lambda(n),$$

where $\Lambda(n)$ is von Mangoldt's function given by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k,\\ 0 & \text{otherwise.}\end{cases}$$

We can then use partial summation to deduce $\pi(x) \sim \mathrm{Li}(x)$ from $\psi(x) \sim x$, obtaining similar error terms.

2) Relate $\psi(x)$ to the logarithmic derivative of $\zeta(s)$ given by

$$\frac{\zeta'}{\zeta}(s) = -\sum_{n=2}^{\infty} \Lambda(n) n^{-s} \text{ for } \mathrm{Re}\, s > 1 \tag{1.4.2}$$

using contour integration. Since $\zeta(s)$ has a simple pole at $s = 1$, its logarithmic derivative has a simple pole with residue -1 at $s = 1$.

3) Move the integral inside the critical strip $0 < \mathrm{Re}\, s < 1$ and use bounds for $(\zeta'/\zeta)(s)$, together with a zero-free region for $\zeta(s)$ to establish that $\psi(x) \sim x$.

As soon as we start developing sieve methods we will need the PNT and one of the basic results from stage 2 above, namely Perron's formula. When we move to applications involving primes in short intervals, we shall need to use all the above stages, including a more powerful zero-free region than was available in 1896. Since we shall need all the details of the proof of the PNT at some point or other, we therefore depart from historical order and use the best results known today to prove the following.

Theorem 1.1. *For any $\epsilon > 0$ we have*

$$\pi(x) = \mathrm{Li}(x) + O\Big(x \exp\Big(-(\log x)^{\frac{3}{5}-\epsilon}\Big)\Big). \tag{1.4.3}$$

Before commencing to assemble the results we shall require to establish this result, we pause to consider the connection between $\zeta(s)$ and the prime numbers. After all, (1.4.1) apparently has no reference to primes. Well, for $\mathrm{Re}\, s > 1$, we have

$$\zeta(s) \sum_{n=1}^{\infty} \mu(n) n^{-s} = \sum_{n=1}^{\infty} n^{-s} \sum_{d|n} \mu(d)$$

by multiplying the terms one by one and gathering the terms in n^{-s} together. From (1.3.1) the coefficients of n^{-s} are all zero, except for $n = 1$. We thus obtain

$$\zeta(s) \sum_{n=1}^{\infty} \mu(n) n^{-s} = 1,$$

and so

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \mu(n) n^{-s} = \prod_{p} \left(1 - \frac{1}{p^s}\right).$$

It follows that, for $\mathrm{Re}\, s > 1$, we have

$$\zeta(s) = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Then

$$\log \zeta(s) = -\sum_{p} \log\left(1 - \frac{1}{p^s}\right),$$

so that

$$\begin{aligned}\frac{\zeta'}{\zeta}(s) = \frac{d}{ds}\log\zeta(s) &= -\sum_p \frac{p^{-s}\log p}{1-p^{-s}} \\ &= -\sum_p (\log p)\left(\sum_{k=1}^{\infty} p^{-ks}\right) \\ &= -\sum_{n=2}^{\infty}\Lambda(n)n^{-s}.\end{aligned} \tag{1.4.4}$$

This establishes (1.4.2). On the other hand,

$$\begin{aligned}\zeta'(s)(\zeta(s))^{-1} &= -\sum_{n=2}^{\infty}(\log n)n^{-s}\sum_{m=1}^{\infty}\mu(m)m^{-s} \\ &= -\sum_{n=2}^{\infty} n^{-s}\sum_{de=n}\mu(e)\log d.\end{aligned}$$

Now

$$\begin{aligned}\sum_{de=n}\mu(e)\log d &= \log n\sum_{d|n}\mu(d) - \sum_{d|n}\mu(d)\log d \\ &= -\sum_{d|n}\mu(d)\log d \\ &= -\log\left(\prod_{d|n} d^{\mu(d)}\right) \\ &= -\log\left(\prod_{p|n} p^{e(p,n)}\right),\end{aligned}$$

where

$$e(p,n) = -\sum_{\substack{d|n/p\\(d,p)=1}}\mu(d) = \begin{cases}1 & \text{if } n = p^k,\\ 0 & \text{otherwise.}\end{cases}$$

We thus obtain a "different" proof that

$$\frac{\zeta'}{\zeta}(s) = -\sum_{n=2}^{\infty}\Lambda(n)n^{-s}.$$

The point we are making is that there are intimate connections between identities involving $\zeta(s)$ and elementary identities between finite sums involving arithmetical functions. Using the fact that the functions n^{-s} are linearly independent, we can immediately deduce from (1.4.4) that

$$-\sum_{d|n}\mu(d)\log d = \Lambda(n), \tag{1.4.5}$$

an identity we proved directly above and which is often proved in textbooks as an example of Möbius inversion applied to the formula

$$\log n = \sum_{d|n} \Lambda(n) \tag{1.4.6}$$

(see [3, pp. 32–33]). We further note the central role played by the formula (1.3.1) in the Eratosthenes-Legendre sieve and in all of the above working. Finally we remark that (1.4.5) gives a simple decomposition of $\Lambda(n)$. We shall reconsider this in Chapter 2 when we consider Vaughan's identity and related expressions.

The first tool we need to prove the PNT will have many other applications. We give a complete proof in the appendix. In the rest of this section, and indeed in much of the rest of this book, we write $s = \sigma + it$, $\sigma, t \in \mathbb{R}$ for a complex variable. Unless otherwise stated, the variables σ, t, s will always be so related.

Lemma 1.1. *When $\sigma > 1$, let*

$$F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

Write

$$f(x) = \max_{x/2<n<2x} |a_n|.$$

Suppose that

$$\sum_{n=1}^{\infty} \frac{|a_n|}{n^\sigma} = O\big(|1-\sigma|^{-\alpha}\big)$$

as $\sigma \to 1^+$. Then, if $c > 0, \sigma + c > 1$, we have

$$\begin{aligned}\sum_{n\le x} \frac{a_n}{n^s} =& \frac{1}{2\pi i}\int_{c-iT}^{c+iT} F(w+s)\frac{x^w}{w}dw + O\bigg(\frac{x^c}{T(\sigma+c-1)^\alpha}\bigg) \\ &+ O\bigg(\frac{f(x)x^{1-\sigma}\log x}{T}\bigg) + O\bigg(f(x)x^{-\sigma}\min\bigg(\frac{x}{T||x||},1\bigg)\bigg).\end{aligned} \tag{1.4.7}$$

Remark. If $F(s) = \zeta(s)$, then $f(x) \equiv 1$ and $\alpha = 1$. When

$$F(s) = \frac{\zeta'}{\zeta}(s)$$

then $f(x) \approx \log 2x$ but still $\alpha = 1$.

Lemma 1.2. *For $|t| < e, 0 < \delta < 1$, we have, for $\delta \le |\sigma - 1| \le 1$,*

$$\left|\frac{\zeta'}{\zeta}(s)\right| \ll \delta^{-1}.$$

Proof. This follows immediately from the fact that $\zeta(s)$ has a simple pole at $s = 1$ and no zeros in the region $|t| < 4, -1 < \sigma < 3$. □

Lemma 1.3. *Suppose that*

$$|t| \geq e, \ \ -1 \leq \sigma \leq 2, \ \ \min_{\rho} |s-\rho| > (\log|t|)^{-1},$$

where the minimum is over zeros ρ of $\zeta(s)$. Then

$$\left|\frac{\zeta'}{\zeta}(s)\right| \ll (\log|t|)^2.$$

Proof. See [27, p. 99]. The crucial formula, whose consequence we shall also need in Chapter 7, is

$$\frac{\zeta'}{\zeta}(s) = \sum_{\substack{\rho \\ |\gamma - t|<1}} \frac{1}{s-\rho} + O(\log t) \tag{1.4.8}$$

for $s = \sigma + it$, with $t \geq e$ and not coinciding with the ordinate of a zero. □

Lemma 1.4. *For $|t| > e^e$ we have $\zeta(s) \neq 0$ for*

$$\sigma > 1 - \frac{A}{(\log|t|)^{\frac{2}{3}}(\log\log|t|)^{\frac{1}{3}}},$$

where A is an absolute constant.

Proof. See [157, p. 135]. This result is now known with $A > \frac{1}{100}$; see [32]. □

Proof. (PNT) Our immediate goal is to establish that

$$\psi(x) = x + O\Big(x\exp\Big(-(\log x)^{\frac{3}{5}-\epsilon}\Big)\Big). \tag{1.4.9}$$

From Lemma 1.1 we have

$$\sum_{n\leq x}\Lambda(n) = -\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\,ds + O\Big(\frac{x(\log x)^2}{T}\Big), \tag{1.4.10}$$

with $c = 1 + (\log x)^{-1}$. We take the integral to the line $\sigma = \sigma'$, where

$$\sigma' = 1 - \frac{A}{2(\log T)^{\frac{2}{3}}(\log\log T)^{\frac{1}{3}}}.$$

The pole at $s = 1$ gives the main term x in (1.4.9). From Lemma 1.3 the horizontal line contours $|t| = T, \sigma' \leq \sigma \leq c$ contribute

$$\ll \frac{x}{T}(\log T)^2, \tag{1.4.11}$$

which is essentially the same error as given by Perron's formula itself if T and x are fixed powers of each other and leads to a slightly smaller order error for our choice of T below. From Lemmas 1.2 and 1.3 we obtain

$$\begin{aligned}\left|\int_{\sigma'-iT}^{\sigma'+iT}\frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\,ds\right| &\ll x^{\sigma'}\int_{-T}^{T}\frac{(\log(|t|+2)^2}{1+|t|}\,dt \\ &\ll x(\log T)^3\exp\left(-\frac{A\log x}{2(\log T)^{\frac{2}{3}}(\log\log T)^{\frac{1}{3}}}\right).\end{aligned} \tag{1.4.12}$$

To balance the error terms, let $T = \exp((\log x)^\alpha)$ with $0 < \alpha < 1$. The error terms in (1.4.10) and (1.4.11) are then

$$O\big(x \exp(-(\log x)^\alpha)(\log x)^2\big),$$

while the error in (1.4.12) is

$$O\Big(x \exp\Big(-A(\log x)^{1-\frac{2\alpha}{3}}(\log\log x)^{-\frac{1}{3}}\Big)(\log x)^{3\alpha}\Big).$$

If we pick $\alpha = 1 - 2\alpha/3$, that is, $\alpha = \frac{3}{5}$, we get both error terms

$$\ll x \exp\Big(-(\log x)^{\frac{3}{5}-\epsilon}\Big)$$

for any $\epsilon > 0$, which completes the proof of (1.4.9).

Partial summation will play a big role in much that we do, so we shall now give the deduction of (1.4.3) from (1.4.9) explicitly. We recall the basic partial summation identity that, for arbitrary functions f, g and integers $A < B$, we have

$$\sum_{n=A}^{B} f(n)g(n) = \sum_{n=A}^{B}(f(n) - f(n+1)) \sum_{m=A}^{n} g(m) + f(B+1)\sum_{m=A}^{B} g(m).$$

We suppose $x \geq 2$ is an integer and begin by writing

$$\pi(x) = \sum_{2\leq n\leq x} \frac{1}{\log n} + \sum_{2\leq n\leq x}\left(\frac{\Lambda(n)-1}{\log n}\right) - \sum_{\substack{2\leq n\leq x\\ p^2\mid n}} \frac{\Lambda(n)}{\log n}$$

$$= S_1 + S_2 - S_3 \quad \text{say.}$$

Let $L = [\log_2 x]$ ($\log_2$ indicating logarithm to base 2 here). Trivial bounds yield

$$S_3 \ \leq \ \sum_{k=2}^{L} x^{\frac{1}{k}} \ \leq x^{\frac{1}{2}} + Lx^{\frac{1}{3}} \ \ll \ x^{\frac{1}{2}}.$$

For any positive function $f(x)$ that is monotonically decreasing we have

$$\sum_{U\leq n\leq V} f(n) = \int_U^V f(y)\,dy + O(f(U)),$$

so

$$S_1 = \mathrm{Li}(x) + O(1).$$

Now write

$$\sum_{2\leq m\leq n} (\Lambda(m) - 1) = E(n),$$

which gives $E(n) \ll x\exp\Big(-(\log x)^{\frac{3}{5}-\epsilon}\Big) = F(x)$, say, when $n \leq x$. Partial summation then gives

$$S_2 = \sum_{2\leq n\leq x}\left(\frac{1}{\log n} - \frac{1}{\log(n+1)}\right)E(n) + \frac{E(x)}{\log(x+1)}.$$

Thus

$$|S_2| \ll F(x)\left(\sum_{2\leq n\leq x}\left(\frac{1}{\log n} - \frac{1}{\log(n+1)}\right) + \frac{1}{\log(x+1)}\right) = \frac{F(x)}{\log 2}.$$

This completes the proof of (1.4.3). □

Before developing the PNT to cover sums over primes and almost-primes, we pause to give a result that will be of crucial importance when we consider primes in short intervals and which is obtained by a slight modification of the proof given above. We shall often use a splitting-up argument, as is common in analytic number theory, to produce variables in ranges that are of the same order of magnitude. For this purpose we introduce the notation $a \sim A$ to mean $A \le a < 2A$.

Lemma 1.5. *Let $V \ge e, P \ge 2$. Then, if $|t| \sim V$, we have*

$$\left|\sum_{p\sim P} p^{-s}\right| \ll P^{1-\sigma}\exp\left(-\frac{\log P}{(\log V)^{\frac{7}{10}}}\right) + \frac{P^{1-\sigma}}{V}(\log P)^3.$$

Remark. In applications V and P will be related in such a way that we will obtain

$$\left|\sum_{p\sim P} p^{-s}\right| \ll P\exp\bigl(-(\log P)^{\beta}\bigr)$$

for some $\beta > 0$.

Proof. We leave this as an exercise for the reader. The major differences are that we take $T = \frac{1}{2}V$ and there is no pole on crossing the line $\sigma = 1$. □

We now need some more notation that will be used throughout the rest of this book. First we define Buchstab's function $\omega(u)$ by $\omega(u) = u^{-1}$ for $1 \le u \le 2$ and then use induction to define $\omega(u)$ for $k \le u \le k+1$ ($k \in \mathbb{N}, k \ge 2$) by

$$\omega(u) = \frac{1}{u}\left(k\omega(k) + \int_k^u \omega(v-1)\,dv\right).$$

It follows immediately from this that $\frac{1}{2} \le \omega(u) \le 1$ for all u and that ω is the continuous solution to the delay/differential equation

$$(u\omega(u))' = \omega(u-1)$$

with the boundary condition $\omega(u) = u^{-1}$ for $1 \le u \le 2$. We shall prove in the appendix that $\omega(u) \to \exp(-\gamma)$ as $u \to \infty$, where γ is Euler's constant.

Write $\mathcal{B} = \mathbb{Z} \cap [x/2, x)$, $\mathcal{A} \subseteq \mathcal{B}$, and put

$$\begin{aligned} S(\mathcal{A}, z) &= |\{n \in \mathcal{A} : p|n \Rightarrow p \ge z\}| \\ &= \sum_{\substack{d|P(z)\\ dn\in\mathcal{A}}} \mu(d). \end{aligned}$$

Clearly the number of primes in $\mathcal{A}$ is $S(\mathcal{A}, x^{1/2})$ and

$$S(\mathcal{B}, x^{\frac{1}{2}}) = \pi(x) - \pi\left(\frac{x}{2}\right) + O(1),$$

where the $O(1)$ arises only if x or $x/2$ is a prime.

We write

$$\mathcal{E}_d = \{n : nd \in \mathcal{E}\}.$$

The inclusion/exclusion principle then quickly gives Buchstab's identity for any set $\mathcal{E}$ and positive reals $z > w > 1$:

$$S(\mathcal{E}, z) = S(\mathcal{E}, w) - \sum_{w \le p < z} S(\mathcal{E}_p, p). \tag{1.4.13}$$

By the PNT we have

$$S(\mathcal{B}, z) \sim \frac{x}{2 \log x} \text{ for } x^{\frac{1}{2}} < z < \tfrac{x}{2}.$$

Then, for $x^{1/3} \le w < x^{1/2}$, (1.4.13) gives

$$\begin{aligned} S(\mathcal{B}, w) &= S(\mathcal{B}, x^{\frac{1}{2}}) + \sum_{w < p \le x^{1/2}} S(\mathcal{B}_p, p) \\ &= S(\mathcal{B}, x^{\frac{1}{2}}) + \sum_{w < p \le x^{1/2}} S(\mathcal{B}_p, (x/p)^{\frac{1}{2}}) \end{aligned}$$

since

$$\frac{x}{p} > p > \left(\frac{x}{p}\right)^{\frac{1}{2}}$$

in this range. Now the PNT gives

$$S(\mathcal{B}_p, (x/p)^{\frac{1}{2}}) = \frac{x}{2p \log(x/p)} \left(1 + O((\log x)^{-1})\right).$$

Hence

$$S(\mathcal{B}, w) = \frac{x}{2 \log x} \left(1 + \sum_{w < p \le z} \frac{\log x}{p \log(x/p)}\right) \left(1 + O((\log x)^{-1})\right).$$

Partial summation as before then gives (the details are contained in the appendix where we consider the error terms in greater detail)

$$S(\mathcal{B}, w) = \frac{x}{2 \log x} \left(1 + \int_w^z \frac{\log x}{y(\log(x/y))(\log y)} dy\right) \left(1 + O((\log x)^{-1})\right).$$

The change of variables $v = (\log x)/(\log y)$ gives

$$\frac{dy}{y \log y} = -\frac{dv}{v},$$

and so

$$\begin{aligned} S(\mathcal{B}, w) &= \frac{x}{2 \log x} \left(1 + \int_2^{\frac{\log x}{\log w}} \frac{1}{v - 1}\, dv\right) \left(1 + O((\log x)^{-1})\right) \\ &= \frac{x}{2 \log x} \left(1 + \log\left(\frac{\log x}{\log w} - 1\right)\right) \left(1 + O((\log x)^{-1})\right). \end{aligned}$$

We thus conclude that

$$S(\mathcal{B}, w) \sim \frac{x}{2 \log x} u\omega(u) = \frac{x}{2 \log w} \omega(u) \tag{1.4.14}$$

when $w = x^{1/u}, 2 \le u \le 3$. It does not take much imagination to see that an inductive argument should establish (1.4.14) for all $u > 1$, and this is indeed established in the appendix.

In general we will be faced with multiple sums over primes of the form

$$\sum_{p_1,\dots,p_k} S(\mathcal{B}_{p_1\dots p_k}, z(p_1,\dots,p_k)),$$

where

$$p_j \in \mathcal{I}(p_1,\dots,p_{j-1})$$

and $\mathcal{I}(p_1,\dots,p_{j-1})$ is some interval like

$$\left[w, \min\left(p_{j-1}, \left(\frac{x}{p_1,\dots,p_{j-1}}\right)^{\frac{1}{2}}\right)\right].$$

Working as above, such as sum will be asymptotically equal to

$$\frac{x}{2}\int\cdots\int \frac{1}{\alpha_1\dots\alpha_k \log z}\omega(u)\, d\alpha_k \dots d\alpha_1,$$

where

$$u = (1-\alpha_1-\cdots-\alpha_k)\frac{\log x}{\log z},$$

and the integration range for α_j is

$$\left[\frac{\log A}{\log x}, \frac{\log B}{\log x}\right],$$

where $\mathcal{I} = [A, B]$. For example, the sum

$$\sum_{x^\nu \le p_2 < p_1 < x^{1/2}} S(B_{p_1p_2}, p_2)$$

becomes

$$\frac{x}{2\log x}\int_\nu^{1/2}\int_\nu^{g(\alpha_1)} \omega\left(\frac{1-\alpha_1}{\alpha_2}-1\right)\frac{1}{\alpha_1\alpha_2^2}\, d\alpha_1\, d\alpha_2. \tag{1.4.15}$$

Here

$$g(\alpha) = \min\left(\alpha, \frac{1-\alpha}{2}\right),$$

where we have noted that $S(B_{p_1p_2}, p_2) = 0$ when $p_1p_2^2 > x$. Thus the summation condition $p_2 < p_1$ becomes $p_2 < \min\left(p_1, (x/p_1)^{1/2}\right)$.

We will often need upper bounds for integrals of the above type, and the following approximation to $\omega(s)$ is then useful (compare Diagram 1.1):

$$\omega(u) \begin{cases} = 1/u & \text{if } 1 \le u \le 2, \\ = (1+\log(u-1))/u & \text{if } 2 \le u \le 3, \\ \le \frac{1}{3}(1+\log 2) & \text{if } u \ge 3. \end{cases} \tag{1.4.16}$$

We note that $\frac{1}{3}(1+\log 2) = 0.56438\dots$, whereas $\lim_{u\to\infty}\omega(u) \approx 0.56146$. The behaviour of $\omega(u)$ for small u is illustrated below. With this scale, the graph is practically flat (local maxima and minima cannot be seen) for larger values of u.

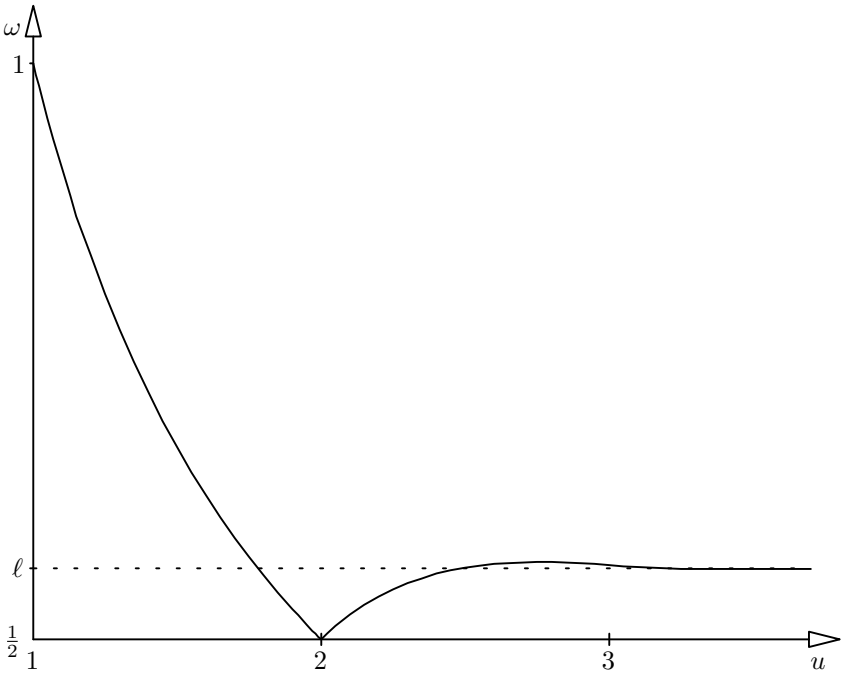

Diagram 1.1: $\omega(u)$ with $\ell = e^{-\gamma}$

Historically, Mertens' prime number theorem preceded the proof of the PNT proper by 22 years [127]. The two equivalent formulations of his result are still of great importance, and we shall need one or other version at various points in this work.

Theorem 1.2. (Mertens' Prime Number Theorem) *As $x \to \infty$, we have*

$$\sum_{p \le x} \frac{1}{p} = \log\log x + C + o(1), \tag{1.4.17}$$

where C is the constant

$$\gamma + \sum_{p} \left(\log\left(1 - \frac{1}{p}\right) + \frac{1}{p} \right)$$

and γ is Euler's constant. Rearranging and taking the exponential of the above formula then gives

$$\prod_{p \le x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x}. \tag{1.4.18}$$

Proof. See [57, Theorems 427–429]. Only real variable techniques are used in this proof with no reference to $\zeta(s)$. □

1.5 BRUN, SELBERG, AND ROSSER-IWANIEC

Historically, the next advance on the sieve of Eratosthenes-Legendre came about from the work of Brun [20], developing ideas in a paper by Merlin [126]. If one takes the view that the problem with (1.3.1) is that d takes on values that are too large, then it is natural to consider upper and lower bounds for the basic sifting function

$$\sum_{d|n} \mu^-(d) \le \sum_{d|n} \mu(d) \le \sum_{d|n} \mu^+(d), \tag{1.5.1}$$

where $\mu^{\pm}(d) = 0$ for $d > D$. If this is possible, then, using $\mathcal{A}$ as in the previous section,

$$S^-(\mathcal{A}, z) \le S(\mathcal{A}, z) \le S^+(\mathcal{A}, z),$$

with

$$\begin{aligned} S^{\pm}(\mathcal{A}, z) &= \sum_{\substack{d|P(z)\\ d\le D}} \mu^{\pm}(d) \sum_{dn\in\mathcal{A}} 1 \\ &= \sum_{\substack{d|P(z)\\ d\le D}} \mu^{\pm}(d)|\mathcal{A}_d|. \end{aligned}$$

If, taking the simplest class of problems, for some X,

$$|\mathcal{A}_d| = \frac{X}{d} + R_d$$

and

$$\sum_{d\le D} |R_d| = o\left(\frac{X}{\log x}\right), \tag{1.5.2}$$

then one might hope to get, as $x \to \infty$,

$$S^-(\mathcal{A}, z) \ge f\frac{X}{\log x}, \qquad S^+(\mathcal{A}, z) \le F\frac{X}{\log x},$$

where f, F depend in some way on z and D.

This line of approach is taken in the standard text by Halberstam and Richert [55] and in its more recent successor by Greaves [51]. There are three basic methods to obtain (1.5.1). The original technique of Brun was based on combinatorial ideas (Chapter 2 in [55], Chapter 3 in [51]). Selberg's idea is, in the first instance, only applicable to obtaining the upper bound in (1.5.1) but is based on the simple idea of ensuring that

$$\sum_{d|n} \mu^+(d)$$

is a square. The reader will find expositions of this method in Chapter 2 of [51] or Chapter 3 in [55]. The third line of attack was provided by Rosser (unpublished) and developed by Iwaniec [102] (see also Chapter 4 of [51]). We shall give a new

derivation of this sieve in Chapter 4 since its construction shares many features with the sieve method we present, indeed our method collapses to the Rosser-Iwaniec approach in the limit. We shall also need to apply this sieve in tandem with our method for two of the later problems (see Chapters 6 and 8).

The sieves of Brun, Selberg, and Rosser have enabled much progress to be made on some of the extremely difficult problems of number theory. Highlights include Chen's theorem [21] stating that all sufficiently large even integers are the sum of a prime and a number with at most two prime factors. Mention must also be made of the recent work on small differences between primes [48] and arithmetic progressions of primes [52]. However, by themselves, these sieves cannot generate primes. This will be discussed in more detail in Chapter 4. It is worthwhile to consider here what arithmetical information is being fed into these sieves. There is some "general" information (which can be used for many different problems) that is used to produce $\mu^{\pm}(d)$ and to estimate sums of the form

$$\sum_{\substack{d|P(z)\\ d\le D}} \frac{\mu^{\pm}(d)}{d}$$

or, more generally,

$$\sum_{\substack{d|P(z)\\ d\le D}} \frac{w(d)\mu^{\pm}(d)}{d},$$

where $w(d)$ is a multiplicative function, subject to certain natural constraints. Then there is more specific information (which varies according to the problems being considered) that is employed to give upper bounds of the form (1.5.2). For example, to consider the problem of integers in short integers without small prime factors, we take $\mathcal{A} = \mathbb{N} \cap [x-y, x)$. We put $X = y$ to give

$$R_d = \sum_{x-y\le n<x} 1 \;-\; \frac{y}{d} \;=\; O(1).$$

It follows that it is possible to take D nearly as large as y. In many applications it is possible to replace (1.5.2) with a multiple sum

$$\sum_{\substack{m\le M\\ n\le N}} a_m b_n R_{mn}$$

and use more sophisticated ways of estimating this term. Indeed, Iwaniec and Jutila [105] were able to improve Huxley's prime number theorem by using flexible error terms in the Rosser-Iwaniec sieve along with extra information on the problem, namely, that one could directly estimate certain multiple sums of the form

$$\sum_{p_1p_2\in\mathcal{A}} 1, \qquad \sum_{p_1p_2p_3\in\mathcal{A}} 1. \tag{1.5.3}$$

This idea was further refined by Heath-Brown and Iwaniec [85]. In this last work, D was taken as large as $x^{0.92}$ when $y = x^{0.55+\epsilon}$. We shall need their results, which lead to this size of D, later in Chapter 10. Again it should be stressed that this is not sufficient in itself to detect primes — it is the additional information specific to the problem from (1.5.3) that is required to supplement the sieve methods.

1.6 ERATOSTHENES-LEGENDRE-VINOGRADOV

In the 1920s Hardy and Littlewood showed that every sufficiently large odd number is the sum of three primes assuming the Generalized Riemann Hypothesis for Dirichlet L-functions. This assumption was necessary to provide nontrivial bounds for exponential sums of the form

$$\sum_{p\le N} e(p\alpha)$$

when α was not too well approximable by a rational with a small denominator. Here, as elsewhere in this book, we write $e(x) = \exp(2\pi i x)$. This sum arises naturally from an application of the circle method [161] to the problem. In this way we have

$$\sum_{p_1+p_2+p_3=N} 1 = \int_0^1 \left(\sum_{p\le N} e(p\alpha)\right)^3 e(-\alpha N)\, d\alpha.$$

Vinogradov (see [163, Chapter 9]) found an ingenious argument that enabled the sieve of Eratosthenes-Legendre to be applied to this sum. We should note that Vinogradov was applying the sieve not to the immediate arithmetical situation (primes p_j such that $n = p_1+p_2+p_3$), as is the case for the sieves described in the previous section, but to the auxiliary functions that arise from the application of a standard number-theoretic method. As a consequence, Vinogradov's technique had to be "precise" — he could not throw away terms of the same size as the main term as happens in the other sieves. Of course, the Eratosthenes-Legendre sieve is just such a precise result.

Suppose we want to count primes with a weight $f(p)$. If $|f(n)| \le 1$ for all n, then the sieve of Eratosthenes-Legendre becomes

$$\sum_{p\le x} f(p) = \sum_{\substack{dn\le x \\ d|P(x^{1/2})}} \mu(d) f(dn) + O(x^{\frac{1}{2}}).$$

At first sight it might seem that we have transformed one difficult problem (a sum over primes) into an equally difficult problem — the Möbius function seems no easier to handle than the characteristic function of the set of primes. Indeed we have turned a neat-looking sum into a rather messy-looking double sum. It is a general principle in number theory that if you cannot estimate a sum, you might be able to estimate it "on average," and the presence of two variables enables us to do just that, as we shall see.

Suppose that the numbers $f(p)$ were complex numbers on the unit circle, and we had some hope that these numbers were fairly evenly distributed, so that there would be some cancellation in the sum over p. We might then want to show that

$$\sum_{p\le x} f(p) = O(F(x)),$$

where we would like $F(x)$ to grow as slowly as possible ($\pi(x)$ is the trivial bound, of course!). We will take the particular example $f(n) = e(\alpha n)$ which, as explained

above, was Vinogradov's original application. This will help us to illustrate the principles at work in estimating double sums and will be applied later to investigate the distribution of αp modulo 1. Using the formula for the sum of a geometric progression, if $\alpha \notin \mathbb{Z}$, we have

$$\sum_{n=1}^{N} e(\alpha n) = \frac{e((N+1)\alpha) - e(\alpha)}{e(\alpha) - 1}.$$

Now,

$$|e(\alpha) - 1| = \left|e\left(\tfrac{1}{2}\alpha\right) - e\left(-\tfrac{1}{2}\alpha\right)\right| = 2|\sin(\pi\alpha)|.$$

Also, for $-\frac{1}{2} \le \alpha \le \frac{1}{2}$ we have $|\sin(\pi\alpha)| \ge 2|\alpha|$. Since $|\sin(\pi\alpha)|$ is periodic with period 1, and recalling the notation

$$\|x\| = \min_{n\in\mathbb{Z}} |x - n|,$$

we have $|\sin(\pi\alpha)| \ge 2\|\alpha\|$. Thus, using the trivial bound for small values of $\|\alpha\|$ we obtain

$$\left|\sum_{n=1}^{N} e(\alpha n)\right| \le \min\left(N, \frac{1}{2\|\alpha\|}\right),$$

with the obvious convention that N is taken as the minimum when $\|\alpha\| = 0$.

Now

$$\sum_{p\le x} e(\alpha p) = \sum_{\substack{dn\le x \\ d|P(x^{1/2})}} \mu(d) e(\alpha dn) + O(x^{\frac{1}{2}}).$$

Vinogradov's vital contribution to the analysis of this problem was to split the consideration of the sum on the right-hand side above according to the size of d. We shall not describe his method since the method we develop in Chapter 3 does essentially the same job with much greater clarity (the obscurity of Vinogradov's method was a great barrier to its application by other authors). The methods we describe in Chapter 2 also lead to equivalent results with far less pain than Vinogradov's procedure.

The first type of sum, which we will call a *Type I sum* makes use of the fact that n runs over consecutive integers. We therefore need to restrict the range of d, say, to $d \le D$. This D corresponds to the distribution level in the sieves described in the previous section. We thus obtain

$$\begin{aligned}\left|\sum_{\substack{dn\le x \\ d|P(x^{1/2}), d\le D}} \mu(d) e(\alpha dn)\right| &\le \sum_{\substack{d\le D \\ d|P(x^{1/2})}} \left|\sum_{n\le x/d} e(\alpha dn)\right| \\ &\le \sum_{d\le D} \min\left(\frac{x}{d}, \frac{1}{\|d\alpha\|}\right).\end{aligned}$$

If α is irrational, or a rational with the right size denominator, we might hope that this final sum is smaller than x in size. Say $\alpha = a/q$ (and the following reasoning

works with $\alpha = a/q + O(q^{-2})$ as well). Then, for $d \equiv 0 \pmod q$, we can only give the trivial bound x/d. On the other hand, as d runs through a set of nonzero residues (mod q), $\|d\alpha\|$ runs through the numbers $\|h/q\|, h = 1, \dots, q-1$ in some order. Hence each of the values $h/q, 1 \le h \le q/2$, is taken at most twice. Thus

$$\sum_{d=1}^{q-1} \frac{1}{\|d\alpha\|} \le 2 \sum_{d=1}^{q/2} \frac{q}{d} < 3q \log q.$$

Hence

$$\begin{aligned} \sum_{d \le D} \min\left(\frac{x}{d}, \frac{1}{\|\alpha d\|}\right) &\le \sum_{\substack{d \equiv 0 \pmod q \\ d \le D}} \frac{x}{d} + \sum_{\substack{d \not\equiv 0 \pmod q \\ d \le D}} \frac{1}{\|\alpha d\|} \\ &\le \frac{x}{q} \log\left(\frac{3D}{q}\right) + \left(\frac{D}{q} + 1\right) 3q \log q \\ &\le 3\left(\frac{x}{q} + D + q\right) \log x. \end{aligned}$$

We have thus established a nontrivial bound for this Type I sum if $x^\epsilon < q < x^{1-\epsilon}, D < x^{1-\epsilon}$.

More generally the above work shows that, for any given function such that $1 \le rf(r) \le N$, we have

$$\sum_{r \le R} a_r \sum_{n \le f(r)} e(\alpha n r) \ll \max_{r \le R} |a_r| \left(\frac{N}{q} + R + q\right) \log(NRq). \tag{1.6.1}$$

The much more difficult question is to deal with

$$\sum_{\substack{dn \le x \\ d | P(x^{1/2}), d \ge D}} \mu(d) e(\alpha d n).$$

Vinogradov's significant contribution was to convert such sums into sums of the following type which we shall call *Type II sums*:

$$\sum_{\substack{m \sim M, h \sim H \\ mh \le x}} a_m b_h e(mh\alpha) = S, \text{ say.}$$

Here a_m, b_n may be a bit messy, but the important feature is that they are bounded by something like the divisor function. That is, we can assume $|a_m| \le \tau(m), |b_n| \le \tau(n)$. We apply Cauchy's inequality, namely,

$$\left(\sum_{r=1}^{H} c_r d_r\right)^2 \le \left(\sum_{r=1}^{H} |c_r|^2\right) \left(\sum_{r=1}^{H} |d_r|^2\right),$$

to the Type II sum above to obtain

$$|S|^2 \le \sum_{m \sim M} |a_m|^2 \sum_{m \sim M} \left| \sum_{\substack{h \sim H \\ mh \le x}} b_h e(\alpha m h) \right|^2.$$

Now since a_m is bounded by the divisor function, we have

$$\sum_{m\sim M} |a_m|^2 \ll M\log^3 M.$$

We treat

$$\sum_{m\sim M}\left|\sum_{\substack{h\sim H\\ mh\le x}} b_h e(\alpha mh)\right|^2$$

by squaring out and changing the order of summation (our idea here is to get a sum over consecutive integers from the m range that we can then treat like the Type I sum above) to get

$$\begin{aligned}\sum_{h_1,h_2\sim H} b_{h_1}b_{h_2}\sum_{\substack{mh_j\le x\\ m\sim M}} e(m(h_1-h_2)\alpha)&\\ \le \sum_{h_1,h_2\sim H}|b_{h_1}b_{h_2}|\left|\sum_{\substack{mh_j\le x\\ m\sim M}} e(m(h_1-h_2)\alpha)\right|&\\ \ll \sum_{h_1,h_2\sim H}|b_{h_1}b_{h_2}|\min\left(\frac{x}{H},\frac{1}{\|\alpha(h_1-h_2)\|}\right).&\end{aligned}$$

Now if we use the arithmetic mean/geometric mean inequality to obtain

$$|b_{h_1}b_{h_2}|\le \tfrac{1}{2}\left(|b_{h_1}|^2+|b_{h_2}|^2\right),$$

we are left to estimate

$$\left(\sum_{h\sim H}|b_h|^2\right)\max_{k\sim H}\sum_{\ell\sim H}\min\left(\frac{x}{H},\frac{1}{\|\alpha(\ell-k)\|}\right).$$

The sum over h above leads to a term $\ll H\log^3 H$ since b_h is bounded by the divisor function. The sum over ℓ can be treated in a similar manner to the sum over d we had in Type I sums. There is the complication now, however, of the maximum over k. This only affects the terms corresponding to $d\equiv 0\,(\mathrm{mod}\,q)$ before. We thus obtain for the sum over ℓ the following expression:

$$\ll\left(\frac{H}{q}+1\right)\left(\frac{x}{H}+q\right)\log x.$$

Drawing all our information together we thereby conclude that

$$|S|\ll x^{\frac{1}{2}}\left(\frac{x}{q}+\frac{x}{H}+q+H\right)^{\frac{1}{2}}(\log x)^{\frac{7}{2}}.$$

This is nontrivial for $x^\epsilon < q < x^{1-\epsilon}$, $x^\epsilon < H < x^{1-\epsilon}$. Note that now we need a *lower* bound on one of the variables as well as an upper bound. Later we will want to apply this estimate with other bounds on the coefficients, so we note that we have actually proved that

$$|S|\ll\left(\sum_{m\sim M}|a_m|^2\sum_{h\sim H}|b_h|^2\right)^{\frac{1}{2}}\left(\frac{x}{q}+\frac{x}{H}+q+H\right)^{\frac{1}{2}}(\log x)^{\frac{1}{2}}. \qquad (1.6.2)$$

Exercises

1. Square-free numbers are easier to detect than primes, and only Type I information is required. Show that

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

Hence deduce that the number of square-free integers between x and $x+y$ (where $y \leq x$) is

$$\frac{6y}{\pi^2} + O\left(\sqrt{x}\right).$$

2. Modify the proof of the PNT to show that

$$\left| \sum_{n \leq N} \mu(n) \right| \ll N \exp(-(\log N)^{\alpha})$$

for some $\alpha > 0$.

3. The Liouville function is given by $\lambda(n) = (-1)^{t(n)}$, where $t(n)$ is the *total* number of prime factors of n (so $\lambda(n)$ is the totally multiplicative version of $\mu(n)$). Show that

$$\sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} = \frac{\zeta(2s)}{\zeta(s)}.$$

Hence, working similarly to Exercise 2, prove that

$$\sum_{n \leq X} \lambda(n) \ll X \exp\left(-(\log X)^{\frac{1}{2}}\right).$$